Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2010, Article ID 305145, 12 pages
doi:10.1155/2010/305145

Research Article

Optimal Policy of Cross-Layer Design for Channel Access and Transmission Rate Adaptation in Cognitive Radio Networks

Hao He, Jun Wang, Jiang Zhu, and Shaoqian Li

National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu, Sichuan Province 611731, China

Correspondence should be addressed to Hao He, hh@uestc.edu.cn

Received 29 April 2009; Revised 14 September 2009; Accepted 2 December 2009

Academic Editor: Rui Zhang

In this paper, we investigate the cross-layer design of joint channel access and transmission rate adaptation in CR networks with multiple channels for both centralized and decentralized cases. Our target is to maximize the throughput of CR network under transmission power constraint by taking spectrum sensing errors into account. In centralized case, this problem is formulated as a special constrained Markov decision process (CMDP), which can be solved by standard linear programming (LP) method. As the complexity of finding the optimal policy by LP increases exponentially with the size of action space and state space, we further apply action set reduction and state aggregation to reduce the complexity without loss of optimality. Meanwhile, for the convenience of implementation, we also consider the pure policy design and analyze the corresponding characteristics. In decentralized case, where only local information is available and there is no coordination among the CR users, we prove the existence of the constrained Nash equilibrium and obtain the optimal decentralized policy. Finally, in the case that the traffic load parameters of the licensed users are unknown for the CR users, we propose two methods to estimate the parameters for two different cases. Numerical results validate the theoretic analysis.

1. Introduction

In recent years, the explosive growth of wireless devices and traffic incurs a dramatic increase of the requirement for radio spectrum resource. Unfortunately, as most of the spectrum resource suitable for wireless communications has been assigned, the available spectrum resources become scare. As today's spectrum is managed under a fixed assignment policy, which is highly inefficient in terms of spectrum utilization [1, 2], cognitive radios are adopted to sense their environments and promptly reconfigure their communication parameters based on their observations [3–5].

In CR networks, a new spectrum access method, namely dynamic spectrum access (DSA), is employed to improve spectrum utilization by allowing CR users to access the idle licensed spectrum bands without colliding with the active licensed users [6]. In multi-user environment, the DSA design should also consider the collision with other CR users. Meanwhile, CR users should take the power consumption into account. Furthermore, the time-varying fading nature of radio channel complicates adaptive access and transmission techniques. The solution to above problems asks for cross-layer design between physical layer and upper layers.

Recently, the issue of cross-layer design for dynamic spectrum access has attracted many researchers' efforts. Zhao et al. present a decentralized cognitive medium access control protocol under the framework of partially observable Markov decision process for ad hoc network [7]. This work is then extended by [8] to maximize the expected number of information bits delivered by an unlicensed user before its total energy is exhausted. Kim and chin [9] propose a MAC-layer sensing framework and energy-efficient dynamic sensing mode selection algorithm. The design of spectrum sensing and access strategies in the presence of spectrum sensing errors has been addressed in [10–13]. On the other hand, spectrum opportunity sharing among CR users is discussed in [14–17]. In [18, 19], the authors consider the power control in CDMA system under power constraint by formulating a stochastic game model to solve the problem. However, their assumptions of the transmission reward

and the channel state in these works are not practical for fading channels. Besides, they do not consider multichannel case. In fact, to the best of our knowledge, there is little work focusing on the cross-layer design under the power constraint by taking the time-varying characteristics of the channel state and the collisions (both the collisions with primary user and the collisions with other CR users) into account. In this paper, we consider cross-layer design of multichannel access and transmission rate adaptation in CR network for both centralized and decentralized cases by taking the time-varying characteristics of channel state into account.

We consider the coexistence of a CR network with a licensed wideband wireless communications network. In centralized case, the cross-layer design problem can be modeled by constrained Markov decision process (CMDP) and solved by a dynamic programming method to achieve the optimal performance. In decentralized case, each CR user only knows its local information and should take actions to maximize the total performance of the whole CR network. We prove the existence of the constrained Nash equilibrium and calculate the optimal decentralized policy.

Another key difference between our approach and that of the above references is that complexity reduction is explicitly taken into account in our method. In both centralized and decentralized cases, the complexity of finding optimal policy increases exponentially with the size of action space and state space, which incurs the so-called curse of dimensionality. To overcome this problem, we perform action set reduction and state aggregation to reduce the complexity without loss of optimality. Under certain condition, we further prove that the multichannel access and transmission rate adaptation policy design can be solved separately in each channel without loss of optimality.

Furthermore, we observe that pure policy is preferred in practical environment due to the convenience of implementation and evaluation. Pure policies take action with deterministic rule. We name all the stationary polices as mixed polices and analyze the difference between pure polices and mixed polices in our proposed cross-layer design. This issue has attracted little attention in exiting researches.

In CR network, the change of white space or spectrum hole utilized for unlicensed communication depends on the spectrum occupancy of licensed users. But the CR users do not know the traffic load parameters of licensed users generally. In this case, we proposed two methods to estimate the traffic load parameters of licensed user.

The remaining part of the paper is organized as follows. In Section 2, the system model is described. The cross-layer design problem is formulated and discussed for both centralized and decentralized cases in Section 3. The complexity reduction of optimal policy design is considered in Section 4, in which we discuss the action set reduction and state aggregation and prove that the multichannel access and transmission rate adaptation policy design can be solved separately in each channel without loss of optimality under certain condition. Section 5 investigates the optimal pure policy design. In Section 6, the estimation methods for the unknown traffic load parameters of licensed user are provided. The numerical result is presented and discussed in Section 7. Finally, we conclude our work and point out the future work in Section 8.

2. System Model

In this paper, we consider the coexistence of a CR network with a licensed wideband wireless communications network. We refer to the wideband channel as licensed channel in this paper. The CR network consists of N CR users and a base station. The wideband channel of the wideband network is divided into M narrowband channels (or subchannels, subcarriers) that are utilized by the N CR users for opportunistic uplink packet transmission. Each narrowband channel can be used by only one CR user in each frame.

The transmission model for every CR user is shown in Figure 1. At the data link layer, for the transmission capacity analysis, the infinite buffer of the transmitter is assumed to be continually backlogged with packets that must be transmitted to the base station and the channel selection is decided by every CR user. At the physical layer, the CR users operate over the selected parallel block fading channels and send data to the CR base station. These channels compose one (or a part of) licensed channel. To maximize the spectral efficiency, adaptive modulation (AM) [20] is utilized for each selected channel. In centralized case, the base station makes the whole decision. In decentralized case, the intelligent controller in each CR user performs cross-layer channel selection and rate selection in frame-by-frame manner. Furthermore, the intelligent controller should include some extra function blocks. First, the controller should calculate the immediate reward and cost for the optimal policy design. Second, in order to reduce complexity, the function block of action set reduction and state aggregation should be included in the intelligent controller. Finally, in the case that the traffic load parameters of licensed users are unknown, the controller should estimate them.

The frame structure is depicted in Figure 2. In Figure 2, the time-axis is divided into contiguous slots of equal duration, which correspond to frames with fixed-length of T_f. Channel sensing time is T_s. For notational convenience and without loss of generality, it is assumed that T_s is fairly small. For CR user i, the probability of sensing false alarm and sensing miss detection are defined as P_{fa}^{i} and P_{md}^{i}, respectively. We assumed that the channel availabilities for the whole CR network are the same. Once the sensing result is idle, the CR user can transmit pilot to the base station to obtain the channel state information (CSI). It is assumed that the CSI could be fed back correctly without delay if the sensing result is correct, otherwise the collision could occur and the CSI cannot be fed back. Finally, if the CSI is fed back correctly, the CR users can take packets from the buffer, map them into symbols and select a transmission rate to send them to base station in both centralized and decentralized cases. At the end of each frame, the base station acknowledges every successful or unsuccessful transmission by error-free ACK or NAK, respectively.

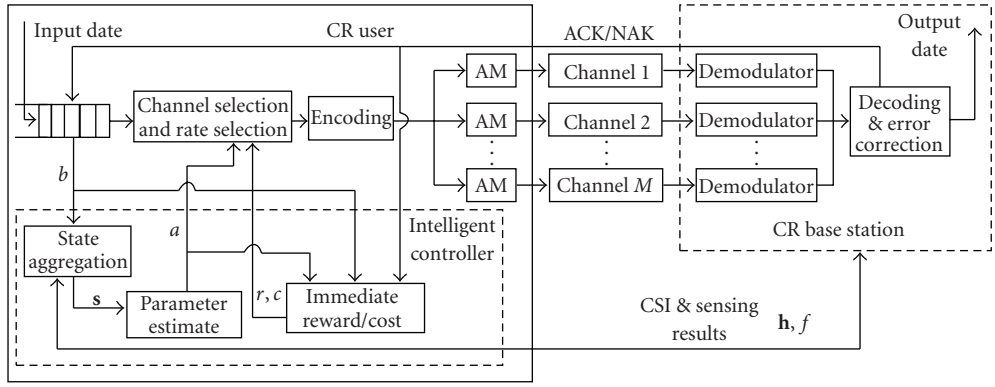

Figure 1: Transmission model.

