Opportunistic Spectrum Access via Periodic Channel Sensing

Qianchuan Zhao, *Member, IEEE,* Stefan Geirhofer, *Student Member, IEEE,*
Lang Tong†, *Fellow, IEEE,* and Brian M. Sadler, *Fellow, IEEE*

***Abstract*— The problem of opportunistic access of parallel channels occupied by primary users is considered. Under a continuous-time Markov chain modeling of the channel occupancy by the primary users, a slotted transmission protocol for secondary users using a periodic sensing strategy with optimal dynamic access is proposed. To maximize channel utilization while limiting interference to primary users, a framework of constrained Markov decision processes is presented, and the optimal access policy is derived via a linear program. Simulations are used for performance evaluation. It is demonstrated that periodic sensing yields negligible loss of throughput when the constraint on interference is tight.**

Index Terms—**Dynamic spectrum access, resource allocation, and constrained Markov decision processes.**

I. Introduction

Opportunistic Spectrum Access (OSA), as part of the hierarchical dynamic spectrum access paradigm [1], allows a secondary user to access channels when primary users are not transmitting. To design the optimal strategy for the secondary access, two conflicting objectives arise: on the one hand, the spectrum utilization is to be optimized by exploiting unused network resources: time, frequency, and codes. On the other hand, opportunistic access of a secondary user must not affect the primary users' communications. Specifically, the level of interference caused by the secondary users needs to be kept below a prescribed tolerance level. Thus there are tradeoffs between being aggressive and being polite, between achieving spectrum efficiency and providing a quality-of-service guarantee.

The first step in the design of optimal OSA is the modeling of the dynamic behavior of the primary users, which depends on the specific application. We assume a simple two-state Markovian model in this paper for primary users on each channel. Coupled with the proposed periodic sensing strategy, this model allows us to formulate and solve the optimal OSA problem practically with reasonable computation cost. Such a model is not always justified, of course, but experimental studies on the IEEE 802.11 Wireless LAN (WLAN) support a semi-Markovian model for various traffic patterns (ftp, http, and VoIP) [4], and the Markovian model can be a reasonable approximation in some if not in all traffic regimes. The benefit of such a model is a simple and practical access strategy that satisfies prescribed interference constraints.

The next step is optimizing the access protocol. To seize transmission opportunities left by the primary users and limit the interference, a secondary user needs to sense before transmitting [5], and it needs to decide on which channel to sense and which channel to transmit. Thus the crux of OSA is to optimize the access policy by exploiting traffic dynamics and sensing history.

A. Related Work and Contributions

There are several recent surveys on opportunistic spectrum access, see *e.g.* , [1], [2] and a recent collection of papers in [3]. We highlight here some related hierarchical access schemes in the taxonomy of dynamic spectrum access [1], [8] and summarize the main contributions of this work.

A substantial amount of work exists in exploiting spectrum opportunities in the spatial domain, where a secondary user transmits at locations where the primary users are not affected. See [1] and references therein. We focus in this paper on the utilization of temporal white space. The framework used here arises from [6], [7] where a Markovian traffic model is first introduced and optimal sensing and access strategies developed. In that work, a secondary user senses only some of the available channels, thus the overall state of the network is partially observable. Assuming that both primary and secondary users have the same transmission slot structure, the authors of [7] derive the optimal and suboptimal spectrum sensing and access strategies under the formulation of finite-horizon Partially Observable Markov Decision Processes (POMDP). The slotted structure makes the problem of imposing constraints on interference trivial unless sensing is unreliable, in which case the the authors of [9] are able to show a separation principle that decouples sensing from accessing.

In this paper, we formulate the problem differently from [7] in several ways; most significant is that the transmissions of primary users are unslotted, and the traffic model of primary users is a continuous-time Markov chain. The use of the continuous-time Markovian model raises several complications. For a slotted network, if a secondary user

†Corresponding author.

Q. Zhao (zhaoqc@tsinghua.edu.cn) is with the Center for Intelligent and Networked Systems, Dept Automation, Tsinghua University, Beijing, 100084, China. S. Geirhofer and L. Tong ({sg355@,ltong@ece.}cornell.edu.) are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853. B. Sadler (bsadler@arl.army.mil) is with the Army Research Laboratory, Adelphi, MD 20783-1197.

This paper was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The work was done when Qianchuan Zhao was visiting Cornell University as a visiting Professor. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. Part of this work is to be presented at the IEEE 2007 WCNC. Qianchuan Zhao received additional support from NSFC grant no. 60574067 and the National 111 International Collaboration Project of China.

correctly senses the channel to be idle, then the transmission of the secondary user will not cause interference to the primary user (assuming of course perfect slot synchronization). For the unslotted network considered here, however, there is always a chance that the transmission of the secondary user interferes that of the primary user since the primary user may start to transmit at any time[1]. Therefore, the problem of finding the optimal access policy under interference constraints is nontrivial.

The optimization and sensing strategies proposed in this paper are also quite different from those in [7]. Zhao *et al.* in [7] develop the optimal policy under the finite-horizon POMDP formulation that has a complexity growing exponentially with the duration of the transmission. Here we consider an infinite-horizon optimization where the complexity does not grow with the length of the transmission. Note that the corresponding infinite-horizon POMDP problem is much more complicated [11], [12].

The main contributions of this paper are as follows. Assuming that multiple primary user channels evolve independently as continuous-time Markov chains, we propose an access scheme referred to as Periodic Sensing Opportunistic Spectrum Access (PS-OSA). The key idea of PS-OSA is to remove the partial observability by sensing the available channels periodically. While restricting to periodic sensing is suboptimal in general, the proposed scheme significantly reduces the complexity required by the optimal OSA proposed in [7] under the POMDP framework. When constraints on interference levels are imposed, we are able to formulate the problem as a constrained Markov decision process (CMDP) [15] and solve for the optimal policy via a linear program. A slight generalization is needed, however, because of the periodicity of the induced Markov chain. Simulation examples are presented to demonstrate a number of properties of the proposed approach, including its performance gap to the optimal (fully observable) scheme and the robustness of the algorithm against parameter perturbation. It is shown that when the constraints on interference are tight, the performance loss of PS-OSA is negligible.

B. Organization and Notation

This paper is organized as follows. The system model is introduced in Section II. The periodic sensing strategy is described in Section III where we specify the sensing protocol and give the mathematical description of the Markovian system induced by the sensing protocol. Properties of the Markov chain are also provided. Next we present the optimal PS-OSA in Section IV. Actions, rewards, and costs are defined first followed by the formulation of the MDP problem. A solution based on linear programming is then presented. In Section VI, we present simulation examples aimed at illustrating the performance and the robustness of the proposed algorithm. The paper concludes by summarizing our results and stating the limitations and future directions.

[1] We assume that primary users do not backoff due to secondary user transmissions. This might be a restrictive assumption if primary users employ random rather than scheduled access protocols.

Notations used in this paper are mostly standard and summarized in the Appendix. In general, random variables are capitalized and their realizations are in lower case. In addition, the indicator function of a set $\mathcal{X}$ is denoted as $1_{\mathcal{X}}$

II. System Model

Assume that there are N parallel channels (indexed from 0 to $N-1$) available for transmissions by the primary and secondary users. Consider a hierarchical access scheme in which the primary users access these channels according to a certain protocol (scheduled or random access) and a secondary user tries to access one of the N channels opportunistically.

We assume that the occupancy of each channel by a primary user evolves independently according to a homogeneous continuous-time Markov chain with idle ($X_i = 0$) and busy ($X_i = 1$) state, respectively. This is motivated by unslotted transmissions of WLANs. Experimental results indicate that the traffic of WLAN users can be adequately modeled as a continuous-time semi-Markov process [10], [13], [14]. We note that the simplifying Markovian assumption, though not necessarily accurate across the entire traffic regime, seems to have a reasonably good fit with measurement data [10].

Due to the Markovian assumption, the holding times are exponentially distributed with parameters λ_i^{-1} for the idle and μ_i^{-1} for the busy states, respectively. We stress that the primary system is *not* slotted; primary users can access the channel at any time.

In contrast to the primary users, the secondary user employs a slotted communication protocol (consider Bluetooth as a practical example). In each slot the secondary user (i) senses one of the N channels at the beginning of the slot, (ii) uses the sensing result to decide if and in which channel to transmit, and (iii) receives an acknowledgement by the secondary receiver if the transmission is successful.

