Dynamic Chinese Restaurant Game in Cognitive Radio Networks

Chunxiao Jiang*†, Yan Chen*, Yu-Han Yang*, Chih-Yu Wang*‡, and K. J. Ray Liu*

*Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA
†Department of Electronic Engineering, Tsinghua University, Beijing 100084, P. R. China
‡Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan
E-mail:{jcx, yan, yhyang}@umd.edu, tomkywang@gmail.com, kjrliu@umd.edu

***Abstract*—In a cognitive radio network with mobility, secondary users can arrive at and leave the primary users' licensed networks at any time. After arrival, secondary users are confronted with channel access under the uncertain primary channel state. On one hand, they have to estimate the channel state, i.e., the primary users' activities, through performing spectrum sensing and learning from other secondary users' sensing results. On the other hand, they need to predict subsequent secondary users' access decisions to avoid competition when accessing the "spectrum hole". In this paper, we propose a Dynamic Chinese Restaurant Game to study such a learning and decision making problem in cognitive radio networks. We introduce a Bayesian learning based method for secondary users to learn the channel state and propose a Multi-dimensional Markov Decision Process based approach for secondary users to make optimal channel access decisions. Finally, we conduct simulations to verify the effectiveness and efficiency of the proposed scheme.**

***Index Terms*—Chinese Restaurant Game, Bayesian Learning, Markov Decision Process, Cognitive Radio, Game Theory.**

I. Introduction

With the emerging of various wireless applications, available electromagnetic radio spectrums are becoming more and more crowded. The traditional static spectrum allocation policy results in a large portion of the assigned spectrum being under utilized [1]. Recently, dynamic spectrum access in cognitive radio networks has shown great potential to improve the spectrum utilization efficiency. In a cognitive radio network, Secondary Users (SUs) can opportunistically utilize the Primary User's (PU's) licensed spectrum bands without harmful interference to the PU [2].

Two essential issues of dynamic spectrum access are spectrum sensing and channel access. Since SUs are uncertain about the primary channel state, i.e., whether the PU is absent, they need to estimate the channel state through performing spectrum sensing [3]. In order to counter the channel fading and shadowing problem, cooperative spectrum sensing technology was proposed recently, in which SUs share their spectrum sensing results to improve the sensing performance [4]. After channel sensing, SUs access one vacant primary channel for data transmission. When choosing a vacant channel, each SU should not only consider the immediate utility, but also take into account the utility in the future since the more subsequent SUs access the same channel, the less throughput can be obtained by each SU. Such a phenomenon is known as negative network externality [5], i.e., the negative influence of other users' behaviors on one user's reward, due to which users tend to avoid making same decisions with others to maximize their own payoffs. However, traditional cooperative sensing schemes simply combine all SUs' sensing results while ignoring the structure of sequential decision making [4], especially in a dynamic scenario where the primary channel state is time-varying and SUs arrive and leave stochastically. Moreover, the negative network externality has not been considered in the previous channel access methods [6]. Therefore, how SUs learn the uncertain primary channel state sequentially and make best channel selections by taking into account the negative network externality are challenging issues in cognitive radio networks.

In our previous work [7], we proposed a new game, called Chinese Restaurant Game, to study how users in a social network make decisions when being confronted with uncertain network state [8][9]. Such a Chinese Restaurant Game is originated from the well-known Chinese Restaurant Process [10], which is widely adopted in non-parametric Bayesian statistics in machine learning to model the distribution which is not parametric but stochastic. In Chinese Restaurant Game, there are finite tables with different sizes and finite customers sequentially requesting tables for meal. Since customers do not know the exact size of each table, they have to learn the table sizes according to some external information. Moreover, when requesting one table, each customer should consider the subsequent customers' decisions due to the limited dining space in each table, i.e., the negative network externality. In [11] and [12], the applications of Chinese Restaurant Game in various research fields are also discussed.

The channel sensing and access problems in cognitive radio networks can be ideally modeled as a Chinese Restaurant Game, where the tables are the primary channels, and customers are SUs who are seeking vacant channels. Moreover, how a SU learns the PU's activities can be regarded as how a customer learns the restaurant state, and how a SU chooses a channel to access can be formulated as how a customer selects a table. One assumption in the Chinese Restaurant Game [7] is the fixed population setting, i.e., there is a finite number of customers choosing the tables sequentially. However, in cognitive radio networks, SUs may arrive and leave at any

978-1-4673-5946-7/13/$31.00 ©2013 IEEE

time, which results in a dynamic population setting. In such a case, a SU's utility will change from time to time due to a dynamic number of SUs in each channel.

Considering these challenges, in this paper, we extend the Chinese Restaurant Game in [7] to a dynamic population setting and propose a Dynamic Chinese Restaurant Game for the cognitive radio networks. In this Dynamic Chinese Restaurant Game, we consider the scenario that SUs arrive at and leave the primary network by Poisson process. Each new coming SU not only learns the channel state according to the information received and revealed by former SUs, but also predicts the subsequent SUs' decisions to maximize the utility. We introduce a channel state learning method based on Bayesian learning rule, where each SU constructs his/her own belief on the channel state according to his/her own signal and the former SU's belief information. Moreover, we formulate the channel access problem as a Multi-dimensional Markov Decision Process (M-MDP) and design a modified value iteration algorithm to find the best strategies. We prove theoretically that there is a threshold structure in the optimal strategy profile for the two primary channel scenario. For multiple primary channel scenario, we propose a fast algorithm with much lower computational complexity while achieving comparable performance. Finally, we conduct simulations to verify the effectiveness and efficiency of the proposed Dynamic Chinese Restaurant Game theoretic scheme.

The rest of this paper is organized as follows. Firstly, our system model is described in Section II. Then, we discuss how SUs learn the primary channel state using Bayesian learning rule in Section III. We introduce an M-MDP model to solve the channel access problem in Section VI and a modified value iteration algorithm in Section V. Finally, we show simulation results in Section VI and draw conclusions in Section VII.