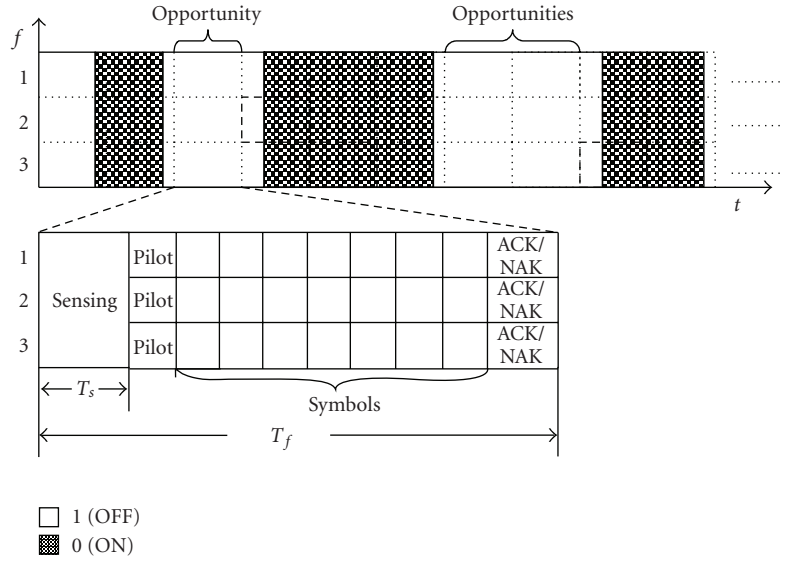

Figure 2: Frame structure, $M = 3$.

For specific spectrum sensing method, such as energy detection, P_{fa}^i and P_{md}^i can be calculated if the average signal-to-noise ratio (SNR) of the licensed user is known [12]. In this paper, we assume the spectrum sensing mechanisms are fixed for both centralized and decentralized cases.

In centralized case, the CR users cooperatively sense the licensed channel. The licensed channel is assumed busy only if the sensing result of every CR user is busy and the probability of false alarm is $\prod_{i=1}^{N} P_{\text{fa}}^i$. By suitable broadcasting mechanism, it is reasonable to assume that the sensing result is identical within the whole CR network. In this case, base station knows the CSI and the power constraint for each CR user and acts as a single controller to design the optimization policy for the whole network.

In decentralized case, the CR users sense the licensed channel separately and only local CSI is available to each CR user and there is no coordination among the CR users. Therefore, the CR users should design cross-layer policy separately.

2.1. Licensed Channel Availability Model. In DSA system, time-varying channel availability should be considered according to the traffic load variation of licensed users, which is assumed to be independent and identically distributed (i.i.d.) alternative renewal process with ON (busy, 0) and OFF (idle, 1) periods [9]. The duration of an ON (OFF) period of channel m is described by an exponentially distributed stochastic variable with parameter λ_0 (λ_1). If CR user sends the pilot and data symbols to the base station without incurring collision with licensed users during a frame, that is, the channel keeps idle during the frame, an opportunity is exploited as a stochastic variable and depends on recent sensing result. Before further discussion, we give the following definition at first.

Definition 1. Duration probability $p_d(i \to j, t, x)$ $i, j \in \{0, 1\}$ denotes the probability that at a specified time instant t ($t \geq 0$), the availability state for a specific licensed channel is j and this licensed channel will keep j for an interval x if the availability state starts with i at $t = 0$.

According to availability state $i \in \{0, 1\}$, the opportunity probability is $p_d(i \to 1, 0, T_f)$, and the collision probability

is $1 - p_d(i \to 1, 0, T_f)$. The expression of $p_d(i \to 1, t, x)$ $i \in \{0, 1\}$ can be expressed as [21, 22]

$$p_d(0 \longrightarrow 1, t, x) = \frac{\exp(-\lambda_0 x)(\lambda_1 - \lambda_1 \exp(-(\lambda_0 + \lambda_1)t))}{(\lambda_0 + \lambda_1)}$$
$$p_d(1 \longrightarrow 1, t, x) = \frac{\exp(-\lambda_0 x)(\lambda_1 + \lambda_0 \exp(-(\lambda_0 + \lambda_1)t))}{(\lambda_0 + \lambda_1)}. \quad (1)$$

Then we can get $p_d(0 \to 1, 0, T_f) = 0$ and $p_d(1 \to 1, 0, T_f) = \exp(-\lambda_0 T_f)$.

2.2. Evolution of Sensing Results. Let $\mathbf{F} = \{0(\text{busy}), 1(\text{idle})\}$ denote the space of sensing results. To derive the state transition probabilities of $\mathbf{F}$, we define point probability as follows.

Definition 2. Point probability $p_p(i \to j, t)$ $i, j \in \{0, 1\}$ denotes the probability that licensed channel availability is j at time t $(t \geq 0)$ if it starts with i at $t = 0$.

According to Definitions 1 and 2, we can consider point probability as duration probability with interval of duration $x = 0$. Meanwhile, it is obvious that the licensed channel availability at the beginning of a frame only depends on the licensed channel availability state obtained in the preceding frame. In addition, the probabilities of sensing false alarm and sensing miss detection are unchanged for each frame. Note that the miss detection of spectrum sensing can be found by pilot symbols. Therefore, the miss detection of spectrum sensing does not affect the change of sensing result and the sensing result of licensed channel can be molded as an ergodic finite state discrete time Markov chain with state space $\mathbf{F}$. Furthermore, according to Definitions 1 and 2, the state transition probabilities of $\mathbf{F}$ can be given as:

$$p_{\mathbf{F}}(f \longrightarrow f') = \begin{cases} p_{\mathbf{F}}\left(0 \longrightarrow 1, T_f\right) = (1 - P_{\text{fa}})\dfrac{\left(\lambda_1 - \lambda_1 \exp\left(-(\lambda_0 + \lambda_1)T_f\right)\right)}{(\lambda_0 + \lambda_1)}, & f = 0,\ f' = 1 \\ p_{\mathbf{F}}\left(1 \longrightarrow 1, T_f\right) = (1 - P_{\text{fa}})\dfrac{\left(\lambda_1 + \lambda_0 \exp\left(-(\lambda_0 + \lambda_1)T_f\right)\right)}{(\lambda_0 + \lambda_1)}, & f = 1,\ f' = 1 \\ p_{\mathbf{F}}\left(1 \longrightarrow 0, T_f\right) = 1 - (1 - P_{\text{fa}})\dfrac{\left(\lambda_1 + \lambda_0 \exp\left(-(\lambda_0 + \lambda_1)T_f\right)\right)}{(\lambda_0 + \lambda_1)}, & f = 1,\ f' = 0 \\ p_{\mathbf{F}}\left(0 \longrightarrow 0, T_f\right) = 1 - (1 - P_{\text{fa}})\dfrac{\left(\lambda_1 - \lambda_1 \exp\left(-(\lambda_0 + \lambda_1)T_f\right)\right)}{(\lambda_0 + \lambda_1)}, & f = 0,\ f' = 0, \end{cases} \quad (2)$$

where $P_{\text{fa}} = \prod_{i=1}^{N} P_{\text{fa}}^i$ in centralized case or $P_{\text{fa}} = P_{\text{fa}}^i$ in decentralized case.