The proposed scheme can easily be generalized to cases when the sensing of and the transmission across multiple channels is possible. For ease of presentation, we restrict ourselves to single channel sensing and transmission in this paper, which gives rise to the partial observability of the Markov process. Such a restriction can occur with existing hardware, so the OSA solution for this case can potentially be implemented with legacy systems.

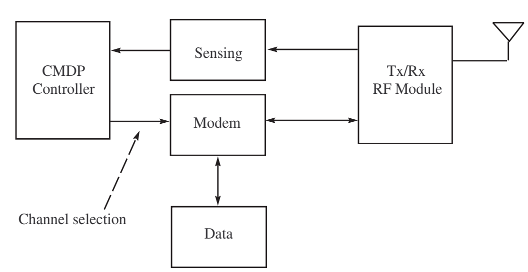

Fig. 1. System block diagram.

A block diagram of the system is shown in Figure 1. The signal captured by the antenna is passed through an analog front end and sampled within the sensing block. A decision is made on whether the primary user is present, and this sensing result is passed on to a controller that decides whether it is safe to transmit (and if yes, in which channel). If a transmission occurs, the secondary user's data are fed to the transmit modem which in turn interfaces the analog front end.

We assume that synchronization is maintained between the secondary sender and receiver. Indeed, periodic sensing simplifies synchronization since sender and receiver need not coordinate their sensing pattern. If the sensor readings (busy or idle) are the same at the secondary user sender and receiver, synchronization is maintained by using the same random seed. When the sender and receiver have different sensing results, there is a probability that the transmitter and the receiver will tune to different channels, and the ensuing transmission of course fails. The lack of acknowledgement, on the other hand, makes both ends aware that a sensing error occurred in the previous slot. They can then set the previous sensing result to a predetermined value. Additionally, acknowledgements and signaling information can be multiplexed with data to ensure synchronization. The implementation details are not considered in this paper, although we do provide simulation results that include cases when sensing errors occur.

III. Periodic Sensing Opportunistic Spectrum Access

We assume that the secondary user cannot sense all channels at the same time. This is motivated by the need of developing access protocols without adding an additional multi-channel sensor to receivers. On the other hand, this assumption makes the problem of finding an optimal access strategy challenging since the state of the system at any time is only *partially observed*. In this paper, we render the problem tractable by postulating a periodic sensing approach, referred to as Periodic Sensing Opportunistic Spectrum Access (PS-OSA). We thus decouple the sensing and the access parts of the problem. While imposing a periodic sensing strategy is in general suboptimal, it leads to a fully observable Markov Decision Process and simplifies the optimal protocol design considerably.

A. Sensing and Transmission Structures of PS-OSA

We describe here the PS-OSA protocol for the secondary user, leaving the optimization of the protocol to Section IV.

Recall that the secondary user operates in a slotted fashion. The sensing protocol is periodic with period N equal to the number of available channels.[2] The access protocol, on the other hand, depends on the sensing result and is not periodic.

Figure 2 illustrates the sensing and transmission events of the secondary user. Each protocol period contains N slots. Without loss of generality, we can assume that the secondary user senses the channel in an increasing order, starting from the smallest index (say channel 0). At the beginning of each slot, the secondary user senses the channel. Based on this and all past sensing results, the secondary user takes an action of either transmitting on one of the N channels or not transmitting at all. Notice that we allow the secondary user to transmit in a different channel from that it has just sensed. See the third slot in Figure 2.

[2]The proposed scheme applies easily to the case when the protocol period is greater than N.

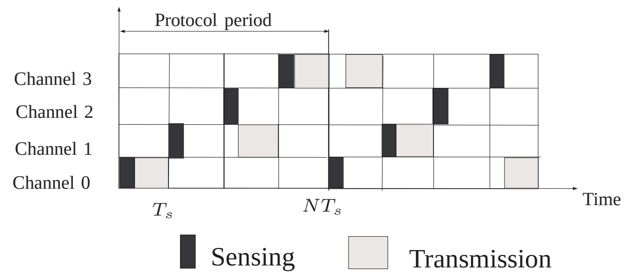

Fig. 2. Sensing and transmission structure for an $N = 4$ channel system.

B. Induced Markov chain

We derive in this section the Markovian structure for PS-OSA. At the beginning of the k-th slot, $I_k \stackrel{\Delta}{=} [kT_s, (k+1)T_s]$, channel $q = k \bmod N$ is sensed, where T_s denotes the slot size, and 'mod' denotes the modulus operation.

With periodic sensing, after sensing is completed in the k-th slot I_k, we define an N-dimensional vector random process $\mathbf{Z}(k) = [Z_0(k), \cdots, Z_{N-1}(k)]^T$ by

$$Z_i(k) = \begin{cases} X_i(kT_s), & \text{if } i = k \text{ mod } N, \\ Z_i(k-1), & \text{otherwise} \end{cases} \quad (1)$$

for $i = 0, 1, \ldots, N-1$ with $k = N, N+1, \ldots$ as its discrete time index. Here N is the number of channels, and $\mathbf{Z}(k)$ contains the sensing results of the most recent N slots. When sensing is active in channel i, the i-th component of $\mathbf{Z}(k)$ is updated with the measurement of the state of i-th channel at the beginning of time slot I_k.

The Markov chain that describes the observed process also depends on the "age" (in terms of number of slots) of sensing result for channel i. Let $q \stackrel{\Delta}{=} k$ mod N be the position of the slot in the current the N-slot protocol period. If channel i is sensed in slot k, then the sensing result has the age of $\tau(i,k) = 0$. In the $(k+1)$th slot, the next channel is sensed, and the age of the sensing result for channel i is $\tau(i,k) = 1$. In general

$$\tau(i,q) \stackrel{\Delta}{=} (N + q - i) \bmod N \quad (2)$$

We are now ready to state the Theorem that gives the Markov chain description of the observed traffic dynamics.

Theorem 3.1: Consider the N parallel channels with traffic modeled by independent binary-state continuous-time Markov chains. For channel i, let λ_i^{-1} be the mean holding time for state 0 and μ_i^{-1} for state 1, and denote the transition rate (generator) matrix by

$$\mathbf{Q}_i \stackrel{\Delta}{=} \begin{pmatrix} -\lambda_i & \lambda_i \\ \mu_i & -\mu_i \end{pmatrix}, \quad i = 0, \cdots, N-1. \quad (3)$$

Then the vector process $\mathbf{Z}(k)$, $k = N, N+1, \ldots$, defined in (1) is a discrete-time Markov chain. Let $q' \stackrel{\Delta}{=} (k+1)$ mod N

be the channel sensed in slot $k+1$. The transition probability of $\mathbf{Z}(k)$ is given by

$$\begin{aligned}&\Pr(\mathbf{Z}(k+1)=\mathbf{z}'|\mathbf{Z}(k)=\mathbf{z})\\ &=\begin{cases}[\exp(\mathbf{Q}_{q'}NT_s)]_{(z_{q'},z'_{q'})} & \text{if } z'_i=z_i, \forall i\neq q'\\ 0 & \text{otherwise}\end{cases}\end{aligned} \tag{4}$$

where $[\exp(\mathbf{Q}_{q'}NT_s)]_{(z_{q'},z'_{q'})}$ is the transition probability of chain $X_{q'}$ (over time NT_s) from state $z_{q'}$ to $z'_{q'}$.
Proof. See appendix.

The periodicity of the Markov chain comes naturally from the periodic sensing employed in PS-OSA. Since every state of $\mathbf{Z}(k)$ is recurrent and $\Pr(\mathbf{Z}(k+1)|\mathbf{Z}(k))$ depends only on q, we also have the following theorem.

Theorem 3.2: The process $\mathbf{Z}(k)$ is irreducible and periodic with period N. For each $q=0,1,\ldots,N-1$, the process $\mathbf{Z}(pN+q)$, $p=1,2,\ldots$, has the stationary distribution

$$f_q(\mathbf{z})=\prod_{i=0}^{N-1}\left(1_{[z_i=0]}v_i(0)+1_{[z_i=1]}v_i(1)\right), \tag{5}$$

where $1_{[\cdot]}$ denotes the indicator function and

$$v_i(0)=\frac{\mu_i}{\lambda_i+\mu_i},\quad v_i(1)=\frac{\lambda_i}{\lambda_i+\mu_i}. \tag{6}$$

Proof. See appendix.