II. System Model

A. Network Entity

We consider a primary network with N independent primary channels, and each channel can maximally host L users, as shown in Fig. 1. The PU has priority to occupy the channels at any time, while SUs are allowed to access the channel under the condition that the PU's communication QoS is guaranteed [6]. We denote the primary channel state as $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_N\}$ (all the subscripts mean the channel number index in the paper), where $\theta_i \in \{\mathcal{H}_0, \mathcal{H}_1\}$ denotes the state of channel i, $\mathcal{H}_0$ means the PU is absent and $\mathcal{H}_1$ means the PU is present. Note that the channel state $\theta_i(t)$ is time-varying since the PU may appear at the primary channel at any time.

For the secondary network, we assume SUs arrive and depart by Poisson process with rate λ and μ, respectively. All SUs can independently perform spectrum sensing using energy detection [1]. For each SU, his/her action set is $\mathcal{A} = \{1, 2, ..., N\}$, i.e., choosing one channel from all N channels. Let us define the grouping state when the jth SU arrives, $\boldsymbol{G}^j = (g_1^j, g_2^j, ..., g_N^j)$ (all the superscripts mean the SU index in the paper), where $g_i^j \in [0, L]$ stands for the number of SUs in channel i. Assuming that the jth SU finally chooses channel i to access, his/her utility function can be given by $U\big(\theta_i(t), g_i^j(t)\big)$, where $\theta_i(t)$ denotes the state of channel i at time t and $g_i^j(t)$ is the number of SUs choosing channel i during the jth SU's access time in channel i. Note that the utility function is a decreasing function in terms of $g_i^j(t)$, which can be regarded as the characteristic of negative network externality since the more subsequent SUs join channel i, the less utility the jth SU can achieve. In the following analysis, the time index (t) is omitted.

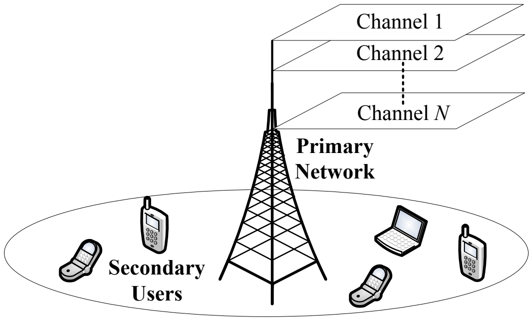

Fig. 1. System model of the cognitive radio network.

As discussed above, the channel state $\boldsymbol{\theta}$ is changing with time. For new arriving SUs, they may not know the exact state of each channel θ_i. Nevertheless, SUs can estimate the state through channel sensing and former SU's sensing result. Therefore, we assume that all SUs know the common prior distribution of the state θ_i for each channel, which is denoted as $\boldsymbol{b}^0 = \{b_i^0 | b_i^0 = \Pr(\theta_i = \mathcal{H}_0), \forall i \in 1, 2, ..., N\}$. Moreover, each SU can receive a signal $\boldsymbol{s}^j = \{s_i^j, \forall i \in 1, 2, .., N\}$ through spectrum sensing, where $s_i^j = 1$ if the jth SU detects some activity on channel i and $s_i^j = 0$ if no activity is detected on channel i. In such a case, the detection and false-alarm probability of channel i can be expressed as $P_i^d = \Pr(s_i = 1 | \theta_i = \mathcal{H}_1)$ and $P_i^f = \Pr(s_i = 1 | \theta_i = \mathcal{H}_0)$, which are considered as common priors for all SUs. Furthermore, we assume that there is a log-file in the server of the secondary network, which records each SU's channel belief and channel selection result. Through querying this log-file, the new coming SU can obtain current grouping state information, i.e., the number of SUs in each channel, as well as the former SU's belief on the channel state.

B. ON-OFF Primary Channel Model

For the PU's behavior in the primary channel, we model it as a general alternating ON-OFF renewal process, where the ON state means the PU is present and the OFF state means the PU is absent. This general ON-OFF switch model can be applied in the scenario when SUs have no knowledge about the PU's communication mechanism [13]. Let T_{ON} and T_{OFF} denote the length of the ON state and OFF state, respectively. According to different types of the primary services (e.g., digital TV broadcasting or cellular communication), T_{ON} and T_{OFF} statistically satisfy different types of distributions. Here we assume that T_{ON} and T_{OFF} are independent and satisfy exponential distributions with parameter r_1 and r_0, respectively, denoted by $f_{\text{ON}}(t)$ and $f_{\text{OFF}}(t)$ as follows:

$$\begin{cases} T_{\text{ON}} \sim f_{\text{ON}}(t) = \frac{1}{r_1}e^{-t/r_1}, \\ T_{\text{OFF}} \sim f_{\text{OFF}}(t) = \frac{1}{r_0}e^{-t/r_0}. \end{cases} \tag{1}$$

In such a case, the expected lengths of the ON state and OFF state are r_1 and r_0 accordingly. These two parameters r_1 and r_0 can be effectively estimated by a maximum likelihood estimator [14]. Such an ON-OFF behavior of the PU is a combination of two Poisson process, which is a renewal process [15]. The renewal interval is $T_p = T_{\text{ON}} + T_{\text{OFF}}$ and the distribution of T_p, denoted by $f_p(t)$, is

$$f_p(t) = f_{\text{ON}}(t) * f_{\text{OFF}}(t). \tag{2}$$

III. Bayesian Learning for the Channel State

In this section, we discuss how a SU estimates the channel state according to his/her own sensing results and the former SU's beliefs. Here, we first introduce the concept of belief to describe the SU's uncertainty about the channel state. The belief b_i^j denotes the jth SU's belief on the state of channel i, θ_i^j. It is assumed that each SU reveals his/her beliefs after making the channel selection. In such a case, the jth SU's belief on channel i is learned from former SU's belief b_i^{j-1}, his/her own signal s_i^j, which can be defined as

$$\boldsymbol{b}^j = \{b_i^j | b_i^j = \Pr(\theta_i^j = \mathcal{H}_0 | b_i^{j-1}, s_i^j), \forall i \in 1, 2, ...N\}. \tag{3}$$