2.3. Rate and Power Adaptation Model. We consider a block fading model to characterize the M parallel narrowband channels, that is, these channels keep constant during each frame. It is well known that block fading channel can be modeled as an ergodic first order finite state discrete time Markov channel (FSMC) [23]. Let $\Gamma_0 = 0 < \Gamma_1, < \cdots < \Gamma_K = \infty$ be a sequence of pre-selected thresholds of received SNR and $\mathbf{H}_i^{(m)} \triangleq \{0, 1, \ldots, (K-1)\}$ denote the channel state space of mth channel for CR user i. The probability distribution of state space $\mathbf{H}_i^{(m)}$ can be given as $p_{\mathbf{H}_i^{(m)}}(k) = \int_{\Gamma^{(k)}}^{\Gamma^{(k+1)}} \exp(-\gamma_i/\overline{\gamma}_i)/\overline{\gamma}_i d\gamma_i$, $k \in \{0, \ldots, K-1\}$, where $\overline{\gamma}_i$ is the average SNR. Therefore, the state transition probabilities of the mth channel are

$$p_{i,\text{FSMC}}^{(m)}(k \longrightarrow (k+1)) = \frac{N\left(\Gamma^{(k+1)}\right)T_f}{p_{\mathbf{H}_i^{(m)}}(k)}, \quad k \in \{0, 1, \ldots, K-2\},$$
$$p_{i,\text{FSMC}}^{(m)}(k \longrightarrow (k-1)) = \frac{N\left(\Gamma^{(k)}\right)T_f}{p_{\mathbf{H}_i^{(m)}}(k)}, \quad k \in \{1, 2, \ldots, K-1\}, \quad (3)$$

where $N(\Gamma^{(k)}) = \sqrt{2\pi\Gamma^{(k)}/\overline{\gamma}^{(m)}} f_{i,\text{Dop}}^{(m)} \exp(-\Gamma^{(k)}/\overline{\gamma}^{(m)})$, and $f_{i,\text{Dop}}^{(m)}$ is maximal Doppler frequency of CR user i [23]. For CR users i, the composite state space of M channels is denoted by $\mathbf{H}_i$, and $\mathbf{H}_i = \mathbf{H}_i^{(1)} \times \mathbf{H}_i^{(2)} \times \cdots \times \mathbf{H}_i^{(M)}$. Correspondingly, the composite channel state is defined as $\mathbf{h}_i \triangleq \{h_i^{(1)}, \ldots, h_i^{(m)}, \ldots, h_i^{(M)}\} \in \mathbf{H}_i$, where $h_i^{(m)} \in \mathbf{H}_i^{(m)}$. If it is assumed that the state transition probabilities between each pair of channels are independent [24] and the transition probability of $\mathbf{h}_i, \mathbf{h}_i' \in \mathbf{H}_i$ is given by $p_{\mathbf{H}_i}(\mathbf{h}_i \to \mathbf{h}_i') = \prod_{m=1}^{M} p_{i,\text{FSMC}}^{(m)}(h_i^{(m)} \to h_i'^{(m)})$.

In the proposed system, adaptive transmission scheme based on M-ary quadrature amplitude modulation (M-QAM) is employed for each channel. Let $\mathbf{V} \triangleq \{0, 1, \ldots, (V-1)\}$ denote the transmission rate space of each channel, in which $\nu \in \{2, 3 \cdots V-1\}$ is corresponding to 2^ν-QAM transmission. Specifically, 0 and 1 are corresponding to no transmission and BPSK transmission, respectively. For given transmission rate, power, and channel state, the bit error rate (BER) can be estimated. Assuming ideal coherent detection, BER bound for $\nu = 1$ is given by [20],

$$p_{\text{BER}}(k, 1) \leq 0.5 \operatorname{erfc}\left(\sqrt{\Gamma^{(k)} P(k, 1)/WN_0}\right). \quad (4)$$

For ν, $\nu \in \{2, 3 \cdots V-1\}$,

$$p_{\text{BER}}(k, \nu) \leq 0.2 \exp\left(\frac{-1.6\Gamma^{(k)} P(k, \nu)}{WN_0(2^\nu - 1)}\right), \quad (5)$$

where WN_0 is noise power. According to (5) and (6), the pessimistic minimum power $P_{\min}(k, \nu)$ can be calculated

to achieve a specified BER bound for channel state k and transmission rate v.

3. Problem Formulation and Discussion

In this section, we consider the cross-layer policy design of the channel access and transmission rate adaptation where each CR user aims at maximizing expected long-term average reward under power constraint. We define the reward as the number of packets successfully transmitted to base station per frame. The policy design will be formulated and discussed in both centralized and decentralized cases.

3.1. Preliminary Definition. To model the stochastic characteristics of the CR networks considered in this paper, we first provide the definition of $\{\mathbf{S}, \mathbf{A}, \mathbf{P}, R, \mathbf{C}\}$, where $\mathbf{S}$ is state space and $\mathbf{A}$,$\mathbf{P}$,R,$\mathbf{C}$ are action space, state transition probability matrix, reward, and cost, respectively.

We can define a composite CR network state space $\mathbf{S} = \mathbf{H}_1 \times \mathbf{H}_2 \times \cdots \times \mathbf{H}_N \times \mathbf{F}$ instinctively, where the notation "$\times$" denotes the Cartesian product. Let $\mathbf{H} = \mathbf{H}_1 \times \mathbf{H}_2 \times \cdots \times \mathbf{H}_N$. Then, $\mathbf{S}$ can be further expressed as $\mathbf{S} = \mathbf{H} \times \mathbf{F}$. The state transition probabilities of $\mathbf{S}$ depend on transition probabilities of composite channel state and sensing result. We assume that sate transition only occurs at the beginning of each frame. Let $\mathbf{s} \triangleq \{\mathbf{h}, f\} \in \mathbf{S}$ denote the whole CR network state at the beginning of a frame. The CR network evolves into a new state $\mathbf{s}' \triangleq \{\mathbf{h}', f'\} \in \mathbf{S}$ at the beginning of the next frame. In this paper, it is assumed that the transition probabilities of composite channel state and channel sensing result are independent of each other [8, 25]. We can express the transition probability as

$$p_{\mathbf{S}}\left(\frac{\mathbf{s} \longrightarrow \mathbf{s}'}{a}\right) = p_{\mathbf{H}}(\mathbf{h} \longrightarrow \mathbf{h}')\, p_{\mathbf{F}}(f \longrightarrow f'). \tag{6}$$

We define $\pi(\mathbf{s})$ to be the steady state probabilities of $\mathbf{s} \in \mathbf{S}$. In the similar way, for each CR user, we can define the local state space $\mathbf{S}_i = \mathbf{H}_i \times \mathbf{F}$ and the steady state probability vector $\pi_i(\mathbf{s}_i)$.

The access action set is defined as $\mathbf{I}_i = I_i^1 \times I_i^j \times \cdots \times I_i^M$, where $I_i^j = 1$ means that CR user i chooses channel j to access and $I_i^j = 0$ means the opposite. We also define the transmission rate space as $\mathbf{V}_i = \mathbf{V}_i^{(1)} \times \mathbf{V}_i^{(2)} \cdots \times \mathbf{V}_i^{(M)}$.

Therefore, for CR user i, we can define the local action space $\mathbf{A}_i = \mathbf{I}_i \times \mathbf{V}_i$, which consists of all access decision and transmission rate decision. The action space of whole CR network can be expressed as $\mathbf{A} = \mathbf{A}_1 \times \mathbf{A}_i \times \cdots \times \mathbf{A}_N$.

In qth decision period (qth frame), if the action of the whole CR network is $\mathbf{a} \in \mathbf{A}$ and the state is $\mathbf{s} \in \mathbf{S}$, the reward for the CR user i is given by

$$r_i(\mathbf{s}, \mathbf{a}) = p_d\left(f_{q,i} \longrightarrow 1, 0, T_f\right) \cdot \sum_{j=1}^{M} I_i^j \cdot \Phi\left(v_i^j\right) \cdot \left(\prod_{k \neq i}\left(1 - I_k^j\right)\right), \quad k, i = 1, 2, \ldots, N, \tag{7}$$

where $f_{q,i}$ is the sensing result at the beginning of qth frame, and $\Phi(v_i^j)$ is the number of transmitted packets in a frame. $\Phi(\cdot)$ is a linear increasing function and is defined as $\Phi(v_i^j) = v_i^j$ for simplification in this paper. If another CR user also accesses channel j, the collision occurs and the reward on the channel j will be 0. We also assume that the packet loss due to transmission failure is only determined by the collision with licensed users when the BER is small enough. $p_d(f_{q,i} \rightarrow 1, 0, T_f)$ is the probability that the licensed channel is idle during T_f when sensing result of CR user i is $f_{q,i}$.