IV. Optimal PS-OSA

Having characterized the Markov chain induced by the primary user and the adopted slot structure for the secondary user, we need to add a control dimension to our problem. Specifically, after each sensing operation, we can either choose to transmit in one of the N channels or, alternatively, not transmit at all. In this section, we formulate the decision problem of the secondary user as a Constrained Markov Decision Process (CMDP). We start with specifying actions and rewards, introduce throughput and interference, and finally convert the CMDP to an equivalent linear programming (LP) problem.

A. Actions and rewards

Let the action chosen in slot I_k under policy π be denoted as $A_k\in\mathcal{A}=\{-1,0,\ldots,N-1\}$; choosing $A_k\geq 0$ symbolizes transmission in the A_k-th channel whereas $A_k=-1$ means no transmission.

If we choose to transmit, we accrue a reward when the transmission is successful or incur a cost otherwise. For simplicity, we assume here that an unsuccessful transmission incurs cost only if there is a collision with the primary user. (One can, of course, include cases when the transmission is not reliable even in the absence of collision.) It is stressed that even if a channel has just been sensed idle, a collision can still occur since the primary user's medium access is *not* slotted.

Let us define the reward $r(\mathbf{z},a,k)$ accrued by a successful transmission in slot k with sensing result $\mathbf{z}$ and action $A_k=a$ as

$$r(\mathbf{z},a,k)\triangleq\begin{cases}\Pr(X_a(t)=0,\forall t\in I_k|\mathbf{Z}(k)=\mathbf{z}) & a\geq 0\\ 0 & a<0\end{cases} \tag{7}$$

Note that the above reward is the (conditional) mean successful rate. Analogously, we can define the cost of choosing action a as

$$c(\mathbf{z},a,k)=\begin{cases}1-r(\mathbf{z},a,k) & a\geq 0\\ 0 & a<0\end{cases}, \tag{8}$$

which is the probability that the transmission leads to a collision with the primary user. The following theorem gives the expression of reward (also for the cost through (8)).

Theorem 4.1: The immediate reward in the k-th slot can be analytically evaluated by

$$r(\mathbf{z},a,k)=\begin{cases}g(\mathbf{z},k\bmod N,a), & a\geq 0\\ 0, & a=-1\end{cases} \tag{9}$$

where

$$g(\mathbf{z},q,i)\triangleq\exp(-\lambda_iT_s)\left[\exp(\mathbf{Q}_i\tau(i,q)T_s)\right]_{(z_i,0)}. \tag{10}$$

Proof. See appendix.

It is worthwhile to note the special case where $a\geq 0$ and channel a is sensed at kT_s, i.e., $\tau(a,k)=0$. In this case, we have

$$r(\mathbf{z},a,k)=1_{[z_a=0]}\exp(-\lambda_aT_s). \tag{11}$$

That is, when $z_a=0$ and we transmit in channel a, the immediate reward will be $\exp(-\lambda_aT_s)$; when $z_a=1$ and we transmit in channel a, no reward will be obtained.

B. The CMDP formulation

Here we aim to maximize the throughput of the secondary system while abiding by hard constraints on the level of interference. Mathematically, we can formulate this goal as maximizing the average number of successful transmissions (of the secondary user),

$$J(\pi)=\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\mathrm{E}_\pi(r(\mathbf{Z}(k),A_k,k)), \tag{12}$$

where the expectation is taken over the probability distribution induced by a policy π.

At the same time, we have to abide by the constraints on interference to individual primary users. Since the interference only occurs when the secondary user is attempting to transmit in a time slot where the channel is not empty, under policy π and for the primary user in channel i, we define the asymptotic ratio of collision and successful transmission slots of the primary user as a measure for the degree of the interference due to the presence of the secondary user. In particular,

$$C_i(\pi)\triangleq\lim_{K\to\infty}\frac{\sum_{k=1}^{K}\mathrm{E}[c(\mathbf{Z}(k),i,k)\Pr(A_k=i|\pi,\mathbf{Z}(k),k)]}{\mathrm{E}(B_i(K))} \tag{13}$$

where $B_i(K)$ is the total number of slots occupied by the primary user in channel i up to time KT_s, and $\Pr(A_k=i|\pi,\mathbf{Z}(k),k)$, the probability that channel i is chosen by policy π for the secondary user to transmit, given sensing result $\mathbf{Z}(k)$ at time kT_s.

The stochastic optimization problem is thus

$$\max_\pi J(\pi) \tag{14}$$

subject to

$$C_i(\pi) \leq \gamma_i, \quad \forall i \in \{0, \ldots, N-1\} \tag{15}$$

where $0 \leq \gamma_i \leq 1$ are given constants.

The problem thus falls into the category of constrained Markov decision processes (CMDPs) [15], [16] and can be solved by a linear program as will be shown in the next section. It is well known that the optimal solution to a CMDP is, in general, randomized. The policy π is thus represented by a mapping from the set of observations $\mathbf{z}$ and $q = k \bmod N$ to the probability that we choose action i.

Notice that our problem is a special type of CMDP in the sense that the underlying Markov chain $\mathbf{Z}(k)$ is not affected by the actions chosen by the decision maker[3]. As a CMDP, it is special also because the rewards $r(\mathbf{z}, a, k)$ and costs $c(\mathbf{z}, a, k)$ at each k are not time independent, instead, they are periodic. Using a similar argument as in [16], it can be shown that our CMDP problem always has an optimal solution.

C. Linear Programming Solution

In this subsection, we will provide a linear programming solution to the CMDP problem formulated above in Equations (14) and (15).

Let the probability that we choose action $i \geq 0$ based on $\mathbf{z}$ and q be denoted by $\beta_{q,i}(\mathbf{z})$. No transmission takes place with probability $\beta_{q,-1}(\mathbf{z}) = 1 - \sum_{i=0}^{N-1} \beta_{q,i}(\mathbf{z})$. We first define a linear programming problem as follows:

$$\max_{\beta} \frac{1}{N} \sum_{q=0}^{N-1} \sum_{\mathbf{z} \in \mathcal{B}^N} f_q(\mathbf{z}) \sum_{i=0}^{N-1} g(\mathbf{z}, q, i) \beta_{q,i}(\mathbf{z}) \tag{16}$$

subject to

$$\sum_{q=0}^{N-1} \sum_{\mathbf{z} \in \mathcal{B}^N} \frac{f_q(\mathbf{z})(1 - g(\mathbf{z}, q, i))\beta_{q,i}(\mathbf{z})}{N(1 - v_i(0)\exp(-\lambda_i T_s))} \leq \gamma_i, \quad \forall i \tag{17}$$

$$\sum_{a \in A} \beta_{q,a}(\mathbf{z}) = 1, \forall q, \mathbf{z}, \quad \beta_{q,i}(\mathbf{z}) \in [0, 1], \forall q, \mathbf{z}, a \tag{18}$$

where $f_q(\mathbf{z})$ is the stationary distribution defined in (5).

We can establish the following theorem.

Theorem 4.2: The linear programming problem in (16)-(18) is equivalent to the CMDP problem in (14)-(15).
Proof. See appendix.

Once the solution $\beta = (\beta_{q,a}(\mathbf{z}), a \in \mathcal{A}, q \in \{0, 1, \ldots, N-1\}, \mathbf{z} \in \mathcal{B}^N)$ has been obtained for this linear program, the secondary user stores it as a table. The secondary user's policy given the observations $\mathbf{z}$ and position q in a period is to flip a biased coin with probability $\beta_{k \bmod N, i}(\mathbf{Z}(k))$; it transmits in channel i, and with probability $\beta_{k \bmod N, -1}(\mathbf{Z}(k))$ no transmission occurs. The optimality of β implies that the optimal performance $J(\pi^*)$ of the CMDP problem (14)-(15) can always be achieved by a randomized periodic policy found through the linear program (16)-(18). Although the optimal value of the linear program is unique, its solution may not be unique. In fact, when the constraints are not tight, there might be feasible solutions allowing transmitting during a busy slot. In this case throughput is the same as the optimal throughput but they have higher collision probabilities although still lower than the given level of γ_is. Among linear program solutions, we always use the one choosing not to transmit in a busy channel for the obvious reason that such a transmission yields no reward and only causes collisions.