From the definition above, we can see that the belief $b_i^j \in [0,1]$ is a continuous parameter. In a practical system, it is impossible for a SU to reveal his/her continuous belief using infinite data bits. Therefore, we quantize the continuous belief into M belief levels $\{\mathbb{B}_1, \mathbb{B}_2, ..., \mathbb{B}_M\}$, which means that if we have $b_i^j \in [\frac{k-1}{M}, \frac{k}{M}]$, then $B_i^j = \mathbb{B}_k$. Since each SU can only reveal and receive the quantized belief, the former SU's quantized belief $\boldsymbol{B}^{j-1}$ is first mapped into the continuous belief $\widehat{\boldsymbol{b}}^{j-1}$ according to the rule that if $B_i^{j-1} = \mathbb{B}_k$ then $\hat{b}_i^{j-1} = \frac{1}{2}\big(\frac{k-1}{M} + \frac{k}{M}\big)$. Note that the mapped continuous belief $\hat{b}_i^{j-1}$ here is not the former SU's real continuous belief b_i^{j-1}. Then, $\widehat{\boldsymbol{b}}^{j-1}$ is combined with the signal $\boldsymbol{s}^j$ to calculate the continuous belief $\boldsymbol{b}^j$. Finally, $\boldsymbol{b}^j$ is quantized into the belief $\boldsymbol{B}^j$. Since all primary channels are assumed to be independent, the learning processes of these channels are also independent. Thus, the learning process for the jth SU on channel i can be summarized as $B_i^{j-1} \xrightarrow{C} \hat{b}_i^{j-1} \xrightarrow{s_i^j} b_i^j \xrightarrow{Q} B_i^j$.

In the learning process, the most important step is how to calculate current belief b_i^j according to current signal s_i^j and the former SU's belief $\hat{b}_i^{j-1}$, which is a classical social learning problem. Based on the approaches to belief formation, social learning can be classified as Bayesian learning [16] and non-Bayesian learning [17]. Bayesian learning refers that rational individuals use Bayes' rule to form the best estimation of the unknown parameters, such as the channel state in our model, while non-Bayesian learning requires individuals to follow some predefined rules to update their beliefs, which inevitably limits the rational users' optimal decision making. Since SUs are assumed to be fully rational, they adopt Bayesian learning to update their beliefs $\boldsymbol{b}^j = \{b_i^j\}$ by

$$b_i^j = \frac{\Pr\big(\theta_i^j = \mathcal{H}_0 | \hat{b}_i^{j-1}\big)\Pr(s_i^j | \theta_i^j = \mathcal{H}_0)}{\sum_{l=0}^{1} \Pr\big(\theta_i^j = \mathcal{H}_l | \hat{b}_i^{j-1}\big)\Pr(s_i^j | \theta_i^j = \mathcal{H}_l)}. \tag{4}$$

As discussed in the system model, the channel state is varying with time. Here, we define the state transition probability as $\Pr(\theta_i^j = \mathcal{H}_0 | \theta_i^{j-1} = \mathcal{H}_0)$, which represents the probability that channel i is currently available when the jth SU arrives given the condition that channel i was available when the $(j-1)$th SU arrived. Similarly, we have $\Pr(\theta_i^j = \mathcal{H}_1 | \theta_i^{j-1} = \mathcal{H}_0)$, $\Pr(\theta_i^j = \mathcal{H}_0 | \theta_i^{j-1} = \mathcal{H}_1)$ and $\Pr(\theta_i^j = \mathcal{H}_1 | \theta_i^{j-1} = \mathcal{H}_1)$. In such a case, the SU can calculate the items $\Pr\big(\theta_i^j = \mathcal{H}_0 | \hat{b}_i^{j-1}\big)$ and $\Pr\big(\theta_i^j = \mathcal{H}_1 | \hat{b}_i^{j-1}\big)$ in (4) as follows:

$$\begin{aligned} \Pr\big(\theta_i^j = \mathcal{H}_0 | b_i^{j-1}\big) = {} & \Pr(\theta_i^j = \mathcal{H}_0 | \theta_i^{j-1} = \mathcal{H}_0) b_i^{j-1} + \\ & \Pr(\theta_i^j = \mathcal{H}_0 | \theta_i^{j-1} = \mathcal{H}_1)(1 - b_i^{j-1}), \quad (5) \\ \Pr\big(\theta_i^j = \mathcal{H}_1 | b_i^{j-1}\big) = {} & \Pr(\theta_i^j = \mathcal{H}_1 | \theta_i^{j-1} = \mathcal{H}_0) b_i^{j-1} + \\ & \Pr(\theta_i^j = \mathcal{H}_1 | \theta_i^{j-1} = \mathcal{H}_1)(1 - b_i^{j-1}). \quad (6) \end{aligned}$$

Since the primary channel is modeled as an ON-OFF process, the channel state transition probability depends on the time interval between the $(j-1)$th and jth SUs' arrival time, t^j. Note that the t^j can be directly obtained from the log-file in the server. For notation simplicity, in the rest of this paper, we use $P_{00}(t^j)$, $P_{01}(t^j)$, $P_{10}(t^j)$ and $P_{11}(t^j)$ to denote $\Pr(\theta_i^j = \mathcal{H}_0 | \theta_i^{j-1} = \mathcal{H}_0)$, $\Pr(\theta_i^j = \mathcal{H}_1 | \theta_i^{j-1} = \mathcal{H}_0)$, $\Pr(\theta_i^j = \mathcal{H}_0 | \theta_i^{j-1} = \mathcal{H}_1)$ and $\Pr(\theta_i^j = \mathcal{H}_1 | \theta_i^{j-1} = \mathcal{H}_1)$, respectively, where $P_{01}(t^j) = 1 - P_{00}(t^j)$ and $P_{11}(t^j) = 1 - P_{10}(t^j)$.

We can derive the close-form expression for $P_{01}(t^j)$ using the renewal theory as follow [18]

$$P_{01}(t^j) = \frac{r_1}{r_0 + r_1}\Big(1 - e^{-\frac{r_0 + r_1}{r_0 r_1} t^j}\Big). \tag{7}$$

Thus, we can have $P_{00}(t^j)$ as

$$P_{00}(t^j) = 1 - P_{01}(t^j) = \frac{r_1}{r_0 + r_1}\Big(\frac{r_0}{r_1} + e^{-\frac{r_0 + r_1}{r_0 r_1} t^j}\Big). \tag{8}$$

Similarly, the close-form expression for $P_{11}(t^j)$ can also be obtained by solving the following renewal equation

$$P_{11}(t) = r_1 f_{\text{ON}}(t) + \int_0^t P_{11}(t - w) f_p(w) dw, \tag{9}$$

where $f_{\text{ON}}(t)$ is the probability density function (*p.d.f*) of the ON state's length given in (1) and $f_p(t)$ is the *p.d.f* of the PU's renewal interval given in (2).