For CR user i, the cost is defined as power consumption and given by

$$w_i(\mathbf{s}, \mathbf{a}) = \sum_{j=1}^{M} P_{\min}\left(h_i^j, v_i^j\right), \tag{8}$$

where $P_{\min}(h_i^j, v_i^j)$ is defined in (4) and (5). For CR user i, the expected long-term average power consumption should be upper bounded by the threshold τ_i, that is,

$$W_i = \lim_{Q \to \infty} \frac{1}{Q} \sum_{q=1}^{Q} \mathbb{E}\left[w_i\left(\mathbf{s}_q, \mathbf{a}_q\right)\right] \leq \tau_i. \tag{9}$$

3.2. Centralized Optimization. We firstly consider maximizing the total expected long-term average reward of the whole CR network in centralized case. This means we should find an optimal policy $u^* \in \mathbf{U}_c$ as

$$u^* = \arg \max_{u \in \mathbf{U}_c} \left\{ \lim_{Q \to \infty} \frac{1}{Q} \sum_{q=1}^{Q} \mathbb{E}\left[\sum_{i=1}^{N} r_i\left(\mathbf{s}_q, u\left(\mathbf{s}_q\right)\right) \right] \right\}$$

$$\text{subject to :} \quad \lim_{Q \to \infty} \frac{1}{Q} \sum_{q=1}^{Q} \mathbb{E}\left[w_i\left(\mathbf{s}_q, u\left(\mathbf{s}_q\right)\right)\right] \leq \tau_i, \quad \forall i \in \{1, \ldots, N\}, \tag{10}$$

where $\mathbf{U}_c$ is the set of centralized policies.

This problem can be formulated as a special Constrained Markov Decision Process (CMDP). That is, the state transition probabilities of the CMDP proposed in this section are not affected by action. According to [26, Theorem 4.3], the standard LP approach for this problem can be formulated as

$$\begin{aligned} \max : \quad & R_u(\rho) = \sum_{\mathbf{s} \in \mathbf{S}} \sum_{\mathbf{a} \in \mathbf{A}} \left(\rho(\mathbf{s}, \mathbf{a}) \sum_{i=1}^{N} r_i\left(\mathbf{s}_q, \mathbf{a}_q\right) \right) \\ \text{subject to :} \quad & \sum_{\mathbf{a} \in \mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) = \pi(\mathbf{s}) = \prod_{i=1}^{N} \pi_i(\mathbf{s}_i), \\ & \sum_{\mathbf{s} \in \mathbf{S}} \sum_{\mathbf{a} \in \mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) w_i(\mathbf{s}, \mathbf{a}) \leq \tau_i, \\ & \sum_{\mathbf{s} \in \mathbf{S}} \sum_{\mathbf{a} \in \mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) = 1, \quad \rho(\mathbf{s}, \mathbf{a}) \geq 0, \\ & \forall \mathbf{s} \in \mathbf{S}, \ \forall \mathbf{a} \in \mathbf{A}, \ \forall i \in \{1, \ldots, N\}, \end{aligned} \tag{11}$$

where occupation measure $\rho(\mathbf{s},\mathbf{a})$ is a variable. By solving (11), we can get the optimal value $\rho^*(\mathbf{s},\mathbf{a})$ for each $(\mathbf{s},\mathbf{a})$ pair and the corresponding optimal policy u^* can be obtained by

$$u_{\mathbf{s}}^*(\mathbf{a}) = \frac{\rho^*(\mathbf{s},\mathbf{a})}{\sum_{\hat{\mathbf{a}}\in\mathbf{A}(\mathbf{s})}\rho^*(\mathbf{s},\hat{\mathbf{a}})} \tag{12}$$

which stands for the probability that the action $\mathbf{a}$ is chosen when the network state is $\mathbf{s}$.

3.3. Decentralized Case. In decentralized case, each CR user does not know any information except the probability of sensing false alarm P_{fa}^i, steady state probabilities π_i, and the power threshold τ_i. We define $u = (u_1,\ldots,u_N) \in \mathbf{U}_d$ as the class of decentralized policies. As mentioned in the previous section, the CR users aim at maximizing the total reward of the whole CR network. It means that all CR users have the common maximizing object

$$R(u) = R_i(u) = \lim_{Q\to\infty}\frac{1}{Q}\sum_{q=1}^{Q}\mathbb{E}\left[\sum_{i=1}^{N} r_i\left(\mathbf{s}_q,\mathbf{a}_q\right)\right], \quad \forall i\in\{1,\ldots,N\}, \tag{13}$$

where $\mathbf{s}_q$ and $\mathbf{a}_q$ can be expressed as $\mathbf{s}_q = \mathbf{s}_{1,q}\times\mathbf{s}_{i,q}\times\cdots\times\mathbf{s}_{N,q}$ and $\mathbf{a}_q = \mathbf{a}_{1,q}\times\mathbf{a}_{i,q}\times\cdots\times\mathbf{a}_{N,q}$, respectively.

For a policy $u = (u_1,\ldots,u_N) \in \mathbf{U}_d$, we define u^{-i} as the subset of u by deleting the ith component. We further define $[z_i, u^{-i}]$ as all CR users except user i use the element in u whereas user i uses the policy z_i.

Definition 3. $u^* \in \mathbf{U}_d$ is a constrained Nash equilibrium [27] if it satisfies the power constraints (9) for all users and

$$R_i(u^*) \geq R_i\left(\left[z_i,(u^*)^{-i}\right]\right), \quad \forall i\in\{1,\ldots,N\} \tag{14}$$

for any $z_i \in \mathbf{U}_{d,i}$ such that the power constraints are satisfied for the policy $[z_i,(u^*)^{-i}]$.

Theorem 1. *Any policy $u = (u_1,\ldots,u_N) \in \mathbf{U}_d$ maximizing (13) while satisfying the power constraints (9) is a constrained Nash equilibrium in this cooperative game.*

Proof. Assume policy u maximizing (13) satisfies the power constraints but is not a constrained Nash equilibrium. According to Definition 3, there must exist a CR user i and the policy z_i, such that the power constraint of this CR user i holds and $R_i(u) < R_i([z_i,u^{-i}])$. Furthermore, based on (8), the power constraints of all CR users can be satisfied by the policy $[z_i,u^{-i}]$. This result contradicts with the assumption that the policy u maximizes (13). Therefore, we conclude that u is a constrained Nash equilibrium. □

Lemma 1. *All CR users can calculate the optimal policy $u^* \in \mathbf{U}_d$ by solving the same optimization problem*

$$u^* = \arg\max_{u\in\mathbf{U}_d} R_i(u) \quad \text{subject to}: \quad W_i(u^*) \leq \tau_i, \quad \forall i\in\{1,\ldots,N\}, \tag{15}$$

where u^* is necessarily a constrained Nash equilibrium and also a globally optimal policy in decentralized case.

Proof. Problem (15) can be formulated as

$$\begin{aligned} \max:\ & R(u) \\ \text{subject to}:\ & \sum_{\mathbf{a}_i\in\mathbf{A}_i}\rho_i(\mathbf{s}_i,\mathbf{a}_i) = \pi(\mathbf{s}_i), \\ & \sum_{\mathbf{s}_i\in\mathbf{S}_i}\sum_{\mathbf{a}_i\in\mathbf{A}_i}\rho_i(\mathbf{s}_i,\mathbf{a}_i)w_i(\mathbf{s}_i,\mathbf{a}_i) \leq \tau_i, \\ & \sum_{\mathbf{s}_i\in\mathbf{S}_i}\sum_{\mathbf{a}_i\in\mathbf{A}_i}\rho_i(\mathbf{s}_i,\mathbf{a}_i) = 1, \quad \rho_i(\mathbf{s}_i,\mathbf{a}_i)\geq 0, \\ & \forall i\in\{1,\ldots,N\},\ \forall\mathbf{s}_i\in\mathbf{S}_i,\ \forall\mathbf{a}_i\in\mathbf{A}_i, \end{aligned} \tag{16}$$

where $\rho_i(\mathbf{s}_i,\mathbf{a}_i)$ is the occupation measure of local state and local action of CR user i. Each CR user solves the same optimization problem (16) and the corresponding optimal policy u_i^* can be obtained similarly as (12). The solution u^* must be a constrained Nash equilibrium and a globally optimal policy in decentralized case. □

