How to Optimally Schedule Cooperative Spectrum Sensing in Cognitive Radio Networks

Ke Lang, Yuan Wu, and Danny H.K. Tsang

Department of Electronic and Computer Engineering
Hong Kong University of Science and Technology
langke@ust.hk, ecewuy@ust.hk, eetsang@ece.ust.hk

Abstract. In cognitive radio (CR) networks, secondary users can be coordinated to perform spectrum sensing so as to detect primary user activities more accurately. However, in a dynamic spectrum environment, more sensing cooperations may induce every secondary user to sense more channels, thus decreasing their transmission time. In this paper, we study this tradeoff by using the theory of partially observable Markov decision process (POMDP). This formulation leads to an optimal sensing scheduling policy that determines which secondary users sense which channels with what miss detection probability and false alarm probability. A myopic policy with lower complexity yet comparable performance is also proposed. Numerical and simulation results are provided to illustrate that our design can utilize the spectrum more efficiently for cognitive radio users.

Keywords: cognitive radio, cooperative sensing scheduling, partially observable Markov decision process.

1 Introduction

Cognitive radio (CR) [1] [2] is a new technology that provides a novel solution to the spectrum inefficiency problem. In CR networks, there are two types of users: primary users and secondary users. A primary user (PU) is a licensed owner of a channel, while a secondary user (SU) periodically scans the PU spectrum, identifies the idle channels and accesses the channels opportunistically without causing intolerable interference to PUs.

IEEE 802.22 wireless regional area networks (WRANs) standard [3] is aimed at using CR techniques to allow sharing of the unused spectrum. The 802.22 system uses base stations (BSs) to manage their cells and all associated SUs. In addition to the traditional role of a BS, the 802.22 BS has a new function of coordinating cooperative sensing [3] to make sensing results more accurate.

Many research works were carried out to analyze the performance of cooperative sensing [4] [5] [6]. Although cooperative sensing can improve the spectrum sensing accuracy, it also has some drawbacks, especially when the number of SUs

X. Jun Hei and L. Cheung (Eds.): AccessNets 2009, LNICST 37, pp. 133–148, 2010.

is limited and only sequential (narrowband) sensing is allowed[1]. A typical example can be given by that in cooperative sensing some SUs may be scheduled to sense several channels sequentially, which in turn will decrease their transmission time significantly. Therefore, there is a tradeoff between cooperative sensing time and transmission time. Meanwhile, the idle spectrum available for SUs to access is time-changing, and the information about the changing idle spectrum can only be partially observed by SUs (due to both the imperfect spectrum sensing and sensing scheduling policy, which we will describe in detail in sections 2 and 3). Based on these considerations, in this paper we study the dynamic scheduling for cooperative sensing under time-varying spectrum environment. Specifically, we formulate our dynamic sensing scheduling problem with the partially observable Markov decision process (POMDP), and derive an optimal sensing scheduling policy (i.e. determining which SUs to sense which set of channels with what miss detection probability and false alarm probability) via solving the formulated problem. To the best of our knowledge, our work is the first one in this direction, and shed light on how to implement cooperative spectrum sensing function proposed in 802.22 standard.

POMDP was used in [7] [8] to study the dynamic spectrum access for an Ad hoc CR networks. An incremental pruning algorithm was proposed in [10] which could solve the POMDP problem. In our paper, we adopt this algorithm to get the optimal sensing scheduling policy of a multi-user cooperative spectrum sensing problem in a centralized manner.

The rest of the paper is organized as follows. We present the network model and propose our protocol in section 2. We then formulate the problem of the tradeoff between cooperative sensing time and transmission time as a POMDP in section 3. We derive the optimal policy and myopic policy for our problem in section 4. Section 5 presents numerical and simulation results. Finally, we conclude this paper in section 6.

2 System Model

2.1 Network Model

In this work, we consider a centralized CR network with a base station (or access point), which manages the cooperative sensing scheduling as well as data transmission. All SUs in a cell need to be synchronized. In the following part, we further assume there exists a set of SUs $\mathcal{M} = \{1, 2, ..., M\}$, and a set of orthogonal frequency channels $\mathcal{N} = \{1, 2, ..., N\}$ with a BS in a cell.

Each SU is equipped with a single radio interface. In this work, we assume all SUs use energy detection mechanism for spectrum sensing and each SU can

[1] According to [11], the wideband spectrum sensing refers to that the sensing device can sense multiple spectrum bands over a wide frequency range at a time. Meanwhile, the sequential spectrum sensing refers to that the sensing device can only sense one spectrum band at a time, and thus different spectrum bands have to be sensed sequentially.

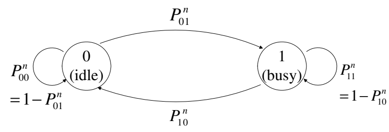

Fig. 1. DTMC model for PU channel n

only carry out the sequential spectrum sensing instead of the wideband spectrum sensing due to some PHY layer limitations.

2.2 Opportunistic Channel Availability Model

In this paper, we assume primary system operates in a time slotted manner with fixed slot length T. In the PU network, each channel's occupancy (from slot to slot) follows a two-state discrete time Markov chain (DTMC) as shown in Figure 1. Let $s_n(t)$ denote the availability state of channel n ($n \in \mathcal{N}$) in time slot t. $s_n(t) = 0$ denotes channel n is idle (available for SU to use) in slot t, while $s_n(t) = 1$ denotes channel n is busy (not available for SU to use) in slot t. Furthermore, let the $1 \times N$ vector $\mathbf{s}(t) = (s_1(t), ..., s_N(t))$ denote the channel availability state vector for all the PU channels in slot t, which has the state space $\Omega^{\mathbf{s}} = \{(\omega_1, \omega_2, ..., \omega_N) | \omega_n = \{0, 1\}, \forall n \in \mathcal{N}\}$. By assuming independence across different channels, the dynamics of $\mathbf{s}(t)$ follow a DTMC with transition probability from state vector ω to state vector ω' given as:

$$P_{\omega\omega'} = \Pr(\mathbf{s}(t+1) = \omega' | \mathbf{s}(t) = \omega) = \prod_{n=1}^{N} P_{\omega_n \omega'_n}^n, \forall \omega, \omega' \in \Omega^s \tag{1}$$

where ω_n, ω'_n denote the nth element of state vector ω and ω', respectively. $P_{\omega_n \omega'_n}^n$ has been shown in Figure 1, representing channel n's state transition probability. We consider the DTMC model as time homogeneous, i.e. $P_{01}^n, P_{10}^n, \forall n \in \mathcal{N}$, are time independent. We assume the PU channels' statistical behavior $P_{01}^n, P_{10}^n, \forall n \in \mathcal{N}$, can be obtained from a long term measurement by some channel parameter estimator [12], and this information is provided to CR BS. Note that for each channel n, the stationary probabilities of being idle and busy $\pi_0^n, \pi_1^n, \forall n \in \mathcal{N}$ can be calculated as $\pi_0^n = \frac{P_{10}^n}{P_{01}^n + P_{10}^n}$, and $\pi_1^n = \frac{P_{01}^n}{P_{01}^n + P_{10}^n}$. In section 5, we use different values of π_0^n to show the impact of PUs' activities on the SUs' performance.

2.3 Spectrum Sensing Technique and Cooperative Detection

Several well-known spectrum sensing techniques have been proposed including matched filter detection, energy detection, cyclostationary feature detection and wavelet detection [1] [5]. In this paper, we adopt the energy detection method [13]. The received signal $x_R(t)$ takes the form

$$x_R(t) = \begin{cases} e(t), & H_0(t) \\ h \cdot x_T(t) + e(t), & H_1(t) \end{cases}$$

where $x_R(t)$ is the received signal at a secondary user on a channel, $x_T(t)$ is the transmitted signal of the primary user, $e(t)$ is the additive white Gaussian noise, h is the channel gain of the sensing channel between the primary user and the secondary user. Hypothesis 0 (H_0) corresponds to no signal transmitted, hypothesis 1 (H_1) corresponds to signal transmitted. Then, in a non-fading environment, the detection probability P_D and false alarm probability P_{FA} are given as follows [13],

$$P_D = \Pr(Y > \lambda | H_1) = Q_u(\sqrt{2\gamma}, \sqrt{\lambda}) \tag{2}$$

and

$$P_{FA} = \Pr(Y > \lambda | H_0) = \Gamma(u, \frac{\lambda}{2})/\Gamma(u) \tag{3}$$

where Y is the test or decision statistic, λ is the decision threshold, u is the time bandwidth product, γ is the SNR, $Q_u(\cdot,\cdot)$ is the generalized Marcum Q-function, $\Gamma(\cdot)$ and $\Gamma(\cdot,\cdot)$ are the complete and incomplete gamma functions. Then, the miss detection probability is $P_{MD} = 1 - P_D$. For a fading environment, P_{MD} and P_{FA} have more complicated expressions [5] [13].

We adopt a simple cooperative sensing scheme, which is called "OR" rule [5] in this paper. This rule works like this: every SU sends its sensing result (0 or 1) of a channel to the BS, and as long as one SU senses the channel as busy, the BS will take this channel as busy. Only if all SUs sense the channel as idle, BS will take the channel as idle. In this case, the miss-detection probability and false alarm probability of channel n are

$$P_{MD}(n) = \prod_{m \in \mathcal{M}(n)} P_{MD}(m, n)$$

$$P_{FA}(n) = 1 - \prod_{m \in \mathcal{M}(n)} (1 - P_{FA}(m, n))$$

where $P_{MD}(m, n)$ and $P_{FA}(m, n)$ are $SU\ m's$ miss detection probability and false alarm probability of channel n, and $\mathcal{M}(n)$ is the set of SUs sensing this channel.

If we further assume that each SU which is scheduled to sense channel n uses the same miss detection probability and false alarm probability, then we have

$$P_{MD}(m, n) = \sqrt[|\mathcal{M}(n)|]{P_{MD}(n)}$$

where $|\mathcal{M}(n)|$ denotes cardinality of set $\mathcal{M}(n)$. This means if we set $P_{MD}(n)$ as a constant target value, and we have $|\mathcal{M}(n)|$ SUs to cooperatively sense channel n instead of having only one SU to sense it, then the tolerable miss detection probability for each of the $|\mathcal{M}(n)|$ SUs $P_{MD}(m, n)$ will increase to $\sqrt[|\mathcal{M}(n)|]{P_{MD}(n)}$.

A larger $P_{MD}(m,n)$ means the decision threshold λ in (2) becomes larger[2]. Then from (3), we will have a smaller $P_{FA}(m,n)$. We can further see that as $P_{MD}(n)$ is set as a constant target value, $P_{FA}(n)$ will become smaller if $|\mathcal{M}(n)|$ SUs cooperatively sense channel n instead of no cooperation. The larger the number of cooperative SUs, the smaller $P_{FA}(n)$ we will get, this means the sensing accuracy becomes better when more SUs sense the channel [5].

2.4 Proposed Protocol

Figure 2 shows an example to illustrate the operation process of the BS and SUs in CR network using our proposed protocol. At the beginning of each slot, each channel will have a state transition according to the DTMC model described in subsection 2.2, and at the same time, the BS decides which SU senses which set of channels with what probabilities of miss detection and false alarm based on our optimal policy which will be presented later. For example, in Figure 2, BS decides SU1 to sense channels $1, 3, 4$, and SU2 to sense channels $1, 2, 5$. After receiving the decisions from the BS, SU1 and SU2 will sequentially sense the assigned channels, and the channel sensing sequence can be arbitrarily determined. Since the sensing duration for each channel is a fixed value ΔL, and we have limited number of SUs, if the BS decides some channels are sensed by more SUs in order to increase these channels' sensing accuracy, then each SU may need to sense more channels accordingly, thus causing less time for transmission (Notice that in Figure 2, the slot length is L. The time duration for BS scheduling cooperative sensing is η_1. The time duration for SUs uploading their sensing results and BS allocating channels to SUs is η_2. All these three values are constant). There is a tradeoff between cooperative sensing time and transmission time.