By solving (9), we can obtain the close-form expression for $P_{11}(t^j)$ given by

$$P_{11}(t^j) = \frac{r_0}{r_0 + r_1}\Big(\frac{r_1}{r_0} + e^{-\frac{r_0 + r_1}{r_0 r_1} t^j}\Big). \tag{10}$$

Then, we can have $P_{10}(t_i)$ as

$$P_{10}(t^j) = 1 - P_{11}(t^j) = \frac{r_0}{r_0 + r_1}\Big(1 - e^{-\frac{r_0 + r_1}{r_0 r_1} t^j}\Big). \tag{11}$$

By substituting (5-6), (7-8) and (10-11) into (4), we can calculate the jth SU's belief b_i^j with the corresponding sensing results $s_i^j = 1$ and $s_i^j = 0$ in (12-13) below. In the following, we denote (12) as $b_i^j|_{s_i^j=1} = \phi(\hat{b}_i^{j-1}, t_i, s_i^j = 1)$, and denote (13) as $b_i^j|_{s_i^j=0} = \phi(\hat{b}_i^{j-1}, t_i, s_i^j = 0)$ for simplicity.

IV. Multi-dimensional Markov Decision Process Based Channel Access

In this section, we investigate the channel access game by modeling it as a Markov Decision Process (MDP) problem [19]. In this game, each SU selects a channel to access, with the objective of maximizing his/her own expected utility during the access time. To achieve this goal, rational SUs not only need to consider the immediate utility, but also need to consider the following SUs' decisions. In our model, SUs arrive by Poisson process and make the channel selection sequentially. When making the decision, one SU is only confronted with current grouping information $\boldsymbol{G}^j$ and belief information $\boldsymbol{B}^j$. In order to take into account the SU's expected utility in the future, we use Bellman equation to formulate the SU's utility and use MDP model to formulate this channel selection problem. In traditional MDP problem, a player can adjust his/her decision when the system state changes. However, in our system, after choosing a certain primary channel, a SU cannot adjust his/her channel selection even if the system state has already changed. Therefore, traditional MDP cannot be directly applied here. To solve this problem, we propose a Multi-dimensional MDP (M-MDP) model, and a modified value iteration method to derive the best response (strategy) for each SU.

A. System State

To construct the MDP model, we first define the system state and verify the Markov property of the state transition. Let the quantized belief $\boldsymbol{B} = (B_1, B_2, ..., B_N) \in [1, M]^N$ be the belief state. Thus, we can define the system state S as the belief state $\boldsymbol{B}$ with the grouping state $\boldsymbol{G} = (g_1, g_2, ..., g_N) \in [0, L]^N$, i.e., $S = (\boldsymbol{B}, \boldsymbol{G})$, where the finite state space is $\mathcal{X} = \Big([1, M]^N \times [0, L]^N\Big)$. When the jth SU arrives, the system state he/she encounters is $S^j = (\boldsymbol{B}^j, \boldsymbol{G}^j)$. In such a case, with multiple SUs arriving sequentially, the system states at different arrival time $\{S^1, S^2, ...S^j, ...\}$ form a stochastic process. In our learning rule, only the $(j-1)$th SU's belief is used to update the jth SU's belief. Therefore, $\boldsymbol{B}^j$ depends only on $\boldsymbol{B}^{j-1}$. Moreover, since SUs arrive by Poisson process, the grouping state $\boldsymbol{G}^j$ is also memoryless. In such a case, we can verify that $\{S^1, S^2, ...S^j, ...\}$ is a Markov process.

B. Belief State Transitions

Note that a SU's belief transition is independent with his/her action, and is only related to the channel state, as well as the Bayesian learning rule. Here, we define the belief state transition probability as $P\big(\boldsymbol{B}^j|\boldsymbol{B}^{j-1}\big)$. Since all channels are independent with each other, we have

$$P\big(\boldsymbol{B}^j\big|\boldsymbol{B}^{j-1}\big) = \prod_{i=1}^{N} P\big(B_i^j|B_i^{j-1}\big), \tag{14}$$

where $P\big(B_i^j|B_i^{j-1}\big)$ is the belief state transition probability of channel i. In such a case, there is an $M \times M$ belief state transition matrix for each channel, which can be derived according to the Bayesian learning rule. To find $P\big(B_i^j = \mathbb{B}_q|B_i^{j-1} = \mathbb{B}_p\big)$, with the quantized belief $B_i^{j-1} = \mathbb{B}_p$, we can calculate the corresponding continuous belief $\hat{b}_i^{j-1} = \frac{1}{2}\big(\frac{p-1}{M} + \frac{p}{M}\big)$. Then, with $B_i^j = \mathbb{B}_q$, we can have the value interval of $b_i^j = [\frac{q-1}{M}, \frac{q}{M}]$. Thus, the belief state transition probability can be computed by

$$P\big(B_i^j = \mathbb{B}_q|B_i^{j-1} = \mathbb{B}_p\big) = \int_{\frac{q-1}{M}}^{\frac{q}{M}} P(b_i^j|\hat{b}_i^{j-1})db_i^j. \tag{15}$$