In the case that the CR network just has two CR users, the globally optimal policy can be simply calculated. By defining $\mathbf{s} = (\mathbf{s}_1,\mathbf{s}_2)$ and $\mathbf{a} = (\mathbf{a}_1,\mathbf{a}_2)$, the problem (16) can be expressed as

$$\begin{aligned} \max:\ & R(u) = \sum_{\substack{\mathbf{s}_1\in\mathbf{S}_1\\ \mathbf{a}_1\in\mathbf{A}_1}}\sum_{\substack{\mathbf{s}_2\in\mathbf{S}_2\\ \mathbf{a}_2\in\mathbf{A}_2}}\rho_1(\mathbf{s}_1,\mathbf{a}_1)(r_1(\mathbf{s},\mathbf{a})+r_2(\mathbf{s},\mathbf{a})) \\ & \qquad\qquad \times\rho_2(\mathbf{s}_2,\mathbf{a}_2) \\ \text{subject to}:\ & \sum_{\mathbf{a}_i\in\mathbf{A}_i}\rho_i(\mathbf{s}_i,\mathbf{a}_i) = \pi(\mathbf{s}_i), \\ & \sum_{\mathbf{s}_i\in\mathbf{S}_i}\sum_{\mathbf{a}_i\in\mathbf{A}_i}\rho_i(\mathbf{s}_i,\mathbf{a}_i)w_i(\mathbf{s}_i,\mathbf{a}_i) \leq \tau_i, \\ & \sum_{\mathbf{s}_i\in\mathbf{S}_i}\sum_{\mathbf{a}_i\in\mathbf{A}_i}\rho_i(\mathbf{s}_i,\mathbf{a}_i) = 1, \\ & \rho_i(\mathbf{s}_i,\mathbf{a}_i)\geq 0, \quad \forall i\in\{1,2\},\ \forall\mathbf{s}_i\in\mathbf{S}_i,\ \forall\mathbf{a}_i\in\mathbf{A}_i. \end{aligned} \tag{17}$$

4. Simplification of Policy Design

With V rates, K channel states, M channels, and CR users $N = 2$, the LP (11) has $2(K^M)^2(2^MV^M)^2$ variables and the LP (17) has $4K^M2^MV^M$ variables. The variables increase exponentially both in centralized and decentralized cases. Consequently, it is impossible to design the optimal policy in real-time in response to evolving parameters (λ_1 and λ_0). In this section, we simplify the LP (11) and (17) by reducing the variable number in two ways without loss of the optimality of the policy. One way is to transform the multichannel policy design to single channel policy design, while the other reduces action set and aggregates states. For the former method, we have the following theorem.

Theorem 2. *Under the condition that the maximal Doppler frequency $f_{i,\mathrm{Dop}}^{(m)}$ is the same on every channel for the CR user i, the policy design can be solved separately for each channel without loss of optimality.*

Proof. See Appendix B. □

In the case of single channel, the action set is $\mathbf{A}_i = \mathbf{I}_i \times \mathbf{V}_i$, where $I_i = 0$ means that the CR user does not access this channel and $v_i = 0$ means that there is no data transmission. In fact, the access action $I_i = 0$ can be obtained in the transmission rate state $v_i = 0$. Therefore, the action set can be further reduced and the reduced action set is defined as $\mathbf{A}_i = \mathbf{V}_i$. On the other hand, any state $\mathbf{s}_i \in \mathbf{H}_i \times \{0\}$ is aggregated into a macro-state $\mathbf{s}_{\mathrm{Agg}}$ by state aggregation. Consequently, we have the following propositions.

Proposition 1. *Based on Theorem 2, for the optimal policy design on each channel, the new action and state space of CR user i are respectively expressed as $\mathbf{A}_i \triangleq \mathbf{V}_i \triangleq \{0, 1, \ldots, (V-1)\}$ and $\overline{\mathbf{S}}_i \triangleq \{\mathbf{H}_i \times \{1\}, \mathbf{s}_{\mathrm{Agg}}\} \triangleq \{\{0, 1, \ldots, (K-1)\} \times \{1\}, \mathbf{s}_{\mathrm{Agg}}\}$. The new action and state space of the whole CR network are respectively expressed as $\mathbf{A} = \mathbf{A}_1 \times \cdots \times \mathbf{A}_N$ and $\overline{\mathbf{S}} \triangleq \{\mathbf{H}_1 \times \cdots \times \mathbf{H}_N \times \{1\}, \mathbf{s}_{\mathrm{Agg}}\}$.*

Proposition 2. *Based on Theorem 2 and Proposition 1, the action set reduction and state aggregation do not affect the optimality of the policy design. In addition, the optimal occupation measures $\rho^*(\overline{\mathbf{s}}, \mathbf{a})$ can be calculated according to the original occupation measures $\rho^*(\mathbf{s}, \mathbf{a})$ which correspond to the original optimal policy u^*, that is,*

$$\begin{aligned}
&\rho^*(\overline{\mathbf{s}}, \mathbf{a}) = \rho^*(\mathbf{s}, \mathbf{a}), \quad \forall \overline{\mathbf{s}} = \mathbf{s} \in \mathbf{H}_1 \times \cdots \times \mathbf{H}_N \times \{1\},\\
&\rho^*(\overline{\mathbf{s}}, \mathbf{a}) = 0, \quad \overline{\mathbf{s}} = \mathbf{s}_{\mathrm{Agg}}, \ \forall \mathbf{a} \neq 0 \times \cdots \times 0,\\
&\rho^*(\overline{\mathbf{s}}, \mathbf{a}) = \sum_{f=0} \rho^*(\mathbf{s}, \mathbf{a}), \quad \overline{\mathbf{s}} = \mathbf{s}_{\mathrm{Agg}}, \ \mathbf{a} = 0 \times \cdots \times 0.
\end{aligned} \tag{18}$$

After the simplification of policy design, the variable number of centralized policy is reduced from $2(K^M)^2(2^M V^M)^2$ to $(K^2+1)V^2$ and the variable number of decentralized policy is reduced from $4K^M 2^M V^M$ to $(K+1)V$. Obviously, the complexity of the optimal policy design is largely reduced.

5. Pure Policy

It is noticed that the previous discussion on the optimal policy is based on the mixed policy. In fact, the controller prefers pure policy to mixed policy due to the convenience of implementation and evaluation. For pure policy, the action selection is not stochastic but only one action could be adopted for each state. The optimal pure policy exists for centralized case and the LP for the optimal pure policy can be formulated as

$$\begin{aligned}
\max: \quad & R_u(\rho) = \sum_{\mathbf{s}\in\mathbf{S}} \sum_{\mathbf{a}\in\mathbf{A}} \left(\rho(\mathbf{s}, \mathbf{a}) \sum_{i=1}^{N} r_i(\mathbf{s}, \mathbf{a}) \right)\\
\text{subject to}: \quad & \sum_{\mathbf{a}\in\mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) = \pi(\mathbf{s}) = \prod_{i=1}^{N} \pi_i(\mathbf{s}_i),\\
& \sum_{\mathbf{s}\in\mathbf{S}} \sum_{\mathbf{a}\in\mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) w_i(\mathbf{s}, \mathbf{a}) \leq \tau_i,\\
& \sum_{\mathbf{s}\in\mathbf{S}} \sum_{\mathbf{a}\in\mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) = 1, \quad \rho(\mathbf{s}, \mathbf{a}) = \{0, \pi(\mathbf{s})\},\\
& \forall i \in \{1, \ldots, N\}, \ \forall \mathbf{s} \in \mathbf{S}, \ \forall \mathbf{a} \in \mathbf{A},
\end{aligned} \tag{19}$$

For decentralized case, we have the following theorems.

Theorem 3. *In decentralized case, the constrained Nash equilibrium exists under the condition of pure polices.*

Proof. Under the condition of pure policy, the state space and action space are finite and countable. Therefore, Theorem 1 still holds and the constrained Nash equilibrium exists under the condition of pure polices. □

When $N = 2$, we can get the optimal pure policy for decentralized case in similar way as (17).

Theorem 4. *In decentralized case, Theorem 2 does not hold under the condition of pure polices.*

Proof. Note that under the condition of pure policy, the state space and action space are finite and countable. If the policy design is solved separately in every channel, the state space and action space actually are reduced and the optimality is not guaranteed. □

6. Parameter Estimation

In CR network, the change of spectrum hole depends on the spectrum occupancy of licensed users. But the CR users do not know the traffic load parameters (λ_0 and λ_1) of licensed users generally. According to the analysis in the above sections, the design of optimal policy requires the knowledge about system state transition probabilities associated with the traffic load parameters. In this section, we estimate the traffic load parameters (λ_0 and λ_1) of licensed users in the following two cases:

(1) The value of parameter pair (λ_0, λ_1) has the constant value $(\lambda_0, \lambda_1)^0$ belongs to a fixed finite set $\boldsymbol{\lambda} = \{(\lambda_0, \lambda_1)^1, \ldots, (\lambda_0, \lambda_1)^q\}$.

(2) There is no prior information about the value of parameter pair (λ_0, λ_1).