[3]This is an idealization under the assumption that the primary users' access protocol is independent of the actions of the secondary users.

V. Suboptimal Static Access Protocols

Under periodic sensing, with the analytical expressions given in Section IV for the immediate reward and collision probability, we introduce two simple heuristic protocols that are easy to implement. They can be used for comparisons as lower bounds of the achievable throughput under constraints on collision with primary users.

A. Memoryless Access (MA)

We consider the following simplified strategy. Under periodic sensing, if in the k-th slot, the secondary user senses a busy channel $q = k \bmod N$, no transmission is made. If the channel is free, it will transmit in the sensing channel q with probability β_q^{MA}. The transmission probability β_q^{MA} is decided such that collision constraints are satisfied while maximizing the throughput for the secondary user. For given levels of allowed collision γ_i, this is equivalent to requiring that the probability of collision in k-th slot is below $\alpha_q = \gamma_q N * (1 - v_q(0)\exp(-\lambda_q T_s))$ Denote this heuristic policy as π^{MA}. It is straightforward to show that the transmission probability is given by

$$\beta_q^{\mathrm{MA}} = \min\left(\frac{\alpha_q}{1 - \exp(-\lambda_q T_s)}, 1\right) \tag{19}$$

and the throughput of this policy is

$$J(\pi^{\mathrm{MA}}) = \frac{1}{N} \sum_{q=0}^{N-1} v_q(0)\beta_q^{\mathrm{MA}} \exp(-\lambda_q T_s), \tag{20}$$

where $v_q(0) = \frac{\mu_q}{\lambda_q + \mu_q}$ is the stationary probability for Channel q to be idle.

B. Greedy Access (GA)

Here we consider a greedy approach to DSA. Given $\mathbf{Z}(k) = \mathbf{z}$ and sensing channel $q = k \bmod N$, compute the probability $g(\mathbf{z}, q, i) = \Pr(X_i(t) = 0, \forall t \in I_k | \mathbf{Z}(k) = \mathbf{z})$ in each channel i being idle in slot I_k. Choose the channel $i^* = \arg\max_i g(\mathbf{z}, q, i)$ which is most likely idle. Transmit in Channel i^* with probability $\beta_q^{\mathrm{GA}}(z)$ such that collision constraints are satisfied while maximizing the throughput for the secondary user. For given levels of allowed collision γ_i, this is equivalent to require that $c(\mathbf{z}, q, i^*)$ in slot I_k is below $\alpha_q = \gamma_q N * (1 - v_q(0)\exp(-\lambda_q T_s))$. Denote this heuristic policy as π^{GA}. It is easy to show that the transmission probability is

$$\beta_q^{\mathrm{GA}}(\mathbf{z}) = \min\left(\frac{\alpha_q}{1 - \max_i g(\mathbf{z}, q, i)}, 1\right) \tag{21}$$

and the throughput of this policy is

$$J(\pi^{\text{GA}}) = \frac{1}{N}\sum_{q=0}^{N-1}\sum_{\mathbf{z}\in S} f_q(\mathbf{z})\beta_q^{\text{GA}}(\mathbf{z}) \max_i g(\mathbf{z}, q, i). \quad (22)$$

This strategy is similar to the greedy approach in [6].

VI. Numerical Examples

In this section we present three numerical and simulation examples: one on the performance of the optimal policy under periodic sensing, the second on the robustness of the optimal policy against perturbations of primary users' traffic parameters, and the third on the robustness of the optimal policy in the presence of sensing errors.

In our experiments and calculations, the choices of λ and μ are motivated from experiments conducted in [4]. In particular, the parameters are chosen based on a VoIP application ("Skype" conference call session) with three participating parties. The idle-times, although showing some heavy-tailed behavior, can be approximated by an exponential distribution with parameter $\lambda^{-1} = 4.2\,\text{ms}$. We assume $\mu^{-1} = 1\,\text{ms}$ for the channel's busy period.

Example 1. Performance of the optimal policy under periodic sensing

In this example, we focus on the case $N = 6$ and consider the tendency of throughput increase as we loosen the interference constraints. By assuming a slot size $T_s = 0.25\text{ms}$, we obtain the throughput characteristics in Fig. 3 and the collision probability (shown only for the first channel) in Fig. 4. We compare with a benchmark protocol that assumes full observability (FO) of all channels at the beginning of every slot[4]. Note that the MDP based on FO gives an upper bound on performance. Two other heuristic protocols, (MA and GA) described in Section V, are also compared; they serve as lower bounds on throughput since they give feasible yet suboptimal solutions to the linear program.

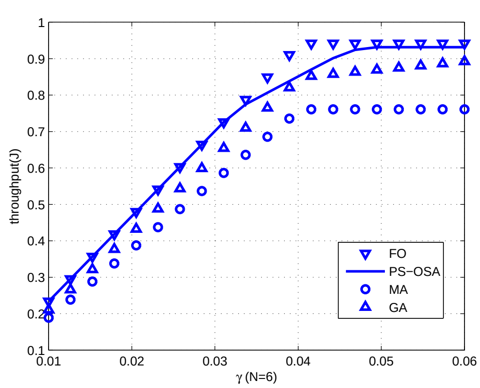

Fig. 3. Throughput of secondary user using optimal periodic sensing

We observe in Fig. 3 that PS-OSA has the performance close to the upper bound (FO) when the constraint is tight, *viz.*,

[4]The full observation case is the standard CMDP problem that admits the same linear programming solution.

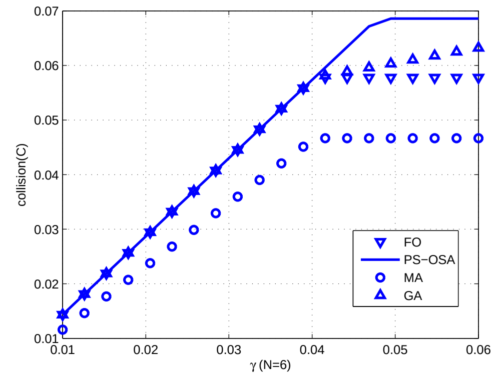

Fig. 4. The first primary user's collision probability with the secondary. The range of interference level is within the interval $[0, 0.06]$ and $\gamma = \gamma_i$.

$\gamma \in [0, 0.033]$. The optimal PS-OSA matches that of the Full Observation (FO), and both curves grow linearly with the value of γ. In the region where γ becomes larger ($\gamma \in [0.034, 0.05]$), there is a loss in the throughput of PS-OSA. When γ becomes large enough, the throughout PS-OSA matches that of the full observation again and approaches a maximum constant value.

The reason behind this trend can be intuitively understood as follows. When γ is small, the constraint on interference in each channel is so restrictive that the maximum achievable throughput is directly limited by the allowed level of collisions. The increase in throughput is proportional to the amount of relaxation in the level of the constraints. When γ is large, there is essentially no constraint on interference. In such a case, both PS-OSA and FO solve unconstrained problems whose solutions are insensitive of the value of γ.

Fig. 3 also includes the performance of two suboptimal but simpler techniques. The GA protocol achieves 80% of the throughput of PS-OSA. One advantage of the heuristic lies in its simplicity when the strategy needs to adjust frequently in response to frequent changes in parameters of the primary channels. The MA protocol, on the other hand, seems too conservative by heavily penalizing a collision in the next slot.

Simulation results on collision probability shown in Fig. 4 further support the above analysis. The first primary user's collision probability with the secondary user is equal to γ in the region $\gamma \in [0, 0.039]$, and less than γ in the region $\gamma \in [0.04, 0.06]$. The reason is that when γ is small, the throughput is limited by the restriction imposed by small collision probability with the primary user; when γ is large enough, the constraint on the first user is no longer active, by taking advantage of the transmission opportunity fully, the secondary user's throughput can be maximized. The maximal value of collision probability is below 1 because we assume that the secondary user never transmits in a channel sensed as busy.