According to (12) and (13), we have $b_i^j = \phi\Big(\hat{b}_i^{j-1} = \frac{1}{2}\big(\frac{p-1}{M} + \frac{p}{M}\big), t^j, s_i^j\Big)$. Therefore, the belief state transition probability can be re-written by (16) below, where the second equality follows the assumption that the arrival interval of two SUs t^j obeys exponential distribution with parameter

$$b_i^j|_{s_i^j=1} = \frac{\Big(r_0 e^{\frac{r_0+r_1}{r_0 r_1}t^j} - r_0 + (r_1+r_0)\hat{b}_i^{j-1}\Big)P_i^f}{\Big(r_0 e^{\frac{r_0+r_1}{r_0 r_1}t^j} - r_0 + (r_1+r_0)\hat{b}_i^{j-1}\Big)P_i^f + \Big(r_1 e^{\frac{r_0+r_1}{r_0 r_1}t^j} + r_0 - (r_1+r_0)\hat{b}_i^{j-1}\Big)P_i^d}, \tag{12}$$

$$b_i^j|_{s_i^j=0} = \frac{\Big(r_0 e^{\frac{r_0+r_1}{r_0 r_1}t^j} - r_0 + (r_1+r_0)\hat{b}_i^{j-1}\Big)(1-P_i^f)}{\Big(r_0 e^{\frac{r_0+r_1}{r_0 r_1}t^j} - r_0 + (r_1+r_0)\hat{b}_i^{j-1}\Big)(1-P_i^f) + \Big(r_1 e^{\frac{r_0+r_1}{r_0 r_1}t^j} + r_0 - (r_1+r_0)\hat{b}_i^{j-1}\Big)(1-P_i^d)}. \tag{13}$$

$$\begin{aligned}
\Pr(B_i^j = \mathbb{B}_q|B_i^{j-1} = \mathbb{B}_p) \;&=\; \iint_{\frac{q-1}{M} \leq \phi\left(\hat{b}_i^{j-1} = \frac{1}{2}(\frac{p-1}{M} + \frac{p}{M}), t^j, s_i^j\right) \leq \frac{q}{M}} \Pr\big(t^j, s_i^j|\hat{b}_i^{j-1}\big)dt^j ds^j, \\
&=\; \int_{\frac{q-1}{M} \leq \phi\left(\hat{b}_i^{j-1} = \frac{1}{2}(\frac{p-1}{M} + \frac{p}{M}), t^j, s_i^j = 0\right) \leq \frac{q}{M}} \lambda e^{-\lambda t^j} \Pr(s_i^j = 0|\hat{b}_i^{j-1})dt^j \\
&\quad+ \int_{\frac{q-1}{M} \leq \phi\left(\hat{b}_i^{j-1} = \frac{1}{2}(\frac{p-1}{M} + \frac{p}{M}), t^j, s_i^j = 1\right) \leq \frac{q}{M}} \lambda e^{-\lambda t^j} \Pr(s_i^j = 1|\hat{b}_i^{j-1})dt^j.
\end{aligned} \tag{16}$$

λ and is independent with the belief. To calculate (16), we need to derive $\Pr(s_i^j|\hat{b}_i^{j-1})$, which represents the distribution of the jth SU's received signal when given the $(j-1)$th SU's belief. Note that given current channel state θ_i^j, signal s_i^j is independent with belief $\hat{b}_i^{j-1}$. Thus, $\Pr(s_i^j|\hat{b}_i^{j-1})$ can be calculated as follows:

$$\begin{aligned}\Pr(s_i^j|\hat{b}_i^{j-1}) &= \Pr(s_i^j, \theta_i^j = \mathcal{H}_0|\hat{b}_i^{j-1}) + \Pr(s_i^j, \theta_i^j = \mathcal{H}_1|\hat{b}_i^{j-1}) \\ &= \Pr(s_i^j|\theta_i^j = \mathcal{H}_0)\Pr(\theta_i^j = \mathcal{H}_0|\hat{b}_i^{j-1}) + \\ &\quad \Pr(s_i^j|\theta_i^j = \mathcal{H}_1)\Pr(\theta_i^j = \mathcal{H}_1|\hat{b}_i^{j-1}). \qquad (17)\end{aligned}$$

Moreover, given the previous channel state θ_i^{j-1}, current state θ_i^j is also independent with the former SU's belief $\hat{b}_i^{j-1}$. Thus, $\Pr(\theta_i^j = \mathcal{H}_0|\hat{b}_i^{j-1})$ in (17) can be obtained as follows:

$$\begin{aligned}\Pr(\theta_i^j = \mathcal{H}_0|\hat{b}_i^{j-1}) &= \Pr(\theta_i^j = \mathcal{H}_0, \theta_i^{j-1} = \mathcal{H}_0|\hat{b}_i^{j-1}) + \\ &\quad \Pr(\theta_i^j = \mathcal{H}_0, \theta_i^{j-1} = \mathcal{H}_1|\hat{b}_i^{j-1}) \\ &= \Pr(\theta_i^j = \mathcal{H}_0|\theta_i^{j-1} = \mathcal{H}_0)\hat{b}_i^{j-1} + \\ &\quad \Pr(\theta_i^j = \mathcal{H}_0|\theta_i^{j-1} = \mathcal{H}_1)(1 - \hat{b}_i^{j-1}), \\ &= P_{00}(t^j)\hat{b}_i^{j-1} + P_{10}(t^j)(1 - \hat{b}_i^{j-1}). \qquad (18)\end{aligned}$$

Similarly, for $\Pr(\theta_i^j = \mathcal{H}_1|\hat{b}_i^{j-1})$, we have

$$\Pr(\theta_i^j = \mathcal{H}_1|\hat{b}_i^{j-1}) = P_{01}(t^j)\hat{b}_i^{j-1} + P_{11}(t^j)(1 - \hat{b}_i^{j-1}). \quad (19)$$

By substituting (18-19) into (17), the conditional distribution of the signal can be obtained as (20-21) below, with which we can calculate the transition probability matrix using (16).

C. Actions and System State Transitions

The finite action set for SUs is the N channel set, i.e., $\mathcal{A} = \{1, 2, ..., N\}$. Let $a \in \mathcal{A}$ denote a new SU's action under the system state $S = (\boldsymbol{B}, \boldsymbol{G})$. Let $P\big(S' = (\boldsymbol{B}', \boldsymbol{G}')\big|S = (\boldsymbol{B}, \boldsymbol{G}), a\big)$ denote the probability that action a in state S will lead to state S'. Since the SU's belief transition is independent with his/her action, we have

$$P\big(S'{=}(\boldsymbol{B}', \boldsymbol{G}')|S{=}(\boldsymbol{B}, \boldsymbol{G}), a\big){=}P\big(\boldsymbol{B}'|\boldsymbol{B}\big)P\big(\boldsymbol{G}'|\boldsymbol{G}, a\big), \quad (22)$$

where $P\big(\boldsymbol{G}'|\boldsymbol{G}, a\big)$ is the system grouping state transition probability. Suppose that $\boldsymbol{G} = (g_1, g_2, ..., g_N)$, if a new SU arrives and accesses channel i, i.e., $a = i$, we have the system state transition probabilities in (23-24) below, where λ is the arrival rate. If no SU arrives, but some SU leaves the channel at state G, we have the remaining system state transition probabilities in (25-26) below, where μ is SUs' departure rate. Note that λ and μ are normalized such that $\lambda + NL\mu \leq 1$. In such a case, the system state transition probabilities $P(S'|S)$ form an $\big(M(L+1)\big)^N \times \big(M(L+1)\big)^N$ state transition matrix when given action a.