6.1. Fixed Finite Set. We construct the following adaptive control rule. At each frame n, the CR users make the maximum likelihood estimate $(\lambda_0, \lambda_1)_n$ after the channel sensing phase

$$\begin{aligned}
&\mathrm{Prob}\{x_0, \ldots, x_n \mid a_0, \ldots, a_{n-1}, (\lambda_0, \lambda_1)_n\}\\
&\quad = \prod_{i=0}^{t-1} p(x_i, x_{i+1}; a_i, (\lambda_0, \lambda_1)_n)\\
&\quad \geq \prod_{i=0}^{t-1} p\left(x_i, x_{i+1}; a_i, (\lambda_0, \lambda_1)^j\right), \quad \forall (\lambda_0, \lambda_1)^j \in \boldsymbol{\lambda},
\end{aligned} \tag{20}$$

where x_i is the sensing result in frame i, $p(x_i,x_{i+1};a_i,(\lambda_0,\lambda_1)_n)$ is the transition probability under the action a_i, and the parameter value is $(\lambda_0,\lambda_1)_n$.

Theorem 5. *The convergence of* $\lim_{n\to\infty}(\lambda_0,\lambda_1)_n = (\lambda_0,\lambda_1)^0$ *is guaranteed by performing the maximum likelihood estimate of* (20).

Proof. Mandi has given the convergence condition in [28] by stating that for each $(\lambda_0,\lambda_1)\neq(\lambda_0,\lambda_1)'$ if there exists $x\in\{0,1\}$ so that

$$\begin{aligned}&\left[p(x,0;a,(\lambda_0,\lambda_1)),p(x,1;a,(\lambda_0,\lambda_1))\right]\\&\quad\neq\left[p\left(x,0;a,(\lambda_0,\lambda_1)'\right),p\left(x,1;a,(\lambda_0,\lambda_1)'\right)\right],\quad\forall a\in\mathbf{A}\end{aligned}\tag{21}$$

then the convergence of $\lim_{n\to\infty}(\lambda_0,\lambda_1)_n = (\lambda_0,\lambda_1)^0$ is guaranteed. Note that the transition probability of the sensing result is not affected by the action choice of CR users. It is obvious that (21) can be satisfied in this system and the convergence property is assured. Furthermore, the maximum likelihood estimation (20) can be simplified as

$$\begin{aligned}&\mathrm{Prob}\{x_0,\ldots,x_n\mid(\lambda_0,\lambda_1)_n\}\\&\quad=\prod_{i=0}^{t-1}p(x_i,x_{i+1};(\lambda_0,\lambda_1)_n)\\&\quad\geq\prod_{i=0}^{t-1}p\left(x_i,x_{i+1};(\lambda_0,\lambda_1)^j\right),\quad\forall(\lambda_0,\lambda_1)^j\in\boldsymbol{\lambda}.\end{aligned}\tag{22}$$

□

6.2. No Prior Information. We consider the first order moment estimation for the case of no prior information about the value of parameter pair (λ_0,λ_1). The statistic transition probability matrix of sensing results can be defined as

$$\mathbf{P}=\begin{pmatrix}p_{00}&p_{01}\\p_{10}&p_{11}\end{pmatrix}.\tag{23}$$

Based on sensing result, the statistic transition probability matrix can be updated frame by frame. In frame n, according to (2), $(\lambda_0^n,\lambda_1^n)$ can be estimated by the following equations

$$\begin{aligned}p_{11}-p_{01}&=(1-P_{\mathrm{fa}})\exp\left(-(\lambda_0^n+\lambda_1^n)T_f\right)\\p_{11}+p_{01}&=(1-P_{\mathrm{fa}})\\&\quad\times\frac{2\lambda_1^n+(\lambda_0^n+\lambda_1^n-2\lambda_1^n)\exp\left(-(\lambda_0^n+\lambda_1^n)T_f\right)}{\lambda_0^n+\lambda_1^n}.\end{aligned}\tag{24}$$

7. Numerical Results

In this section, we present numerical results to evaluate the performances of the proposed policies in both centralized and decentralized cases. In the numerical computation, the number of CR users is set to be $N = 2$. Traffic load parameters of licensed channel are given by $1/\lambda_0 = 100$ ms and $1/\lambda_1 = 25$ ms. The length of frame $T_f = 2$ ms and the spectrum sensing false alarm is set to be 0.1 for each CR user. Licensed channel is divided into M narrow channels. To illustrate the influence of M on the performances of pure policies, we consider 3 different cases: $M = 1$, $M = 2$, and $M = 4$. Each channel has $K = 3$ states, and $\Gamma^{(1)} = -8.47$ dB, $\Gamma^{(2)} = -0.08$ dB. Noise power WN_0 and target BER are set to be 1 mW, 10^{-3}, respectively. Average SNR and maximal Doppler frequency of each channel are 0 dB and 50 Hz, respectively. We adopt four modulation schemes which are BPSK, 4-QAM, 8-QAM, and 16-QAM. Then we have $\mathbf{V}\triangleq\{0,1,2,3,4\}$.

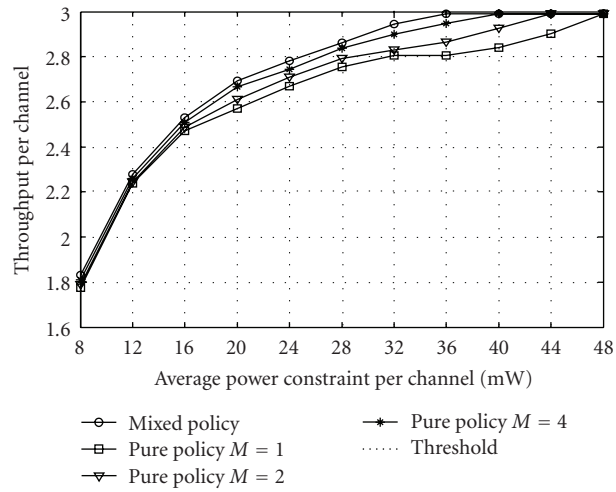

Figure 3: Performance comparison between centralized optimal mixed policies and pure policies under different average power constraints.

In Figure 3, throughput versus average power constraint for different policies in centralized case is presented. Due to the cooperative spectrum sensing, the sensing false alarm is 0.01 and the throughput is almost as same as the perfect spectrum sensing case (Note that the miss detection can be found by the pilot symbols). It can be seen that the mixed polices are not affected by the value of M and this result coincides with Theorem 2. We further notice that the throughput of pure policies is improved with the increase of M. This coincides with Theorem 4. The reason which leads to this difference between mixed polices and pure polices is as follows. For pure polices, the number of occupation measure $\rho(\mathbf{s},\mathbf{a})$ in (19) is $((K^2+1)V^2)^M$ and increases exponentially with M. Then, the larger M is, the more feasible solutions to (19) are provided. On the other hand, the feasible solutions of mixed polices are infinite and uncountable. Moreover, we can find that all polices reach the throughput threshold as the power constraint increases. In this case, if one of the CR users gets the transmission chance, it always chooses the maximum transmission rate to transmit data.

In Figure 4, the throughputs of different polices in decentralized case are plotted with the different average

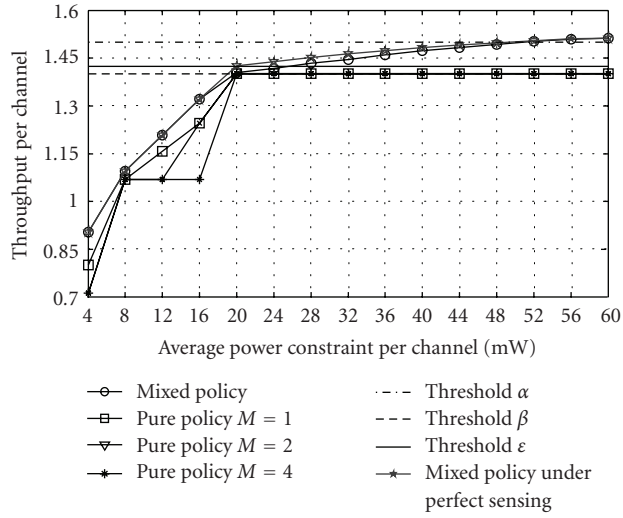

Figure 4: Performance comparison of decentralized optimal mixed policies and pure policies under different average power constraints.