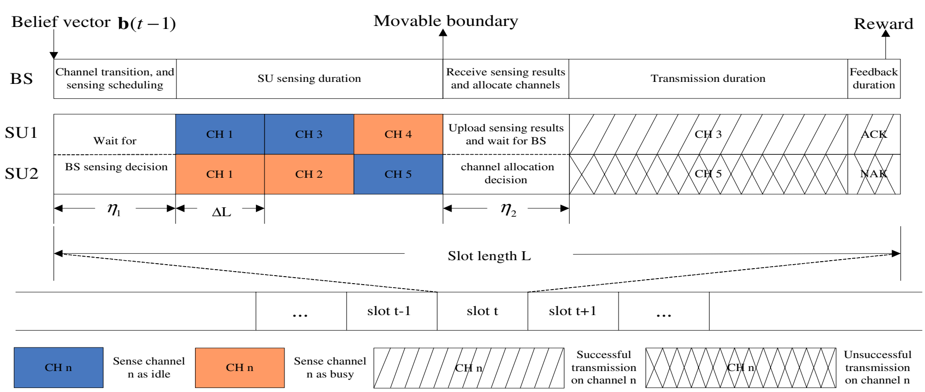

Fig. 2. An example of the operation process of our proposed protocol

[2] In our work both the values of receiver SNR γ in a non-fading environment or average receiver SNR in a fading environment, and time bandwidth product u are fixed.

Since we adopt in our protocol the cooperative sensing scheme using "OR" rule, although SU1 senses channel 1 as idle, and SU2 senses it as busy, BS determines channel 1 is busy and not to use it in the example of Figure 2.

In our proposed protocol, data transmission works as follows: if a channel's sensing result is idle, then although there exists the probability of miss detection, the BS will allocate this channel to one of SUs. As seen from Figure 2, channel 5 turns out to be busy when SU2 tries to use it. Our protocol requires sensing synchronization for all SUs, i.e., each SU senses the same number of channels.

At the end of a slot, the SU using the channel will send an ACK or NAK to the BS, which will be used as an important information for future decisions. In this paper, we only consider the case of downlink transmission, but our proposed protocol can also be applied to the uplink transmission.

3 Problem Formulation

At the beginning of each time slot, based on previous actions and observations, the BS could have a belief state over every channel, which is the probability of a PU channel being in that state in the previous time slot. This is different from traditional Markov decision process, because in our case, the BS may not know the exact state of a channel. For instance, if BS determines the channel as busy, it can not be sure if it is busy due to the probability of false alarm. Besides, if some channels are not sensed by any SU, the exact state of these channels will not be known either.

Before we present our complete problem formulation, we first describe our optimal cooperative sensing scheduling problem from the perspective of POMDP.

(i) Action

In our formulation, at the beginning of time slot t there are two actions, a^I and a^{II}. a^I determines which SU senses which channels. a^{II} determines how to tune the sensor operating point of each SU (i.e. miss detection probability and false alarm probability) when sensing a channel.

$$a^I(t) = \begin{bmatrix} a^1_{11}(t) & ... & a^1_{1N}(t) \\ a^1_{21}(t) & ... & a^1_{2N}(t) \\ ... & ... & ... \\ a^1_{M1}(t) & ... & a^1_{MN}(t) \end{bmatrix}$$

where $a^1_{mn}(t) \in \{0, 1\}$, $a^1_{mn}(t) = 1$ denotes SU m senses channel n in time slot t, and $a^1_{mn}(t) = 0$ means the opposite. We define the set of SUs that are scheduled to sense channel n in slot t as $\mathcal{M}(n, t) = \{m|a^1_{mn}(t) = 1\}$.

$$a^{II}(t) = \begin{bmatrix} a^2_{11}(t) & ... & a^2_{1N}(t) \\ a^2_{21}(t) & ... & a^2_{2N}(t) \\ ... & ... & ... \\ a^2_{M1}(t) & ... & a^2_{MN}(t) \end{bmatrix}$$

where $a^2_{mn}(t) = P_{MD}(m, n, t)$ denotes the specified miss detection probability for SU m on channel n in time slot t. Notice that by setting the value of miss detection probability $P_{MD}(m,n,t)$, we actually determine the sensor operating point for SU m on channel n in slot t, because both the sensing decision threshold λ and the false alarm probability $P_{FA}(m,n,t)$ for SU m on channel n in slot t can be calculated via (2) and (3), respectively. Specifically, in this work we choose the value of miss detection probability $P_{MD}(m,n,t)$ from a set of discrete values, which is practical for most spectrum sensing modules' operation because the sensing operation point cannot be tuned continuously.

We define $\mathbf{a}(t)= [a^I(t), a^{II}(t)]$.

(ii) Observation

Let $\theta_n(t)$ denote the observation result of channel n in time slot t. Then $\theta_n(t)$ could have the following 4 possible observations,

- BS determines the channel as idle and receives ACK after transmission; denote this as observation 0.
- BS determines the channel as idle and receives NAK after transmission due to miss detection; denote this as observation 1.
- BS determines the channel as busy, does not use the channel, and thus the BS receives no ACK or NAK; denote this as observation 2.
- BS decides not to sense the channel and thus observes nothing; denote this as observation 3.