D. Expected Utility

The immediate utility of SUs in channel i at state S is

$$U_i(S) = \hat{b}_i \cdot R_i\big(g_i\big), \qquad (27)$$

where $\hat{b}_i$ is the continuation mapping of B_i and R_i is a decreasing function with respect to the number of SUs in channel i, g_i. In general, each SU will access the selected channel for a period of time, during which the system state may change. Therefore, when making the channel selection, SUs should not only consider the immediate utility, but also take into account the future utility. Here, we define a SU's expected utility in channel i, $V_i(S)$, based on Bellman equation [19] as follow

$$V_i(S) = U_i(S) + (1 - \mu) \sum_{S' \in \mathcal{X}} P_i(S'|S)V_i(S'), \qquad (28)$$

where $(1 - \mu)$ is the discount factor, which can be regarded as the probability that the SU keeps staying at the selected channel since μ is the departure probability, and $P_i(S'|S)$ is the state transition probability defined as

$$P_i\big(S'{=}(\boldsymbol{B}', \boldsymbol{G}')|S{=}(\boldsymbol{B}, \boldsymbol{G})\big){=}P\big(\boldsymbol{B}'|\boldsymbol{B}\big) \cdot P_i\big(\boldsymbol{G}'|\boldsymbol{G}\big), \quad (29)$$

where $P\big(\boldsymbol{B}'|\boldsymbol{B}\big)$ is the belief state transition probability, and $P_i\big(\boldsymbol{G}'|\boldsymbol{G}\big)$ is the grouping state transition probability conditioned on that SUs in channel i still stay in channel i in the next state S', which is different with $P\big(\boldsymbol{G}'|\boldsymbol{G}\big)$ in (23-26). Note that $P_i\big(\boldsymbol{G}'|\boldsymbol{G}\big)$ is closely related to the new arriving SU's action. Suppose that the new SU's action $a_S = k$, i.e., accessing channel k at state S, we have the state transition probability in (30) below. For the leaving transition probability, since we have considered the discount factor $(1 - \mu)$ in the future utility, i.e., the SU will not leave the channel, thus we have state transition probabilities in (31-33) below, where the item $(g_i - 1)$ is because the grouping in channel i, g_i, already

$$\Pr(s_i^j = 0|\hat{b}_i^{j-1}) = \big(1 - P_i^f\big)\Big(P_{00}(t^j)\hat{b}_i^{j-1} + P_{10}(t^j)(1 - \hat{b}_i^{j-1})\Big) + \big(1 - P_i^d\big)\Big(P_{01}(t^j)\hat{b}_i^{j-1} + P_{11}(t^j)(1 - \hat{b}_i^{j-1})\Big), \qquad (20)$$

$$\Pr(s_i^j = 1|\hat{b}_i^{j-1}) = P_i^f\Big(P_{00}(t^j)\hat{b}_i^{j-1} + P_{10}(t^j)(1 - \hat{b}_i^{j-1})\Big) + P_i^d\Big(P_{01}(t^j)\hat{b}_i^{j-1} + P_{11}(t^j)(1 - \hat{b}_i^{j-1})\Big). \qquad (21)$$

$$P\big(\boldsymbol{G}' = (g_1, g_2, ..., g_i + 1, ..., g_N)|\boldsymbol{G} = (g_1, g_2, ..., g_i, ..., g_N), a = i\big) = \lambda, \qquad (23)$$

$$P\big(\boldsymbol{G}' \neq (g_1, g_2, ..., g_i + 1, ..., g_N)|\boldsymbol{G} = (g_1, g_2, ..., g_i, ..., g_N), a = i\big) = 0, \qquad (24)$$

$$P\big(\boldsymbol{G}' = (g_1, g_2, ..., g_i - 1, ..., g_N)|\boldsymbol{G} = (g_1, g_2, ..., g_i, ..., g_N)\big) = g_i\mu, \big(\forall i \in [1, N]\big), \qquad (25)$$

$$P\big(\boldsymbol{G}' = \boldsymbol{G}|\boldsymbol{G} = (g_1, g_2, ..., g_i, ..., g_N)\big) = 1 - \lambda - \sum_{i=1}^{N} g_i\mu. \qquad (26)$$

includes this SU who will not leave the channel at state S'. In such a case, we can have an M-dimensional expected utility function set in (34), where $\mathbf{P}_i(S'|S) = \big[P_i(S'|S)|\forall S' \in \mathcal{X}\big]$ and $\mathbf{V}_i(S'|S) = \big[V_i(S'|S)|\forall S' \in \mathcal{X}\big]$.

E. Best Strategy

The strategy profile $\pi = \{a_S|\forall S \in \mathcal{X}\}$ is a mapping from the state space to the action space, i.e., $\pi : \mathcal{X} \to \mathcal{A}$. Due to the selfish nature, each SU will choose the best strategy to maximize his/her own expected utility. Suppose that one SU arrives at the primary network with system state $S = \big(\boldsymbol{B}, \boldsymbol{G} = (g_1, g_2, ..., g_i, ..., g_N)\big)$, his/her best strategy can be defined as

$$a_S = \underset{i\in[1,N]}{\operatorname{argmax}} \Big\{ V_i\big(\boldsymbol{B}, \boldsymbol{G} = (g_1, ..., g_i + 1, ..., g_N)\big)\Big\}. \tag{35}$$

Since the strategy profile satisfying (34) and (35), denoted by $\pi^\star$, maximizes every arriving SU's utility, $\pi^\star$ is a Nash equilibrium of the proposed game.