Example 2. Robustness to parameter perturbations

In this example, we evaluate the robustness of the optimal solution when the parameters of primary users deviate from their assumed norms. The setting of the experiment is the same as in Example 1 except we allow $\pm 5\%$ deviations of

λ. Fig. 5 and Fig. 6 show the results for throughput and collisions, respectively. It is clear that both throughput and

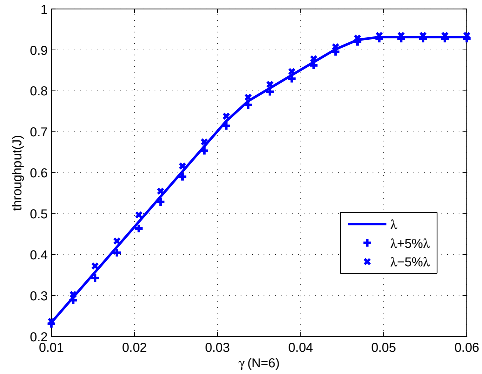

Fig. 5. The effect of primary user traffic parameter change on throughput

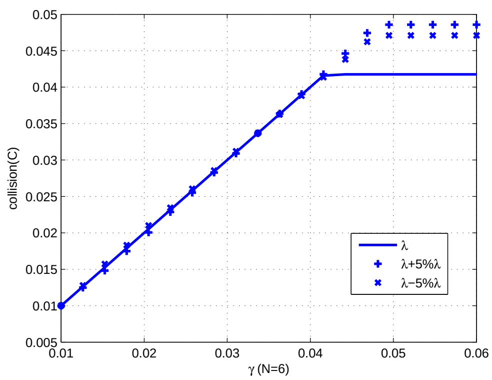

Fig. 6. The effect of primary user traffic parameter change on collision probability

interference change slightly as the parameter λ increases or decreases slightly. It is also reasonable that λ^{-1} represents the average length of idle period, so the increase in λ leads to a decrease in length of idle period, resulting in lower throughput.

Example 3. Robustness to traffic model

In this example, we evaluate the robustness of the optimal solution when the Markovian traffic model is violated. Based on the analysis in [4], the following more realistic traffic model is used. The busy period is constant and equal to 1ms. The idle periods follow a mixture distribution

$$F(t) = p_c F_c(t) + p_f F_f(t)$$

where $p_c = p_f = 0.5$, $F_c(t)$ is the uniform distribution on the interval $[0, 0.7\,\text{ms}]$ and $F_f(t)$ is the generalized Pareto distribution with parameter $k = -0.255$ and $\sigma = 0.01$, $F_f(t) = 1 - \left(1 + \frac{kt}{\sigma}\right)^{-1/k}$. The mean value of the idle time is 4.2 ms. The other experimental settings are the same as in Example 1. The simulation results for a total of 20,000 slots are shown in Fig. 7 and Fig. 8 where the Markovian benchmark is labeled as (Th). The non-Markovian curve is labeled as (NM).

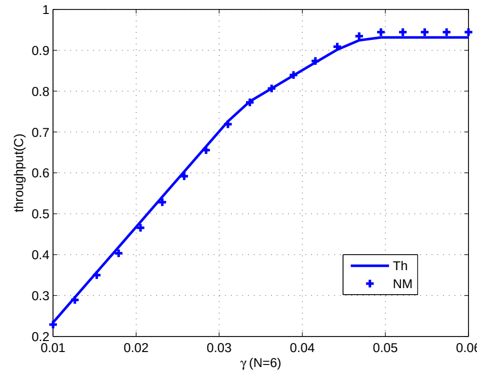

Fig. 7. The throughput for non-Markovian traffic model

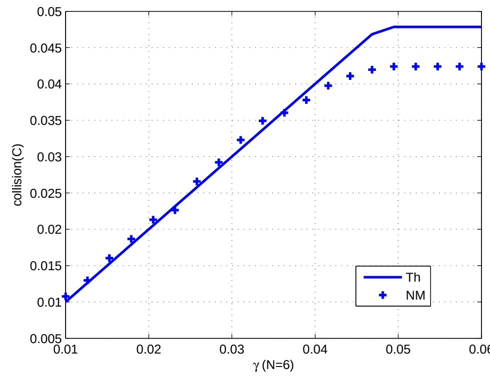

Fig. 8. The collision for non-Markovian traffic model

When the Markovian traffic model is violated, our simulation shows that the throughput only varies slightly. The difference is less than 4% over the region $\gamma \in [0, 0.06]$. There are about the same number of collisions over the region $\gamma \in [0, 0.04]$ and less collisions over the region $\gamma \in (0.04, 0.06]$.

Example 4. Robustness to sensing errors

In this example, we evaluate the robustness of the optimal solution when the channel sensing is not perfect. The probability of sensing the state of each channel correctly is 0.95. Other settings of the experiment are the same as in Example 1. The simulation results for a total of 20,000 slots are shown in Fig. 9 and Fig. 10 where the noiseless benchmark is labeled as (Th).

When observation noise is added, as expected, our simulation shows that the throughput (SN) decreases. The degradation caused by noise is less than 17% over the region $\gamma \in [0.00, 0.03]$ and less than 6% over the region $\gamma \in [0.03, 0.06]$. Due to the sensing noise, the collisions increase to some extent. This may be problematic when the collision constraint is restrictive. One way to deal with this problem is to require tighter γs in the linear program.

VII. Conclusion

We have considered the problem of sharing spectrum in the time domain by exploiting idle periods between bursty

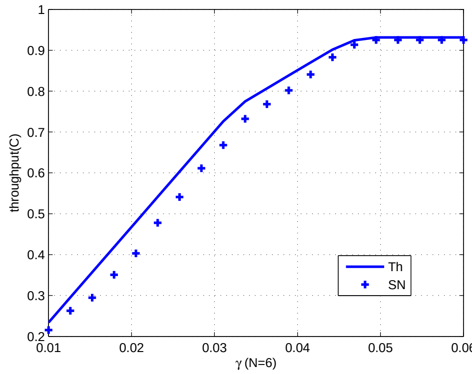

Fig. 9. The effect of observation noise on throughput

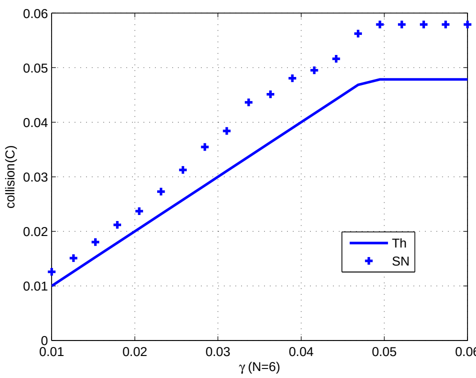

Fig. 10. The effect of observation noise on collision

transmissions of a primary user. By focusing on a periodic sensing scheme, we are able to formulate the problem as a constrained Markov decision process (CMDP), and find the optimal randomized control policy using a linear programming technique. We have also introduced two heuristic protocols which are easier to implement (without need to solve the linear program). We have evaluated the methods' performance numerically. Our results show that the periodic sensing, while limiting the set of admissible policies, is close to the best achievable performance when all channels can be sensed simultaneously.

We have omitted a number of issues in favor of a simpler presentation. Some of these issues can be easily addressed, but others require a more elaborate investigation. For example, the results of this paper can be easily generalized to the case when multiple channels can be sensed simultaneously [17], resulting in improved performance. We have also examined how performance improves with the increase of the number of sensing channels. The framework considered in this paper is also sufficiently general to include other reward and cost functions for specific applications.

The models considered in this paper, though analytically tractable, have limitations. The Markovian traffic assumption may not be sufficiently accurate, and more general traffic models are preferred. We have not considered formally the presence of sensing error except that we have used simulation to demonstrate the robustness of the optimal PS-OSA. To this end, the modeling considered in [5] and the ideas presented in [9] are most relevant. The presence of more than two secondary users is not treated in this paper, which requires the modeling of contention. There are also practical protocol issues of synchronization and the estimation and tracking of the traffic parameters. These are topics for further investigation.