Further, let the $1 \times N$ vector $\theta(t) = (\theta_1(t), ...\theta_N(t))$ denote the channel observation vector for all the PU channels at the end of slot t, which has the observation space $\mathbf{Z}^\theta = \{(z_1, z_2, ..., z_N)|z_n = \{0, 1, 2, 3\}, \forall n \in \mathcal{N}\}$.

The individual channel observation probability is defined as the probability of the observation given the action we take and the current state of channel n, *i.e.*

$$\Pr(\theta_n(t)|\mathbf{a}(t), s_n(t)) = \begin{cases} 1, \quad if \ \sum_{m=1}^{M} a^1_{mn}(t) = 0,\ \theta_n(t) = 3 \\ 1 - P_{FA}(n,t), \\ \quad if \ \sum_{m=1}^{M} a^1_{mn}(t) > 0,\ s_n(t) = 0,\ \theta_n(t) = 0 \\ P_{FA}(n,t), \\ \quad if \ \sum_{m=1}^{M} a^1_{mn}(t) > 0,\ s_n(t) = 0,\ \theta_n(t) = 2 \\ P_{MD}(n,t), \\ \quad if \ \sum_{m=1}^{M} a^1_{mn}(t) > 0,\ s_n(t) = 1,\ \theta_n(t) = 1 \\ 1 - P_{MD}(n,t), \\ \quad if \ \sum_{m=1}^{M} a^1_{mn}(t) > 0,\ s_n(t) = 1,\ \theta_n(t) = 2 \\ 0, \quad otherwise \end{cases}$$

where $P_{MD}(n,t)$ is the miss detection probability of channel n in time slot t. $P_{FA}(n,t)$ is the false alarm probability of channel n in time slot t. Because of the "OR" rule in cooperative sensing, we have

$$P_{MD}(n,t) = \prod_{m \in \mathcal{M}(n,t)} P_{MD}(m,n,t),\ P_{FA}(n,t) = 1 - \prod_{m \in \mathcal{M}(n,t)} (1 - P_{FA}(m,n,t))$$

In the above observation probability function, if $\sum_{m=1}^{M} a^1_{mn}(t) = 0$, which means SUs do not sense channel n, then no matter what the current state is, we will have observation 3 with probability 1, and we will not have any other observations.

If $\sum_{m=1}^{M} a_{mn}^1(t) > 0,\ s_n(t) = 0$, which means when SUs sense channel n and the current state is idle, we will have observation 0 when no false alarm happens, or we will have observation 2 when false alarm happens.

If $\sum_{m=1}^{M} a_{mn}^1(t) > 0,\ s_n(t) = 1$, which means when SUs sense channel n and the current state is busy, we will have observation 1 when miss detection happens, or we will have observation 2 when no miss detection happens.

Then, the observation probability is

$$\Pr(\theta(t) = \mathbf{z}|\mathbf{a}(t), \mathbf{s}(t) = \omega) = \prod_{n=1}^{N} \Pr(\theta_n(t) = z_n|\mathbf{a}(t), s_n(t) = \omega_n) \tag{4}$$

where $\mathbf{z} \in \mathbf{Z}^\theta$ is the observation vector, and z_n denotes the nth element of observation vector $\mathbf{z}$.

For the sake of simplicity, we assume every SU being scheduled to sense channel n should tune to the same sensor operating point (i.e. $P_{MD}(m, n, t) = \sqrt[|\mathcal{M}(n,t)|]{P_{MD}(n, t)}, \forall m \in \mathcal{M}(n, t)$).

(iii) Belief vector

Because of the partial spectrum sensing decisions and the presence of sensing errors, a BS may not observe the true system state. However, the BS can infer the system state based on all its past decisions and observations, and summarize this information into a belief vector [8], $\mathbf{b}(t) \triangleq \{b_\omega(t)\}_{\omega\in\Omega^{\mathbf{s}}}$[3] where $b_\omega(t) \triangleq \Pr(\mathbf{s}(t) = \omega|\mathbf{b}(0), \{\mathbf{a}(\tau), \theta(\tau)\}_{\tau=1}^{t}) \in [0, 1]$ is the conditional probability (given all past decisions and observations) that the system state is ω in the current time slot t. $b_\omega(t)$ can only be computed at the end of the current time slot t when $\theta(t)$ is known (as shown in Figure 2). The BS will make actions at slot $t + 1$ based on its belief vector of the system state $\mathbf{b}(t)$.

We define the updated belief vector as follows:

$$\mathbf{b}(t) \triangleq \mathcal{T}(\mathbf{b}(t-1), \mathbf{a}(t), \theta(t)) \triangleq \{b_{\omega'}(t)\}_{\omega'\in\Omega^{\mathbf{s}}} \tag{5}$$

where $\mathcal{T}(\mathbf{b}(t-1), \mathbf{a}(t), \theta(t))$ represents the updated knowledge of the network state after incorporating the action and observation obtained in slot t. Then, from Bayes rule, we have

$$\begin{aligned} b_{\omega'}(t) &= \Pr(\mathbf{s}(t) = \omega'|\mathbf{b}(t-1), \mathbf{a}(t), \theta(t)) \\ &= \frac{\sum\limits_{\omega\in\Omega^{\mathbf{s}}} b_\omega(t-1)P_{\omega\omega'} \Pr(\theta(t)|\mathbf{a}(t), \mathbf{s}(t) = \omega')}{\sum\limits_{\omega\in\Omega^{\mathbf{s}}} \sum\limits_{\omega''\in\Omega^{\mathbf{s}}} b_\omega(t-1)P_{\omega\omega''} \Pr(\theta(t)|\mathbf{a}(t), \mathbf{s}(t) = \omega'')} \end{aligned} \tag{6}$$

From these equations, we know that we have regained the Markov property for the belief state in that the next belief state depends only on the previous belief state, the current action and the current observation received.

[3] Here we abuse the notation a little since we just want to list all the elements in the set $\Omega^{\mathbf{s}}$ and assign them to the vector $\mathbf{b}(t)$, and the element order is not important.