V. Modified Value Iteration Algorithm

As discussed at the beginning of Section IV, although the channel access problem can be modeled as an MDP problem, it is different from the traditional MDP problem that a SU cannot adjust action even the system state changes. In traditional MDP problem, there is only one Bellman equation associated with each system state, and the optimal strategy is directly obtained by optimizing the Bellman equation. In our Multi-dimensional MDP problem, there is a set of Bellman equations as shown in (34) and the optimal strategy profile should satisfy (34) and (35) simultaneously. Therefore, the traditional dynamic programming method in [19] cannot be directly applied. To solve this problem, we design a modified value iteration algorithm.

Given an initial strategy profile π, the conditional state transition probability $\mathbf{P}_i(S'|S)$ can be calculated by (29-33), and thus the conditional expected utility $\mathbf{V}_i(S)$ can be found by (34). Then, with $\mathbf{V}_i(S)$, the strategy profile π can be updated again using (35). Through such an iterative way, we can finally find the optimal strategy $\pi^\star$. In Algorithm 1, we summarize the proposed modified value iteration algorithm for the Multi-dimensional MDP problem.

Algorithm 1 Modified Value Iteration Algorithm for Multi-dimensional MDP Problem.

```
• Given tolerance $\eta_1$ and $\eta_2$, set $\epsilon_1$ and $\epsilon_2$.
• Initialize $\{V_i^{(0)}(S) = 0, \forall S \in \mathcal{X}\}$ and randomize
   $\pi = \{a_S, \forall S \in \mathcal{X}\}$.
while $\epsilon_1 > \eta_1$ or $\epsilon_2 > \eta_2$ do
    for all $S \in \mathcal{X}$ do
        • Calculate transition probability $\mathbf{P}_i(S'|S)$,
          $\forall i \in [1, N]$ using $\pi$ and (29-33).
        • Update utility function $\mathbf{V}_i^{(n+1)}(S), \forall i \in [1, N]$
          using (34).
    end for
    for all $S \in \mathcal{X}$ do
        • Update $\pi^\star = \{a_S\}$ using (35).
    end for
    • Update the parameter $\epsilon_1$ by $\epsilon_1 = \|\pi - \pi^\star\|_2$.
    • Update the parameter $\epsilon_2$ by $\epsilon_2 = \big\|\mathbf{V}_i^{(n+1)}(S) -$
       $\mathbf{V}_i^{(n)}(S)\big\|_2$.
    • Update the strategy file $\pi = \pi^\star$.
end while
• The optimal strategy profile is $\pi^\star$.
```

A. Channel Access: Two Primary Channels Case

In this subsection, we discuss the case where there are two primary channels. In such a case, the system state $S = (B_1, B_2, g_1, g_2)$, where B_1 and B_2 are beliefs of two channels, g_1 and g_2 are numbers of SUs in two channels. We define SUs' immediate utility in channel i, $U(B_i, g_i)$, as

$$U(B_i, g_i) = \hat{b}_i R(g_i) = \hat{b}_i \log\left(1 + \frac{\text{SNR}}{(g_i - 1)\text{INR} + 1}\right), \tag{36}$$

where $\hat{b}_i$ is the continuous mapping of quantized belief B_i.

According to (34), the expected utility functions of two channels can be written as

$$V_1(S) = U(B_1, g_1) + (1-\mu)\sum_{S'} P_1(S'|S)V_1(S'), \tag{37}$$

$$V_2(S) = U(B_2, g_2) + (1-\mu)\sum_{S'} P_2(S'|S)V_2(S'), \tag{38}$$

$$P_i\big(\boldsymbol{G}' = (g_1, ..., g_k + 1, ..., g_N)|\boldsymbol{G} = (g_1, ..., g_k, ..., g_N)\big) = \lambda, \tag{30}$$

$$P_i\big(\boldsymbol{G}' = (g_1, g_2, ..., g_i - 1, ..., g_N)|\boldsymbol{G} = (g_1, g_2, ..., g_i, ..., g_N)\big) = (g_i - 1)\mu, \tag{31}$$

$$P_i\big(\boldsymbol{G}' = (g_1, g_2, ..., g_{i'\neq i} - 1, ..., g_N)|\boldsymbol{G} = (g_1, g_2, ..., g_{i'\neq i}, ..., g_N)\big) = g_{i'}\mu, \big(\forall i' \in [1, N]\big), \tag{32}$$

$$P_i\big(\boldsymbol{G}' = \boldsymbol{G}|\boldsymbol{G} = (g_1, g_2, ..., g_N)\big) = 1 - \lambda - \left(\sum_{i=1}^{N} g_i - 1\right)\mu. \tag{33}$$

$$\begin{bmatrix} V_1(S) \\ V_2(S) \\ \vdots \\ V_N(S) \end{bmatrix} = \begin{bmatrix} U_1(S) \\ U_2(S) \\ \vdots \\ U_N(S) \end{bmatrix} + (1-\mu) \begin{bmatrix} \mathbf{P}_1(S'|S) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2(S'|S) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}_N(S'|S) \end{bmatrix} \begin{bmatrix} \mathbf{V}_1(S')^T \\ \mathbf{V}_2(S')^T \\ \vdots \\ \mathbf{V}_N(S')^T \end{bmatrix}. \tag{34}$$

where $S = (B_1, B_2, g_1, g_2)$ is the system state, P_1 and P_2 are the state transition probabilities conditioned on the event that SUs stay in the channels they has chosen, which can be calculated according to (29-33).

According to (35), the best strategy a_S for SUs arriving with system state $S = (B_1, B_2, g_1, g_2)$ is as follows:

$$a_S = \begin{cases} 1, \ V_1(B_1, B_2, g_1{+}1, g_2) \geq V_2(B_1, B_2, g_1, g_2{+}1), \\ 2, \ V_1(B_1, B_2, g_1{+}1, g_2) < V_2(B_1, B_2, g_1, g_2{+}1). \end{cases} \quad (39)$$

Thus, with (36-39), we can find the optimal strategy profile $\pi^\star = \{a_S, \forall S \in \mathcal{X}\}$ using the modified value iteration method in Algorithm 1. In the following, we will show that when given the beliefs of two channel, there exists a threshold structure in the optimal strategy profile $\pi^\star$.