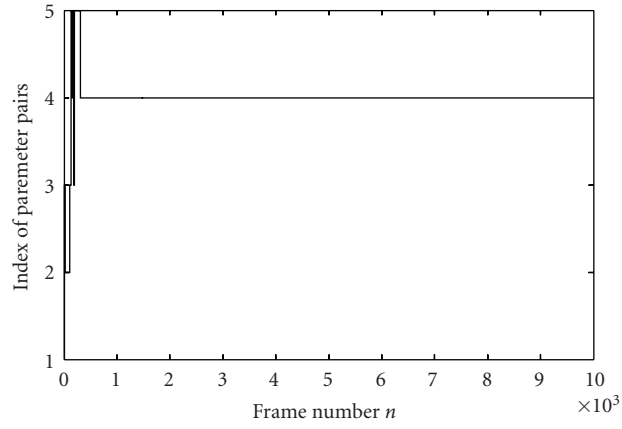

Figure 5: Parameter pair estimation in the fixed finite set.

power constraints. For sake of comparison, we also provide the result for perfect sensing. Generally, the throughputs of pure policies are improved with the increase of M. It is found that the optimal solutions of different M are the same under certain power constraints. The reason is that in decentralized case, the number of feasible solutions is much less than that of the centralized case. Moreover, we can find that mixed policy and pure policy reach different throughput thresholds with the increase of power constraints. Note that in pure polices, if the channel state is $h^{(1)}$ (middle state) and the sensing result is idle, there are only two choices for each CR user which are transmission with probability one or keep silence with probability one. The latter choice is better than the first one and the threshold β is achieved in this case. This is why the threshold α is different with the threshold β.

Different with the centralized case, the sensing false alarm affects the performance obviously. Due to the sensing false alarm, the throughput threshold of pure polices β is less than the threshold ε, which can only be reached by pure

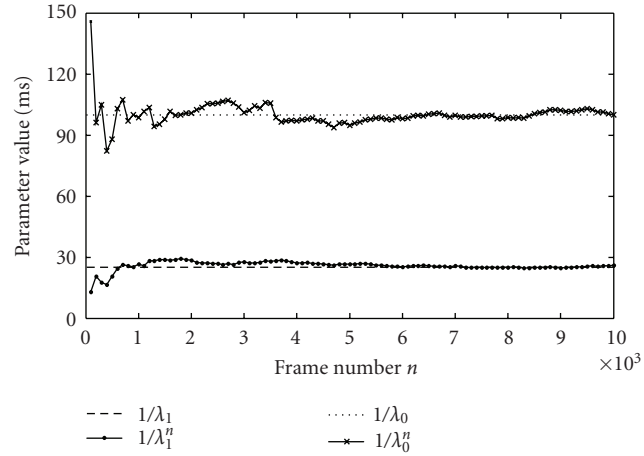

Figure 6: Parameter pair estimation without prior information.

polices in perfect spectrum sensing case. For mixed polices, the throughput thresholds are the same. But in generally, the performance of imperfect sensing is less than that of the perfect sensing case.

In Figure 5, the traffic load parameters pair (λ_0 and λ_1) of licensed users is estimated. In this case, we assume that the CR users have the prior knowledge that the constant value $(\lambda_0, \lambda_1)^0 = (10, 40)$ belongs to a fixed finite set $\boldsymbol{\lambda}$. $\boldsymbol{\lambda}$ is set to be $\{(10, 10)^1, (15, 30)^2, (20, 50)^3, (10, 40)^4, (12, 60)^5\}$.

Figure 5 shows the change of the estimated parameters pair (λ_0, λ_1) with the increase of estimation frame number n. As expected, the convergence of $\lim_{n\to\infty}(\lambda_0, \lambda_1)_n = (\lambda_0, \lambda_1)^0$ is observed and coincides with Theorem 5.

In Figure 6, we estimate the traffic load parameters (λ_0 and λ_1) of licensed user in the case that no prior information is known about the value of parameter pair (λ_0, λ_1). The constant value of $(1/\lambda_0, 1/\lambda_1)$ is set to be (25 ms, 100 ms). It can be seen that, the estimated parameters $(\lambda_0^n, \lambda_1^n)$ approach to the constant value (λ_0, λ_1) with the increase of estimation frame number n. This observation coincides with the previous analysis.

8. Conclusions and Future Works

In this paper, we consider cross-layer design of joint channel access and transmission rate adaptation in CR networks in both centralized and decentralized cases. In centralized case, the cross-layer design can be solved by formulated as a special CMDP, and the optimal policy can be achieved. In decentralized case, we prove the existence of the constrained Nash equilibrium and characterize the structure of optimal decentralized policy. We point out that in both the centralized and decentralized cases, the complexity of finding optimal policy increases exponentially with the size of action space and state space, which incurs so-called curse of dimensionality. Therefore, we apply action set reduction and state aggregation to reduce the complexity without loss of optimality. We further prove that under certain condition, the multichannel access and transmission rate adaptation policy

Table 1

Notation	Description	Notation	Description
T_f	Frame length	T_s	Channel sensing time
P_{fa}	Probability of sensing false alarm	P_{md}	Probability of sensing miss detection
$\lambda_0(\lambda_1)$	Traffic load parameters of licensed user	WN_0	Noise power
$\mathbf{S}$	State space	$\mathbf{H}$	Composite state space of M channels
$\mathbf{H}^{(m)}$	State space of channel m	$\mathbf{F}$	Space of sensing results
$\mathbf{S}_i$	State space of user i	$\mathbf{H}_i$	Composite state space of M channels for user i
$\mathbf{H}_i^{(m)}$	State space of channel m for user i	$\mathbf{V}$	Transmission rate space
$p_{\text{BER}}(\cdot,\cdot)$	Bit error rate	$\pi(\mathbf{s})$	Steady state probability
$\pi_i(\mathbf{s}_i)$	Steady state probability of user i	$\mathbf{I}_i$	Access action set of user i
$\mathbf{V}_i$	Transmission rate space of user i	$\mathbf{A}_i$	Action space of user i. $\mathbf{A}_i = \mathbf{I}_i \times \mathbf{V}_i$
$\mathbf{A}$	Action space of whole CR network	f_q	Sensing result at the beginning of qth frame
$\Phi(\ )$	Mapping from the number of packets to transmission rate	$P_{\min}(h_i^j, v_i^j)$	Minimum transmission power to achieve a specified BER for channel state h_i^j and transmission rate v_i^j.
$r_i(\ )$	Immediate reward of user i	$w_i(\ )$	Immediate power consumption of user i
$R(\ )$	Expected reward of whole CR network	W_i	Expected power consumption of user i
τ_i	average power consumption threshold of user i	$\mathbf{s}_q$	System state in qth decision period
$\mathbf{U}_c$	The set of centralized policies	$u(\mathbf{s}_q)$	Action choice for state $\mathbf{s}_q$ according to policy u
$\mathbf{U}_d$	The set of decentralized policies	u^*	Optimal policy
$\rho(\mathbf{s},\mathbf{a})$	Occupation measure	$\rho_i(\mathbf{s}_i,\mathbf{a}_i)$	Local occupation measure of CR user i
$\rho^*(\mathbf{s},\mathbf{a})$	Optimal occupation measure	$\rho_i^*(\mathbf{s}_i,\mathbf{a}_i)$	Optimal local occupation measure of CR user i
$\mathbf{S}'$	State space based on state aggregation of $\mathbf{S}$	$\mathbf{S}_i'$	State space based on state aggregation of $\mathbf{S}$
$f_{i,\text{Dop}}^{(m)}$	Maximal Doppler frequency of mth channel for user i	$\overline{\gamma}_i$	Average SNR
$p_d(\cdot \to \cdot,\cdot,\cdot)$	Duration probability	$p_p(\cdot \to \cdot,\cdot)$	Point probability
$p_{\mathbf{S}}(\cdot \to \cdot/\cdot)$	State transition probability of $\mathbf{S}$	$p_{\mathbf{H}}(\cdot \to \cdot)$	State transition probability of $\mathbf{H}$
$p_{\mathbf{F}}(\cdot \to \cdot)$	Transition probability of $\mathbf{F}$	$\times$	Cartesian product

design can be solved separately with respect to every channel without loss of optimality. Furthermore, the pure polices are investigated and compared with the mixed polices. Finally, under the condition that the traffic load parameters of the licensed user are unknown for the CR users, we provide two different methods to estimate the parameters in two different cases.

In the future, we will extend our work to finite buffer with stochastic packet arrival and also concern the learning mechanism for the unknown CR environments.

Appendices

A. Notations Table

Notations are listed in Table 1.