To illustrate this concept clearly, we take a simple example. Suppose we focus on one channel (e.g. channel n), according to the above equation, we have the following belief updating rule.

If we take the action of sensing channel n, we may have observation $0,\ 1$, or 2.

When we have observation 0, the BS receives an ACK, so it knows the channel state is idle, and according to (6), $b_0(t) = 1$.

When we have observation 1, the BS receives a NAK, so it knows the channel state is busy, and according to (6), $b_0(t) = 0$.

However, for the case of observation 2, the situation becomes partially observable, because we do not know the exact channel state. When we have observation 2, due to the existence of false alarm probability, although the BS determines the channel as busy, it may be actually idle. Then according to (6), $b_0(t) = \{b_0(t-1)\cdot P_{00}^n\cdot P_{FA}(n,t)+b_1(t-1)\cdot P_{10}^n\cdot P_{FA}(n,t)\}/\{b_0(t-1)\cdot P_{00}^n\cdot P_{FA}(n,t)+b_1(t-1)\cdot P_{10}^n\cdot P_{FA}(n,t)+b_0(t-1)\cdot P_{01}^n\cdot(1-P_{MD}(n,t))+b_1(t-1)\cdot P_{11}^n\cdot(1-P_{MD}(n,t))\}$. If false alarm probability becomes 0, we can see that $b_0(t)$ becomes 0.

Similarly, if we take the action of not sensing channel n, we will have observation 3, since we do nothing on the channel in this time slot, we do not know the exact channel state, its belief state just updates according to the DTMC model, and according to (6), $b_0(t) = b_0(t-1)\cdot P_{00}^n + b_1(t-1)\cdot P_{10}^n$.

Here, $b_0(t)$ is the belief of the state being idle in slot t, and the belief of the state being busy is $b_1(t) = 1 - b_0(t)$.

(iv) Reward function

There will be a reward when the channel is sensed and finally the BS receives an ACK, the immediate reward for channel n ($n \in \mathcal{N}$) is

$$R_n(\mathbf{a}(t), \theta_n(t)) = \begin{cases} \frac{L-k-\eta}{L}, & if\ \sum_{m=1}^{M} a_{mn}^1(t) > 0, \theta_n(t) = 0 \\ 0, & otherwise \end{cases} \tag{7}$$

where $k = \Delta L \cdot \sum_{n=1}^{N} a_{mn}^1(t),\ \forall m \in \mathcal{M}$, is the sensing duration (note that to keep synchronization, each SU should sense same number of channels, then $\sum_{n=1}^{N} a_{1n}^1(t) = \sum_{n=1}^{N} a_{2n}^1(t) = ... = \sum_{n=1}^{N} a_{Mn}^1(t)$), ΔL is the sensing duration for one channel, and $\eta = \eta_1 + \eta_2$ is a constant time used for BS decisions and getting sensing results from SUs. Then, the immediate reward for all the channels in time slot t is

$$R(\mathbf{a}(t), \theta(t)) = \sum_{n=1}^{N} R_n(\mathbf{a}(t), \theta_n(t)) \tag{8}$$

Finally, the expected reward for BS to make a decision at the beginning of slot t is

$$\begin{aligned} &\widetilde{R}(\mathbf{a}(t), \mathbf{s}(t-1) = \omega) \\ &= \sum_{\omega' \in \Omega^{\mathbf{s}}} \sum_{\mathbf{z} \in \mathbf{Z}^{\theta}} P_{\omega\omega'} \Pr(\theta(t) = \mathbf{z} | \mathbf{a}(t), \mathbf{s}(t) = \omega') R(\mathbf{a}(t), \theta(t) = \mathbf{z}) \end{aligned} \tag{9}$$

(v) Complete problem formulation
We aim to develop the optimal policy that can maximize the expected total throughput of the SUs over a finite time horizon T, and at the same time it must satisfy the synchronization constraint and primary user interference constraint. The problem is formulated as follows:

$$\max \ E\{\sum_{t=1}^{T} R(\mathbf{a}(t), \theta(t))|\mathbf{b}(0) = \mathbf{b}\} \tag{10}$$

$$\text{subject to: } \sum_{n=1}^{N} a_{1n}^{1}(t) = \sum_{n=1}^{N} a_{2n}^{1}(t) = ... = \sum_{n=1}^{N} a_{Mn}^{1}(t) \tag{11}$$

$$L - \Delta L \cdot \sum_{n=1}^{N} a_{mn}^{1}(t) - \eta > 0, \ \forall m \in \mathcal{M} \tag{12}$$

$$P_c(n,t) \leq \zeta, \ \forall n \in \mathcal{N} \tag{13}$$

In the above formulation, $\mathbf{b}$ is the initial belief vector which could be set according to the channels' statistical behavior. Constraint (11) is a synchronization constraint, which guarantees that each SU senses the same number of channels. Constraint (12) guarantees the transmission time is positive. Constraint (13) is the interference constraint, which aims to satisfy primary users' interference tolerance. Here, $P_c(n,t)$ is the collision probability of channel n in slot t, if we want to guarantee this value below the prescribed primary channel collision probability ζ, it is equivalent to require the miss detection probability of channel n in slot t below this threshold ζ,

$$P_{MD}(n,t) \leq \zeta, \ \forall n \in \mathcal{N} \tag{14}$$

Then the formulation with objective (10) constrained by (11), (12), and (13) changes into the formulation with objective (10) constrained by (11), (12), and (14).