VIII. Appendix

Before we present the proofs of results, let us introduce a notation $\varphi(i,k)$ which will be used frequently below. Define $\varphi(i,k)$ as the slot index where channel i was last sensed before the k-th slot. As a convention, if channel i is sensed at kT_s, we assume $\varphi(i,k)=k$. With this notation, $Z_i(k)=X_i(\varphi(i,k)T_s)$. It is clear that $\tau(i,k)+\varphi(i,k)=k$ since $\tau(i,k)$ is the number of time slots passed at time kT_s since the last sensing was made in channel i. So we can determine $\varphi(i,k)$ as $\varphi(i,k)=k-\tau(i,k \bmod N)=k-[(N+q-i) \bmod N]$.
Proof of Theorem 3.1: Note that process $Z(k)$ starts at time N. Thus we need to prove [18]

$$\Pr(\mathbf{Z}(k+1)|\mathbf{Z}(k),\ldots,\mathbf{Z}(N))=\Pr(\mathbf{Z}(k+1)|\mathbf{Z}(k)). \quad (23)$$

In fact, for $k+1$, with $k\geq N-1$, sensing channel is $q'=k+1 \bmod N$. Note, our process starts from time $k=N$. Since only this channel's state is updated, $\mathbf{Z}(k+1)$ should be different from $\mathbf{Z}(k)$ in only the q'-th component. The q'-th component of $\mathbf{Z}(k+1)$ is $X_{q'}((k+1)T_s)$. Thus, we have

$$\begin{aligned}&\Pr(\mathbf{Z}(k+1)|\mathbf{Z}(k),\ldots,\mathbf{Z}(N))\\&\quad=\Pr(X_{q'}((k+1)T_s)|\mathbf{Z}(k),\ldots,\mathbf{Z}(N)),\end{aligned}$$

if for all $i\neq q'$, $Z_i(k+1)=Z_i(k)$;

$$\Pr(\mathbf{Z}(k+1)|\mathbf{Z}(k),\ldots,\mathbf{Z}(N))=0, \quad (24)$$

otherwise. Due to the independency of channels, we have

$$\begin{aligned}&\Pr(\mathbf{Z}(k+1)|\mathbf{Z}(k),\ldots,\mathbf{Z}(N))\\&=\Pr(X_{q'}((k+1)T_s)|X_{q'}(\varphi(i,k)T_s),\ldots,X_{q'}(\varphi(i,N)T_s)),\end{aligned}$$

if $\forall i\neq q', Z_i(k+1)=Z_i(k)$. Recall that $\tau(i,k)(\leq N-1)$ is the number of slots passed at time kT_s since the last observation in channel i. Furthermore, since every channel is Markovian, we have

$$\begin{aligned}&\Pr(X_i((k+1)T_s)|X_i(\varphi(i,k)T_s),\ldots,X_i(\varphi(i,N)T_s))\\&\quad=\Pr(X_i((k+1)T_s)|X_i(\varphi(i,k)T_s)),\forall i.\end{aligned}$$

This implies that Eq. (23) holds.

The above discussion also enables us to reduce the determination of the transition probability

$$\Pr(\mathbf{Z}(k+1)=\mathbf{z}'|\mathbf{Z}(k)=\mathbf{z})$$

to the determination of

$$\Pr(X_{q'}(\varphi(q',k+1)T_s)=z'_{q'}|X_{q'}((\varphi(',k))T_s)=z_{q'}).$$

This turns out can be done since for continuous Markov Chains $X_i(t)$ with parameters λ_i (idle) and μ_i (busy), we can obtain expressions of $\Pr(X_i(t_2)=z'_i|X_i(t_1)=z_i)$ for all $t_2\geq t_1$.

Let the transition rate matrix for each channel be $\mathbf{Q}_i$, then we have

$$\mathbf{Q}_i = \begin{pmatrix} -\lambda_i & \lambda_i \\ \mu_i & -\mu_i \end{pmatrix} \tag{25}$$

The matrix exponential $\exp(\mathbf{Q}_i t)$ evaluates to be

$$\begin{aligned}&\exp(\mathbf{Q}_i t)\\ &= \begin{bmatrix} 1-\frac{\lambda_i}{\lambda_i+\mu_i}(1-\exp(-(\lambda_i+\mu_i)t)) & \frac{\lambda_i}{\lambda_i+\mu_i}(1-\exp(-(\lambda_i+\mu_i)t)) \\ \frac{\mu_i}{\lambda_i+\mu_i}(1-\exp(-(\lambda_i+\mu_i)t)) & 1-\frac{\mu_i}{\lambda_i+\mu_i}(1-\exp(-(\lambda_i+\mu_i)t)) \end{bmatrix}\end{aligned} \tag{26}$$

Then for $z_i', z_i \in \{0,1\}$, we have

$$\begin{aligned}&\Pr(X_i(\varphi(i,k+1)T_s) = z_i' | X_i(\varphi(i,k)T_s) = z_i)\\ &= [\exp(\mathbf{Q}_i(\varphi(i,k+1) - \varphi(i,k))T_s)]_{(z_i,z_i')}.\end{aligned}$$

For the special case $i = q'$, channel q' is sensed at slot I_{k+1}, thus $\varphi(q',k+1) = k+1$. Furthermore, since we are carrying out periodic sensing with period length being N slots, $\varphi(q',k) = k+1-N$. Thus, we have

$$\varphi(q',k+1) - \varphi(q',k) = N.$$

As a result, we can establish that

$$\begin{aligned}&\Pr(X_{q'}(\varphi(q',k+1)T_s) = z_i' | X_{q'}(\varphi(q',k)T_s) = z_i)\\ &= [\exp(\mathbf{Q}_{q'} N T_s)]_{(z_i,z_i')}.\end{aligned}$$

The proof is completed. □

Proof of Theorem 3.2: The steady-state probabilities of the observations generated by periodic sensing $\mathbf{Z}(pN+q)$, $p = 1,2,\ldots$, for any $q \in \{0,1,\ldots,N-1\}$ are given by

$$f_q(\mathbf{z}) = \lim_{P\to\infty} \frac{1}{P} f_q^P(\mathbf{z}), \tag{27}$$

where $f_q^P(\mathbf{z})$ represents the number of times $\mathbf{z}$ appears in the sequence $\{\mathbf{Z}(pN+q), p = 1,\ldots,P\}$.

The existence of (27) is guaranteed for all q and $\mathbf{z} \in \mathcal{B}^N$ since Markov chains $\mathbf{Z}(pN+q)$, $p = 1,2,\ldots$, are irreducible and aperiodic. In fact, we have transition probability

$$\begin{aligned}&\Pr(\mathbf{Z}(pN+q+N) = \mathbf{z}' | \mathbf{Z}(pN+q) = \mathbf{z})\\ &= \prod_{i=0}^{N-1} \Pr(X_i(N+\varphi(i,pN+q)T_s) = z_i' | X_i(\varphi(i,pN+q)T_s) = z_i)\\ &= \prod_{i=0}^{N-1} [\exp(\mathbf{Q}_i N T_s)]_{(z_i,z_i')},\end{aligned}$$

where the second equality is due to the periodicity of indices $\tau(i,k)$, the first equality is due to independent of the primary user processes X_i. This implies $\Pr(\mathbf{Z}(pN+q+N) = \mathbf{z}' | \mathbf{Z}(pN+q) = \mathbf{z}) > 0$ for all pairs of vectors $\mathbf{z}', \mathbf{z} \in \mathcal{B}^N$. In terms of chain structure, $\mathbf{Z}(pN+q)$, this means that all states are immediately reachable from each state, thus the chain is irreducible and aperiodic. Furthermore, based the transition probability expression, we can derive the stationary distribution f_q in product form as

$$f_q(\mathbf{z}) = \prod_{i=0}^{N-1} \left(1_{[z_i=0]} v_i(0) + 1_{[z_i=1]} v_i(1)\right), \tag{28}$$

where $1_{[\cdot]}$ denotes the indicator function and

$$v_i(0) = \frac{\mu_i}{\lambda_i+\mu_i}, \quad v_i(1) = \frac{\lambda_i}{\lambda_i+\mu_i}. \tag{29}$$

In fact, it is not hard to show that f_q is an invariant distribution of the sequences $\mathbf{Z}(pN+q)$, $p = 1,2,\ldots$, for all $q \in \{0,1,\ldots,N-1\}$. It is interesting to note from Eq. (28) that the stationary distributions $f_q(\mathbf{z})$ are identical for all q. This is intuitive: the processes of all primary user channels are stationary, as a result the distribution of the observation made by the secondary user should not depend on the specific time in a period. □

Proof of Theorem 4.1: Observe that

$$Z_i(k) = X_i(\varphi(i,k)T_s),$$

an analytical expression for the reward is derived as follows

$$\begin{aligned}&r(\mathbf{z},a,k)\\ &= \Pr(X_i(t) = 0, \forall t \in I_k | X_i(\varphi(i,k)T_s) = z_i))\\ &= \Pr(X_i(t) = 0, \forall t \in I_k | X_i(kT_s) = 0, X_i(\varphi(i,k)T_s) = z_i)\\ &\quad \cdot \Pr(X_i(kT_s) = 0 | X_i(\varphi(i,k)T_s) = z_i)\\ &= \Pr(X_i(t) = 0, \forall t \in I_k | X_i(kT_s) = 0)\\ &\quad \cdot \Pr(X_i(kT_s) = 0 | X_i(\varphi(i,k)T_s) = z_i)\\ &= \exp(-\lambda_i T_s) \left[\exp(\mathbf{Q}_i \tau(i,k) T_s)\right]_{(z_i,0)}\\ &= \exp(-\lambda_i T_s) \left[\exp(\mathbf{Q}_i \tau(i,q) T_s)\right]_{(z_i,0)}\end{aligned} \tag{30}$$

where recall that the subscript notation $[\exp(\mathbf{Q}_i t)]_{(z_i,0)}$ indicates the transition probability form state z_i to 0 in channel i (over time t) and $q = k \bmod N$.