B. Proof of Theorem 1

Assume u^* is the optimal policy and the $\rho^*(\mathbf{s},\mathbf{a})$ is the solution of LP (11), the state s and action $\mathbf{a}$ can be expressed

as $\mathbf{s} = \mathbf{s}_1 \times \cdots \times \mathbf{s}_m \times \cdots \times \mathbf{s}_M$ and $\mathbf{a} = \mathbf{a}_1 \times \cdots \times \mathbf{a}_m \times \cdots \times \mathbf{a}_M$, respectively. Here, $\mathbf{s}_m$ and $\mathbf{a}_m$ are the corresponding state and action on channel m. Based on $\rho^*(\mathbf{s}, \mathbf{a})$, we can obtain the new occupation measure on channel m as

$$\begin{aligned}&\rho(\mathbf{s}_m, \mathbf{a}_m)\\ &= \sum_{k \neq m} \sum_{j \neq m} \rho^*\\ &\quad \times \left(\mathbf{s}_1 \times \cdots \times \mathbf{s}_j \times \cdots \times \mathbf{s}_M, \mathbf{a}_1 \times \cdots \times \mathbf{a}_k \times \cdots \times \mathbf{a}_M\right).\end{aligned} \tag{B.1}$$

Because the maximal Doppler frequency $f_{i,\mathrm{Dop}}^{(m)}$ of every channel is the same, each channel can be constructed as the same Markov chain and then the state transition matrix on each channel is the same. By defining

$$\rho^*(\mathbf{s}_m, \mathbf{a}_m) = \frac{\sum_{m=1}^{M} \rho(\mathbf{s}_m, \mathbf{a}_m)}{M} \tag{B.2}$$

we have

$$\begin{aligned}\max: \quad & R = M \sum_{\mathbf{s}_m \in \mathbf{S}_m} \sum_{\mathbf{a}_m \in \mathbf{A}_m} \left(\rho^*(\mathbf{s}_m, \mathbf{a}_m) \sum_{i=1}^{N} r_i(\mathbf{s}_m, \mathbf{a}_m)\right)\\ \text{subject to}: \quad & \sum_{\mathbf{a}_m \in \mathbf{A}_m} \rho^*(\mathbf{s}_m, \mathbf{a}_m) = \pi(\mathbf{s}_m),\\ & M \sum_{\mathbf{s}_m \in \mathbf{S}_m} \sum_{\mathbf{a}_m \in \mathbf{A}_m} \rho^*(\mathbf{s}_m, \mathbf{a}_m) w_i(\mathbf{s}_m, \mathbf{a}_m) \leq \tau_i,\\ & \sum_{\mathbf{s}_m \in \mathbf{S}_m} \sum_{\mathbf{a}_m \in \mathbf{A}_m} \rho^*(\mathbf{s}_m, \mathbf{a}_m) = 1, \quad \rho^*(\mathbf{s}_m, \mathbf{a}_m) \geq 0,\\ & \forall i \in \{1, \ldots, N\}, \ \forall \mathbf{s}_m \in \mathbf{S}_m, \ \forall \mathbf{a}_m \in \mathbf{A}_m.\end{aligned} \tag{B.3}$$

We define the new state space and action space: $\mathbf{A} = \mathbf{A}_m$, $\mathbf{S} = \mathbf{S}_m$ and formulate the new LP as

$$\begin{aligned}\max: \quad & R_u(\rho) = \sum_{\mathbf{s} \in \mathbf{S}} \sum_{\mathbf{a} \in \mathbf{A}} \left(\rho(\mathbf{s}, \mathbf{a}) \sum_{i=1}^{N} r_i(\mathbf{s}, \mathbf{a})\right)\\ \text{subject to}: \quad & \sum_{\mathbf{a} \in \mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) = \pi(s) = \prod_{i=1}^{N} \pi_i(s_i),\\ & \sum_{\mathbf{s} \in \mathbf{S}} \sum_{\mathbf{a} \in \mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) w_i(\mathbf{s}, \mathbf{a}) \leq \frac{\tau_i}{M},\\ & \sum_{\mathbf{s} \in \mathbf{S}} \sum_{\mathbf{a} \in \mathbf{A}} \rho(\mathbf{s}, \mathbf{a}) = 1,\\ & \rho(\mathbf{s}, \mathbf{a}) \geq 0,\\ & \forall i \in \{1, \ldots, N\}, \ \forall \mathbf{s} \in \mathbf{S}, \ \forall \mathbf{a} \in \mathbf{A}.\end{aligned} \tag{B.4}$$

The occupation measure defined by (B.3) is also obtained in the feasible solution set of (B.4). Therefore, the policy design can be solved separately with respect to every channel without loss of optimality.

In the similar way, we can prove that Theorem 2 can also be applied to the decentralized case.

Acknowledgments

The authors would like to thank Prof. Xuesong Tan for his valuable suggestions and help in preparing this paper and the two anonymous reviewers for their very helpful suggestions and comments. This work is supported in part by High-Tech Research and Development Program of China under Grant no. 2007AA01Z209, 2009AA011801, and 2009AA012002, National Fundamental Research Program of China under Grant A1420080150, and National Basic Research Program (973 Program) of China under Grant no. 2009CB320405, Nation Grand Special Science and Technology Project of China under Grant no. 2008ZX03005-001, National Natural Science Foundation of China under Grant no. 60702073, 60972029, and Special Project on Broadband Wireless Access sponsored by Huawei co., LTD.

References

[1] D. Cabric, I. D. O'Donnell, M. S.-W. Chen, and R. W. Brodersen, "Spectrum sharing radios," *IEEE Circuits and Systems Magazine*, vol. 6, no. 2, pp. 30–45, 2006.

[2] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," *IEEE Communications Magazine*, vol. 44, no. 6, pp. 44–49, 2006.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[4] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[5] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[6] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, 2007.

[7] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, 2007.

[8] Y. Chen, Q. Zhao, and A. Swami, "Distributed spectrum sensing and access in cognitive radio networks with energy constraint," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 783–797, 2009.

[9] H. Kim and K. G. Shin, "Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 533–545, 2008.

[10] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2053–2071, 2008.

[11] Y. Pei, A. T. Hoang, and Y.-C. Liang, "Sensing-throughput tradeoff in cognitive radio networks: how frequently should spectrum sensing be carried out?" in *Proceedings of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, pp. 1–5, Athens, Greece, September 2007.

[12] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE*

Transactions on Wireless Communications, vol. 7, no. 4, pp. 1326–1337, 2008.

[13] A. Ghasemi and E. S. Sousa, "Optimization of spectrum sensing for opportunistic spectrum access in cognitive radio networks," in *Proceedings of the 4th Annual IEEE Consumer Communications and Networking Conference (CCNC '07)*, pp. 1022–1026, Las Vegas, Nev, USA, January 2007.

[14] H. Liu, B. Krishnamachari, and Q. Zhao, "Cooperation and learning in multiuser opportunistic spectrum access," in *Proceedings of the IEEE International Conference on Communications (ICC '08)*, pp. 487–492, Beijing, China, May 2008.

[15] Z. Liang, W. Liu, P. Zhou, and F. Gao, "Randomized multi-user strategy for spectrum sharing in opportunistic spectrum access network," in *Proceedings of IEEE International Conference on Communications (ICC '08)*, pp. 477–481, Beijing, China, May 2008.

[16] K. Liu, Q. Zhao, and Y. Chen, "Distributed sensing and access in cognitive radio networks," in *Proceedings of the 10th IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA '08)*, pp. 23–27, Bologna, Italy, August 2008.

[17] H. Liu and B. Krishnamachari, "Randomized strategies for multi-user multi-channel opportunity sensing," in *Proceedings of IEEE Workshop on Cognitive Radio Networks (CCNC '08)*, January 2008.

[18] E. Altman, K. Avratchenkov, N. Bonneau, M. Debbah, R. El-Azouzi, and D. S. Menasché, "Constrained stochastic games in wireless networks," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 315–320, Washington, DC, USA, November 2007.

[19] E. Altman, K. Avrachenkov, G. Miller, and B. Prabhu, "Discrete power control: cooperative and non-cooperative optimization," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 37–45, Anchorage, Alaska, USA, May 2007.

[20] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, 2001.

[21] R .E. Barlow and L. C. Hunter, "Reliability analysis of a one-unit system," *Operations Research*, vol. 9, no. 2, pp. 200–208, 1961.

[22] L. A. Baxter, "Availability measures for a two-state system," *Journal of Applied Probability*, vol. 18, pp. 227–235, 1981.

[23] H. S. Wang and N. Moayeri, "Finite-state Markov channel—a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, 1995.

[24] Md. J. Hossain, D. V. Djonin, and V. K. Bhargava, "Delay limited optimal and suboptimal power and bit loading algorithms for OFDM systems over correlated fading channels," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 5, pp. 2787–2792, St. Louis, Mo, USA, November-December 2005.

[25] A. K. Karmokar, D. V. Djonin, and V. K. Bhargava, "Optimal and suboptimal packet scheduling over time-varying flat fading channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 446–457, 2006.

[26] E. Altman, *Constrained Markov Decision Process: Stochastic Modeling*, Chapman & Hall/CRC, London, UK, 1999.

[27] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave N-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.

[28] P. Mandi, "Estimation and control in Markov chains," *Advances in Applied Probability*, vol. 6, pp. 40–60, 1974.