4 Optimal Policy and Myopic Policy

4.1 Optimal Policy

In order to solve the objective function (10), we could solve the following value function $V_t(\mathbf{b}(t-1))$ which denotes the maximum expected remaining reward that can be obtained from the beginning of slot t when the current belief vector is $\mathbf{b}(t-1)$. We use backward induction method to calculate the value function from two parts. One part is the expected immediate reward $\widetilde{R}(\mathbf{a}(t), \mathbf{s}(t-1) = \omega)$ in the current time slot, and the other part is the expected future reward $V_{t+1}(\mathcal{T}(\mathbf{b}(t-1), \mathbf{a}(t), \theta(t) = \mathbf{z}))$. The optimal policy finds a balance between gaining immediate reward and gaining future reward.

(i) When $t = 1, 2, ...T - 1$,

$$V_t(\mathbf{b}(t-1)) = \max_{\mathbf{a}(t)}\{\sum_{\omega \in \Omega^{\mathbf{s}}} b_\omega(t-1)[\widetilde{R}(\mathbf{a}(t), \mathbf{s}(t-1) = \omega) +$$
$$\sum_{\omega' \in \Omega^{\mathbf{s}}} P_{\omega\omega'} \sum_{\mathbf{z} \in \mathbf{Z}^\theta} \Pr(\theta(t) = \mathbf{z}|\mathbf{a}(t), \mathbf{s}(t) = \omega') V_{t+1}(\mathcal{T}(\mathbf{b}(t-1), \mathbf{a}(t), \theta(t) = \mathbf{z}))]\}$$

subject to: (11), (12), (14)

(ii) When $t = T$,

$$V_t(\mathbf{b}(t-1)) = \max_{\mathbf{a}(t)} \sum_{\omega \in \Omega^{\mathbf{s}}} b_\omega(t-1)\widetilde{R}(\mathbf{a}(t), \mathbf{s}(t-1) = \omega)$$

subject to: (11), (12), (14)

where $\widetilde{R}(\mathbf{a}(t), \mathbf{s}(t-1) = \omega)$ is given by (9), $P_{\omega\omega'}$ is given by (1), $\Pr(\theta(t) = \mathbf{z}|\mathbf{a}(t), \mathbf{s}(t) = \omega')$ is given by (4), and $\mathcal{T}(\mathbf{b}(t-1), \mathbf{a}(t), \theta(t) = \mathbf{z})$ is given by (5) and (6).

Therefore, the optimal policy could be obtained as follows:
(i) When $t = 1, 2, ...T - 1$,

$$\mathbf{a}^*(t) = \arg\max_{\mathbf{a}(t)}\{\sum_{\omega \in \Omega^{\mathbf{s}}} b_\omega(t-1)[\widetilde{R}(\mathbf{a}(t), \mathbf{s}(t-1) = \omega) +$$
$$\sum_{\omega' \in \Omega^{\mathbf{s}}} P_{\omega\omega'} \sum_{\mathbf{z} \in \mathbf{Z}^\theta} \Pr(\theta(t) = \mathbf{z}|\mathbf{a}(t), \mathbf{s}(t) = \omega') V_{t+1}(\mathcal{T}(\mathbf{b}(t-1), \mathbf{a}(t), \theta(t) = \mathbf{z}))]\} \quad (15)$$

subject to: (11), (12), (14)

(ii) When $t = T$,

$$\mathbf{a}^*(t) = \arg\max_{\mathbf{a}(t)} b_\omega(t-1)\widetilde{R}(\mathbf{a}(t), \mathbf{s}(t-1) = \omega) \quad (16)$$

subject to: (11), (12), (14)

We use the incremental pruning algorithm to solve (15) and (16). Detailed algorithm and its complexity analysis could be referred to [10].

4.2 Myopic Policy

Although the optimal scheduling policy for cooperative spectrum sensing can be derived via (15) and (16), the required computation complexity grows tremendously with the numbers of SUs and channels, even using the incremental pruning algorithm. Moreover, the optimal scheduling policy requires maintaining a table that specifies the optimal actions in every time slot. Therefore, the table could become very large as the time horizon increases. One solution for this computational complexity problem is the divide and conquer approach. For example, we separate all SUs into two smaller SU groups and also separate all channels into two smaller channel groups. Then, we can carry out two POMDP algorithms each consisting of one group of SUs and one group of channels.

To further address the computational complexity problem, in this work we also consider a myopic scheduling policy for cooperative spectrum sensing, which can be expressed as follows:

$$\mathbf{a}^*(t) = \arg\max_{\mathbf{a}(t)} \sum_{\omega \in \Omega^{\mathbf{s}}} b_\omega(t-1)\widetilde{R}(\mathbf{a}(t), \mathbf{s}(t-1) = \omega)$$

subject to: (11), (12), (14)

Essentially, in our myopic policy, SU BS aims to maximize its instantaneous expected reward in each time slot t. (Notice that SU BS also uses (5) and (6) to update its belief state $b_\omega(t)$). Our following simulation results show that the myopic scheduling policy can achieve a comparable performance as that of the optimal policy based on POMDP formulation.

5 Numerical and Simulation Results

We set the CR network with a set of SUs $\mathcal{M} = \{1, 2\}$, and a set of orthogonal frequency channels $\mathcal{N} = \{1, 2\}$. Both of the channels have the same unit bandwidth. We set $\Delta L = 0.2L$ as the sensing duration for each channel. We also set $\eta = 0.1L$ as the duration for BS's sensing decision, receiving sensing results from SUs, and channel allocation decision. We assume Rayleigh fading channels with the same average receiver SNR=10dB [5] [13] and we use the same time bandwidth product $u = 5$. In Figure 3, the time horizon is 15 slots, and in the other simulations the time horizon is 10 slots. In these simulations we set every channel as homogeneous for different SUs. In this case there may exist several optimal solutions for (15) and (16), the BS will just pick one of them randomly.

To compare with the optimal policy and myopic policy, we consider a simple random policy which randomly picks an action as long as it can satisfy all the constraints (11), (12), (14).