If we introduce a table g indexed by $\mathbf{z}, q, i$,

$$g(\mathbf{z},q,i) = \exp(-\lambda_i T_s) \left[\exp(\mathbf{Q}_i \tau(i,q) T_s)\right]_{(z_i,0)}, \tag{31}$$

then based on Equation (26), we have

$$\begin{aligned}g(\mathbf{z},q,i) = \exp(-\lambda_i T_s)\Big(&\tfrac{\mu_i}{\lambda_i+\mu_i} + \big(1_{[z_i=0]}\tfrac{\lambda_i}{\lambda_i+\mu_i} - 1_{[z_i=1]}\tfrac{\mu_i}{\lambda_i+\mu_i}\big)\\ &\cdot \exp(-(\lambda_i+\mu_i)\tau(i,q)T_s)\Big)\end{aligned} \tag{32}$$

for $i = \{0,\ldots,N-1\}$, the immediate reward and cost in k-th slot can be analytically evaluated by

$$r(\mathbf{z},a,k) = \begin{cases} g(\mathbf{z}, k \bmod N, a), & a \geq 0 \\ 0, & a = -1 \end{cases} \tag{33}$$

and

$$c(\mathbf{z},a,k) = \begin{cases} 1 - g(\mathbf{z}, k \bmod N, a), & a \geq 0 \\ 0, & a = -1 \end{cases} \tag{34}$$

□

Proof of Theorem 4.2: The proof is based on the application of the existing CMDP theory [16]. Compared with the standard CMDP formulation, our model in Eq. (14,15) has two major differences. One difference is that the reward function $r(\mathbf{z},a,k)$ and the cost function is $c(\mathbf{z},a,k)$ are periodic instead of constant for a given state and action pair $(\mathbf{z},a)$. However, if we extend state vector to include the position in a period, $q = k \bmod N$, we will obtain a recurrent Markov chain with time invariant reward and cost. The other difference is that our constraints are not in form of a time average. This difference is superficial in the sense that we can view $C_i(\pi)$ as

$$\lim_{K\to\infty} \frac{\sum_{k=1}^{K} \mathrm{E}c(\mathbf{Z}(k),i,k)\Pr(A_k = i|\pi,\mathbf{Z}(k),k)/K}{\mathrm{E}B_i(K)/K}$$

and note that the limit $\lim_{K\to\infty} \mathrm{E}B_i(K)/K$ always exist ($= 1 - v_i(0)\exp(-\lambda_i T_s)$). So, if we redefine the state as $(\mathbf{z},q)$

and the constants in the right hand of the constants as $\gamma_i(1 - v_i(0)\exp(-\lambda_i T_s))$, we can convert our CMDP problem to the standard form of CMDP formulation in [16]. According to CMDP theory, when the state and action space are both finite, for unichain (including recurrent) chains, the optimal value is always achieved at some "stationary" randomized policy. Here stationary is in terms of the extended state space which means periodicity in the original state space.

First we show the optimal throughput of our CMDP is no greater than that of the optimal value of the LP. Let us consider a fixed optimal periodic policy π^* of the CMDP. If we classify transmissions according to the position q in a period, the objective function in Eq.(14) can then be written in form of

$$J(\pi^*) = \lim_{P\to\infty} \frac{\sum_{p=1}^{P}\sum_{q=0}^{N-1} \mathrm{E}_{\pi^*} r(\mathbf{Z}(pN+q), A_{pN+q}, pN+q)}{PN+q}.$$

Denote $\beta_{q,a}^{\pi^*,P}(\mathbf{z}) \in [0,1]$ as the frequency of action $a \in \mathcal{A}$ chosen by π^* in slot I_{pN+q} when the observed value of $\mathbf{Z}(pN+q)$ equals to $\mathbf{z} \in \mathcal{B}^N$ in a sample path with $p = 1, 2, \ldots, P$. In other words,

$$\beta_{q,a}^{\pi^*,P}(\mathbf{z}) = P^{-1}\sum_{p=1}^{P} \Pr(A_{pN+q} = a | \mathbf{Z}(pN+q) = \mathbf{z}, \pi^*). \tag{35}$$

According to CMDP theory, for the chosen policy, the frequency $\lim_{P\to\infty}\beta_{q,a}^{\pi^*,P}(\mathbf{z})$ of the state-action pair $(\mathbf{z}, q, a)$ exists. Let us denote collectively these frequencies as

$$\beta^{\pi^*} = (\beta_{q,a}^{\pi^*}(\mathbf{z}), a \in \mathcal{A}, q \in \{0, 1, \ldots, N-1\}, \mathbf{z} \in \mathcal{B}^N)).$$

Given a position q in one sensing period N, under policy π^*, for the process $\mathbf{Z}(pN+q)$, $p = 1, 2, \ldots, P$, the expected total number of successful transmission

$$\sum_{p=1}^{P} \mathrm{E}_{\pi^*} r(\mathbf{Z}(pN+q), A_{pN+q}, pN+q)$$

equals to

$$\begin{aligned} &\sum_{p=1}^{P}\sum_{\mathbf{z}\in\mathcal{B}^N}\sum_{i=0}^{N-1} g(\mathbf{z}, q, i)\cdot \Pr(\mathbf{Z}(pN+q) = \mathbf{z}) \\ &\qquad\qquad \cdot \Pr(A_{pN+q} = i | \mathbf{Z}(pN+q) = \mathbf{z}, \pi^*) \\ &= \sum_{\mathbf{z}\in\mathcal{B}^N}\sum_{i=0}^{N-1} g(\mathbf{z}, q, i)\cdot \sum_{p=1}^{P}\Pr(\mathbf{Z}(pN+q) = \mathbf{z}) \\ &\qquad\qquad \cdot \Pr(A_{pN+q} = i | \mathbf{Z}(pN+q) = \mathbf{z}, \pi^*). \end{aligned} \tag{36}$$

Since the sensing results on primary users are not affected by the transmission policy of the secondary user. Assume the processes of primary users are in stationary states at the beginning, that is, $X_i(0)$ has the distribution

$$v_i = (v_i(0), v_i(1)),$$

where

$$v_i(0) = \frac{\mu_i}{\lambda_i + \mu_i}$$

and

$$v_i(1) = \frac{\lambda_i}{\lambda_i + \mu_i}.$$

Then we have

$$v_i \mathbf{Q}_i = v_i$$

and

$$v_i \exp(\mathbf{Q}_i t) = v_i.$$

As a result,

$$\Pr(X_i(t) = 0) = v_i(0)$$

and

$$\Pr(X_i(t) = 1) = v_i(1)$$

for all $t \geq 0$. Especially, we have

$$\Pr(X_i(\varphi(i,k)T_s) = 0) = v_i(0)$$

and

$$\Pr(X_i(\varphi(i,k)T_s) = 1) = v_i(1)$$

for all $k \geq N$. Since the processes of primary user channels are independent, for any given $\mathbf{z} \in \mathcal{B}^N$, we have

$$\begin{aligned} &\Pr(X_i(\varphi(i,k)T_s) = z_i, \forall i = 0, 1, \ldots, N-1) \\ &= \textstyle\prod_{i=0}^{N-1}\Pr(X_i(\varphi(i,k)T_s) = z_i). \end{aligned}$$