Figure 3 shows the throughput comparison of the theoretical results of optimal policy, the simulation results of myopic policy and random policy. We set the scenario that both channels have the same statistical behavior (i.e. $P_{00}^n = 0.8, P_{10}^n = 0.2,\ n = 1, 2$), and the same prescribed collision probability $\zeta = 0.1$. This figure shows the advantages of the optimal policy and myopic policy over the random one with time horizon increasing. It also shows the optimal policy and myopic policy have very similar throughput performance in this scenario.

In Figure 4, we also set the scenario that both channels have the same statistical behavior (i.e. $P_{00}^n = 0.8, P_{10}^n = 0.2,\ n = 1, 2$), and the same prescribed collision probability ζ from 0.05 to 0.3. This figure shows that with the increase of the prescribed collision probability ζ, SUs' throughput becomes larger because of PUs' more collision tolerance. Nevertheless, when the prescribed collision probability reaches some level, SUs' throughput will stop increasing, this is because

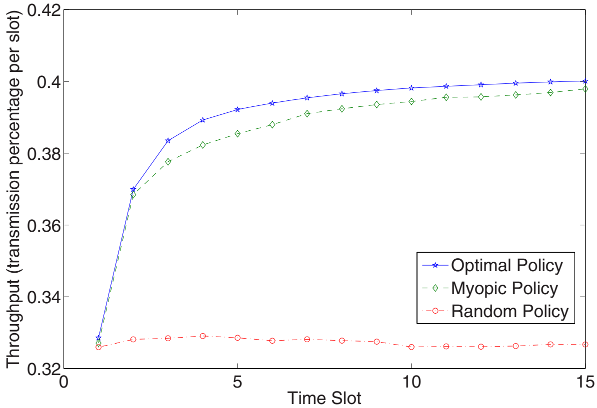

Fig. 3. SUs throughput performance comparison with $P_{00}^1 = P_{00}^2 = 0.8, P_{10}^1 = P_{10}^2 = 0.2$, and the same $\zeta = 0.1$

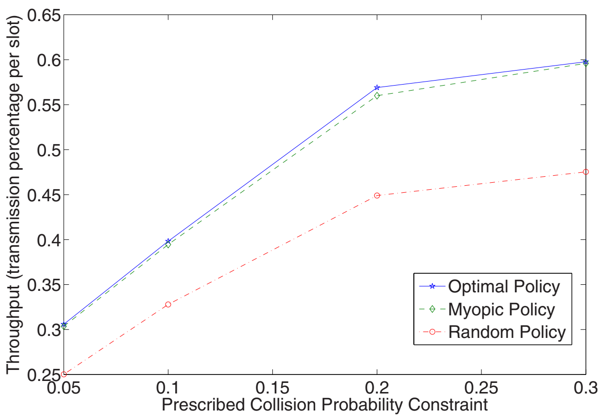

Fig. 4. SUs throughput performance comparison with $P_{00}^1 = P_{00}^2 = 0.8, P_{10}^1 = P_{10}^2 = 0.2$

it has already arrived at the maximum point of the primary channels' unutilized opportunity.

In Figure 5, we study the SUs' throughput performance under different memories of PU channel transition process. According to [9], the memory of channel n's transition process is defined as $\mu_n = 1 - P_{01}^n - P_{10}^n,\ n \in \mathcal{N}$, which is the probability of remaining in the same channel state. In this paper, we set $\mu_n > 0,\ n \in \mathcal{N}$, which means all the channels have positive transition process memories. The larger the memory, the higher tendency a channel will remain in the same state. We also consider the case of both channels having the same

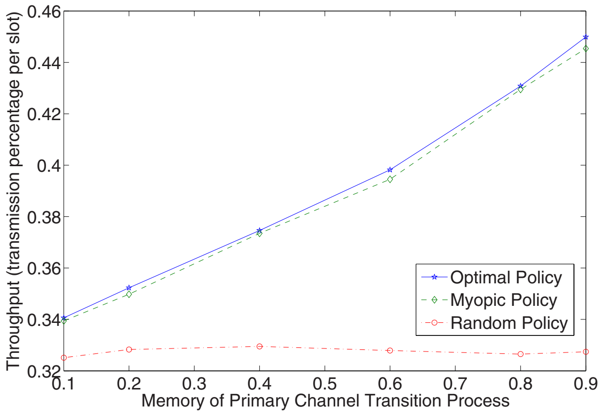

Fig. 5. SUs throughput performance comparison with $\pi_0^1 = \pi_0^2 = 0.5$, and the same $\zeta = 0.1$

statistical behavior (i.e. $P_{00}^1 = P_{00}^2, P_{10}^1 = P_{10}^2$), the same stationary idle probability (i.e. $\pi_0^n = 0.5,\ n = 1, 2$), and the same prescribed collision probability $\zeta = 0.1$. Figure 5 shows that when the channels' transition process memories grow larger, the throughput performance of optimal policy and myopic policy grow much better than the random policy. This indicates that if all the channels have positive transition process memories, then the larger the memories, the better throughput performance we can get by using our optimal and myopic policies. In fact, if some channels have negative transition process memories, then the larger the absolute value of the channel transition process memories, the better throughput performance we can get by using our optimal and myopic policies.

In Figure 6, we study the SUs' throughput performance when the two channels' statistical behaviors become different. Here, we set the prescribed collision probability ζ as 0.1 for each channel, and we set the sum of the two channels' transition process memories as a constant ($i.e.$ $\mu_1 + \mu_2 = 1.2$). Besides, their stationary idle probabilities are the same ($i.e.$ $\pi_0^1 = \pi_0^2 = 0.5$). This figure shows that although their stationary idle probabilities are the same and the sum of the two channels' transition process memories does not change, using the optimal policy can obtain a better throughput performance than the myopic policy when the diversity of the two channels' transition process memories (i.e. $|\mu_1 - \mu_2|$) grows larger. This is because that when the two channels' statistical behaviors are similar, the myopic policy will be similar to the optimal policy. However, when the two channels' statistical behaviors become more and more different, the myopic policy will get different decisions with the optimal policy.