It then follows from Eq. (28) and

$$\Pr(\mathbf{Z}(pN+q) = \mathbf{z}) = \Pr(X_i(\varphi(i, pN+q)T_s) = z_i, \forall i)$$

that

$$\Pr(Z(pN+q) = \mathbf{z}) = f_q(\mathbf{z})$$

for all $p = 1, 2\ldots$. Furthermore we establish from Eq.(35) that Eq.(36) can be rewritten as

$$\sum_{\mathbf{z}\in\mathcal{B}^N}\sum_{i=0}^{N-1} g(\mathbf{z}, q, i)\cdot f_q(\mathbf{z})\cdot \beta_{q,i}^{\pi*,P}(z)\cdot P \tag{37}$$

As a result, the asymptotical transmission rate under policy π^* at the position q in a period is given by

$$\begin{aligned} &\lim_{P\to\infty}\textstyle\sum_{p=1}^{P}\mathrm{E}_{\pi*} r(\mathbf{Z}(pN+q), A_{pN+q}, pN+q)/P \\ &= \textstyle\sum_{\mathbf{z}\in\mathcal{B}^N}\sum_{i=0}^{N-1} g(\mathbf{z}, q, i)\cdot f_q(\mathbf{z})\cdot \beta_{q,i}^{\pi^*}(\mathbf{z}). \end{aligned}$$

Thus sum over q, we have

$$\begin{aligned} &J(\pi^*) \\ &= N^{-1}\textstyle\sum_{q=0}^{N-1}\sum_{\mathbf{z}\in\mathcal{B}^N}\sum_{i=0}^{N-1} g(\mathbf{z}, q, i)\cdot f_q(\mathbf{z})\beta_{q,i}^{\pi^*}(\mathbf{z}). \end{aligned}$$

Similarly to previous derivations, the constraints on secondary user's interference to primary users in individual channel can be converted to the following inequalities

$$\begin{aligned} &N^{-1}\textstyle\sum_{q=0}^{N-1}\sum_{\mathbf{z}\in\mathcal{B}^N} f_q(\mathbf{z})(1 - g(\mathbf{z}, q, i))\beta_{q,i}(\mathbf{z}) \\ &\leq \gamma_i(1 - v_i(0)\exp(-\lambda_i T_s)), \quad \forall i, \end{aligned}$$

where $v_i(0) = \frac{\mu_i}{\lambda_i + \mu_i}$. In fact, for policy π^*, we have

$$\begin{aligned} &C_i(\pi^*) \\ &= \lim_{K\to\infty}\tfrac{\sum_{k=1}^{K}\mathrm{E}c(\mathbf{Z}(k), i, k)\Pr(A_k = i|\pi^*, \mathbf{Z}(k), k)}{\mathrm{E}B_i(K)} \\ &= \lim_{K\to\infty}\tfrac{K^{-1}\sum_{k=1}^{K}\mathrm{E}c(\mathbf{Z}(k), i, k)\Pr(A_k = i|\pi^*, \mathbf{Z}(k), k)}{\mathrm{E}B_i(K)/K} \\ &= \tfrac{N^{-1}\sum_{q=0}^{N-1}\sum_{\mathbf{z}\in\mathcal{B}^N} f_q(\mathbf{z})(1 - g(\mathbf{z}, q, i))\beta_{q,i}^{\pi^*}(\mathbf{z})}{1 - v_i(0)\exp(-\lambda_i T_s)}. \end{aligned}$$

Now put everything together, we have verified that β^{π^*} is a feasible solution to our linear programming problem defined in Eqs. (16,18).

Second, we will show that the optimal value of the linear programming problem is no greater than that of the CMDP.

It is sufficient to show that any optimal solution to the LP is feasible to the CMDP. In fact, given an optimal solution $\beta = (\beta_{q,a}(\mathbf{z}), a \in \mathcal{A}, q \in \{0, 1, \ldots, N-1\}, \mathbf{z} \in \mathcal{B}^N)$ to the LP, the secondary user need only do the following to establish a feasible solution $\pi(\beta)$ to the CMDP. Store β as a table. Given the observations $\mathbf{z}$ and position q in a period, the secondary user's policy is simply to flip a biased coin such that with probability $\beta_{k \bmod N, i}(\mathbf{Z}(k))$ we transmit in channel i and with probability $\beta_{k \bmod N, -1}(\mathbf{Z}(k))$ no transmission occurs. Let us call this random policy $\pi(\beta)$. It is straightforward to verify that the policy $\pi(\beta)$ satisfying

$$J(\pi(\beta)) = N^{-1} \sum_{q=0}^{N-1} \sum_{\mathbf{z} \in \mathcal{B}^N} \sum_{i=0}^{N-1} g(\mathbf{z}, q, i) f_q(\mathbf{z}) \beta_{q,i}^{\pi(\beta)}(\mathbf{z})$$

and

$$C_i(\pi(\beta)) = \sum_{q=0}^{N-1} \sum_{\mathbf{z} \in \mathcal{B}^N} \frac{f_q(\mathbf{z})(1 - g(\mathbf{z}, q, i))\beta_{q,i}(\mathbf{z})}{N(1 - v_i(0)\exp(-\lambda_i T_s))}$$

since the frequency of the state-action pair $(\mathbf{z}, q, a)$ of this policy is exactly $\beta_{q,a}(\mathbf{z})$. It then follows from the feasibility of β i.e., Eq.(17), that

$$C_i(\pi(\beta)) \leq \gamma_i \tag{38}$$

which means that Eq. (15) holds. The proof is completed. □

REFERENCES

[1] Q. Zhao and B.M. Sadler, "A survey of dynamic spectrum access: signal processing, networking, and regulatory policy," *IEEE Signal Processing Magazine*, May 2007.

[2] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, volume 50, Issue 13, Pages 2127-2159, Sept., 2006.

[3] *Proc. First IEEE Internatial Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005.

[4] S. Geirhofer, L. Tong, and B. M. Sadler, “”Dynamic spectrum access in WLAN channels: empirical model and its stochastic analysis," in *Proc. First International Workshop on Technology and Policy in Accessing Spectrum (TAPAS)*, Boston, MA, August, 2006.

[5] A. Leu, M. McHenry, and B. Mark, "Modeling and analysis of interference in Listen-Before-Talk spectrum access schemes," *Intl. J. Network Mgmt*, vol. 16, pp. 131-147, 2006.

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

[7] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," to appear in *IEEE Journal on Selected Areas in Communications: Special Issue on Adaptive, Spectrum Agile and Cognitive Wireles Networks,* vol. 25, no. 3, pp. 589-600, April, 2007.

[8] Q. Zhao and A. Swami, "A decision-theoretic framework for opportunistic spectrum access," to appear in *IEEE Wireless Communications Magazine: Special Issue on Cognitive Wireless Netowrks*, 2007.

[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*, November, 2006.

[10] S. Geirhofer, L. Tong, and B. M. Sadler, "A measurement-based model for dynamic spectrum access," in *Proc. IEEE Conference on Military Communications (MILCOM)*, 2006.

[11] E. Sondik, "The optimal control of partially observable markov processes over the infinite horizon: discounted costs," *Operations Research*, Vol. 26, No. 2 (Mar. - Apr., 1978), pp. 282-304

[12] H. Yu, *Approximation solution methods for partially observable Markov and semi-Markov decision processes*, Ph.D. Dissertation, MIT, 2006.

[13] S. Geirhofer, L. Tong, and B. M. Sadler,, "Dynamic spectrum access in the time domain: Modeling and exploiting whitespace," *IEEE Communications Magazine*, Vol. 45, No. 2, May 2007.

[14] S. Jones, N. Merheb, and I.-J. Wang, "An experiment for sensingbased opportunistic spectrum access in csma/ca networks," in *First IEEE Internatial Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2005, pp. 593–596.

[15] E. Altman, *Constrained Markov Decision Processes*. Chapman & Hall/CRC, 1999.

[16] M. L. Puterman, *Markov Decision Processes.Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.

[17] Q. Zhao, S. Geirhofer, L. Tong, and B. Sadler, "Periodic Sensing Opportunistic Spectrum Access," Technical Report, ACSP-12-06-01, December, 2006.

[18] E. Çinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Inc., Englewood Cliffs, New Jersy, 1975.