Figure 7 shows the SUs' throughput performance under different stationary idle probability of PU channels. Here we set the two PU channels with the same statistical behavior (i.e. $P_{00}^1 = P_{00}^2, P_{10}^1 = P_{10}^2$) and the same prescribed collision

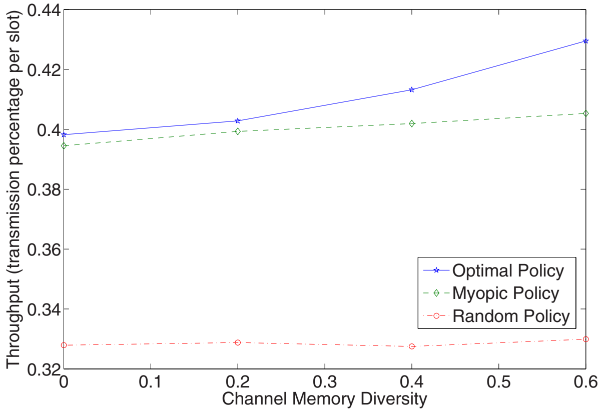

Fig. 6. SUs throughput performance comparison with $\pi_0^1 = \pi_0^2 = 0.5$, and the same $\zeta = 0.1$, and $\mu_1 + \mu_2 = 1.2$

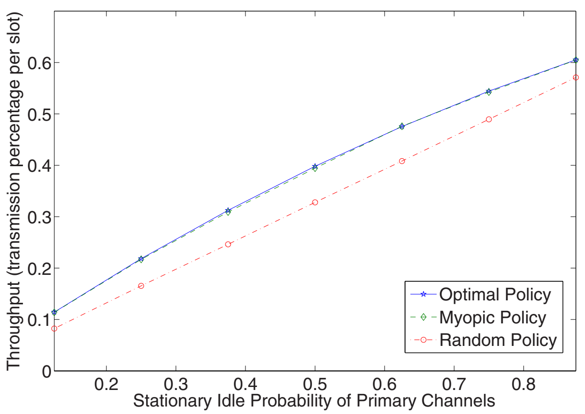

Fig. 7. SUs throughput performance comparison with the same $\zeta = 0.1$, and $\mu_1 = \mu_2 = 0.6$

probability $\zeta = 0.1$, and then we change their stationary idle probability while maintaining their channel transition process memories ($i.e.$ $\mu_1 = \mu_2 = 0.6$). It is shown from the figure that when PU channels' stationary idle probability increases, SUs' throughput increases accordingly. This is because SUs will get more opportunities as the PU channels' idle probability increases.

6 Conclusion

In this paper, we study the cooperative sensing scheduling problem in cognitive radio networks. We first formulate this problem as a POMDP which aims to maximize the total CR system throughput with the guarantee of primary users' prescribed collision probability. Then, we derive the optimal policy and a myopic policy that determines which SUs sense which channels with what miss detection probability and false alarm probability. Numerical and simulation results are provided to illustrate the throughput performance of our optimal and myopic scheduling policies for cooperative spectrum sensing.

References

1. Akyildiz, I.F., Lee, W.-Y., Vuran, M.C., Mohanty, S.: Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. Comput. Netw. Int. J. Comput. Telecommun. Netw. 50(13), 2127–2159 (2006)
2. Haykin, S.: Cognitive radio: Brain-empowered wireless communications. IEEE J. Sel. Areas Commun. 23(2), 201–220 (2005)
3. Cordeiro, C., Challapali, K., Birru, D., Shankar, S.: IEEE 802. 22: The First Worldwide Wireless Standard based on Cognitive Radios. J. Commun. 1(1), 60–67 (2006)
4. Mishra, S.M., Sahai, A., Brodersen, R.W.: Cooperative Sensing among Cognitive Radios. In: Proc. IEEE Int. Conf. Commun., Turkey, vol. 4, pp. 1658–1663 (2006)
5. Letaief, K.B., Zhang, W.: Cooperative Spectrum Sensing. In: Cognitive Wireless Communication Networks, pp. 115–138. Springer, Heidelberg (2007)
6. Zhang, W., Letaief, K.B.: Cooperative Spectrum Sensing with Transmit and Relay Diversity in Cognitive Radio Networks. IEEE Trans. Wireless Commun. 7(12), 4761–4766 (2008)
7. Zhao, Q., Tong, L., Swami, A., Chen, Y.: Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework. IEEE J. Sel. Areas Commun. 25(3) (2007)
8. Chen, Y., Zhao, Q., Swami, A.: Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors. IEEE Trans. Inf. Theory. 54(5), 2053–2071 (2008)
9. Hoang, A.T., Liang, Y.C., Wong, D.T.C., Zeng, Y.H., Zhang, R.: Opportunistic Spectrum Access for Energy-Constrained Cognitive Radios. IEEE Trans. Wireless Commun. 8(3) (2009)
10. Cassandra, A., Littman, M.L., Zhang, N.L.: Incremental Pruning: A simple, Fast, Exact Method for Partially Observable Markov Decision Processes. In: Proc. 13th Conf. Uncertainty in Artificial Intelligence (UAI), Providence RI, pp. 54–61 (1997)
11. Lee, W.-Y., Akyildiz, I.F.: Optimal Spectrum Sensing Framework for Cognitive Radio Networks. IEEE Trans. Wireless Commun. 7(10) (2008)
12. Kim, H., Shin, K.G.: Efficient Discovery of Spectrum Opportunities with MAC-Layer Sensing in Cognitive Radio Networks. IEEE, Trans. Mobile Computing 7(5) (2008)
13. Digham, F.F., Alouini, M.-S., Simon, M.K.: On the Energy Detection of Unknown Signals over Fading Channels. In: Proc. IEEE ICC 2003, vol. 5, pp. 3575–3579 (2003)