Lemma 1: For $g_1 \geq 0$ and $g_2 \geq 1$,

$$V_1(B_1, B_2, g_1, g_2) \geq V_1(B_1, B_2, g_1 + 1, g_2 - 1), \quad (40)$$

$$V_2(B_1, B_2, g_1, g_2) \leq V_2(B_1, B_2, g_1 + 1, g_2 - 1). \quad (41)$$

Proof: Due to page limitation, we show the proof in the supplementary information [20]. ■

Lemma 1 shows that given the beliefs of two channels, V_1 is non-decreasing and V_2 is non-increasing along the line of $g_1 + g_2 = m, \forall m \in \{0, 1, ..., 2L\}$. Based on *Lemma 1*, we will show the threshold structure in the optimal strategy profile $\pi^\star$ by *Theorem 1*.

Theorem 1: For the two-channel case, given the belief state, the optimal strategy profile $\pi^\star = \{a_S\}$ derived from the modified value iteration algorithm has threshold structure.

Proof: Due to page limitation, we show the proof in the supplementary information [20]. ■

Note that the optimal strategy profile $\pi^\star$ can be obtained off-line and the profile can be stored in a table for SUs. We can see that for some spectific belief state, the number of system states is $(L+1)^2$, which means the corresponding strategy file has $(L+1)^2$ strategies. With the proved threshold structure on each line $g_1 + g_2 = m, \forall m \in \{0, 1, 2, ..., 2L\}$, we just need to store the threshold point on each line. In such a case, the storage of the strategy profile can be reduced from $\mathcal{O}(L^2)$ to $\mathcal{O}(2L)$.

B. Channel Access: Multiple Primary Channels Case

In this subsection, we discuss the case where there are multiple primary channels. Although the optimal strategy profile of the multi-channel case can also be obtained using Algorithm 1, the computation complexity grows exponentially in terms of the number of primary channels N. Besides, the storage and retrieval of the strategy profile are also challenging when the number of system states exponentially increases with N. Therefore, it is important to develop a fast algorithm for the multi-channel case.

Suppose the channel number N is even, we can randomly divide these N primary channels into $N/2$ pairs. For each pair, SUs can choose one channel using the threshold strategy derived in the previous subsection. Then, SUs can further divide the selected $N/2$ channels into $N/4$ pairs and so on so forth. In such a case, SUs can finally select one suboptimal channel to access. On the other hand, if the channel number N is odd, the suboptimal channel can be selected by a similar way. With such an iterative dichotomy method, a SU can find one suboptimal primary channel only by $\log N$ steps and the complexity of each step is same with that of the two-channel case. This fast algorithm is summarized in Algorithm 2. In the simulation section, we will compare the performance of this fast algorithm with the optimal algorithm using modified value iteration method.

Algorithm 2 Fast Algorithm for the Multi-channel Case.

```
if $N$ is even then
    while $N > 1$ do
        • Randomly divide the $N$ channels into $N/2$ pairs.
        for all $N/2$ pairs do
            • Select one channel according to Algorithm 1.
        end for
        • $N = N/2$.
    end while
end if
if $N$ is odd then
    while $N > 1$ do
        • Randomly divide the $N$ primary channels into
          $(N-1)/2$ pairs and one channel.
        for all $(N-1)/2$ pairs do
            • Select one channel according to Algorithm 1.
        end for
        • $N = (N-1)/2 + 1$.
    end while
end if
```

VI. Simulation Results

In this section, we conduct simulations to evaluate the performance of proposed scheme in cognitive radio networks. Specifically, we evaluate the performance of channel sensing and access, as well as the interference to the PU.

A. Bayesian Channel Sensing

In this simulation, we evaluate the performance of channel sensing with Bayesian learning. We first generate one primary channel based on the ON-OFF model, and the channel parameters are set to be $r_0 = 55$s and $r_1 = 50$s, respectively. Then, a number of SUs with some arrival rate λ sequentially sense the primary channel and construct their own beliefs by combining the sensing result with the former SU's belief. In Fig. 2, we compare the detection and false-alarm probabilities between channel sensing with Bayesian learning and sensing without learning under different arrival rate λ. Overall, the detection probability is enhanced and false-alarm probability decreases when the Bayesian learning is adopted. Moreover, we can see that with Bayesian learning, the larger the arrival rate λ, the higher detection probability and the lower the false-alarm probability. This is because a larger λ means a shorter arrival interval between two SUs, and thus the former SU's belief information is more useful for current SU's belief construction.

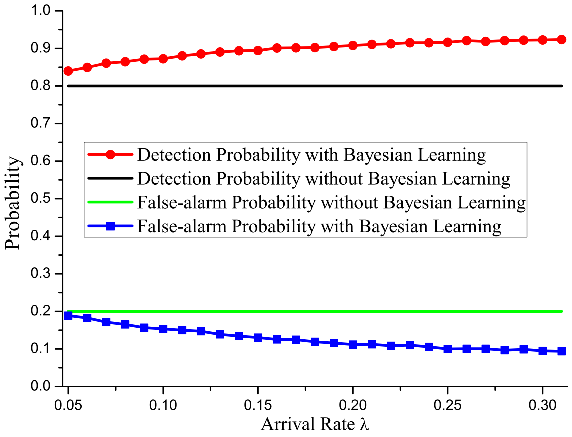

Fig. 2. Detection and false-alarm probability.

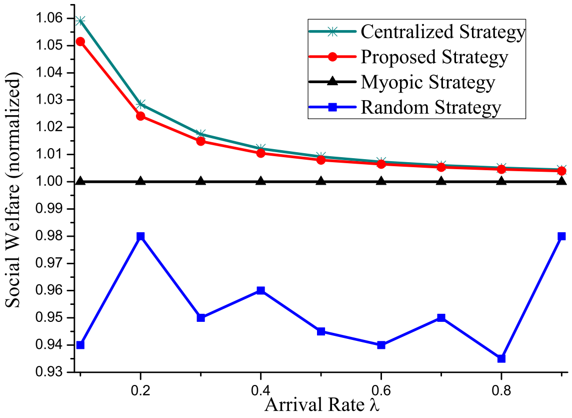

Fig. 3. Social welfare under 2-channel with $M = 2$ and $L = 2$.

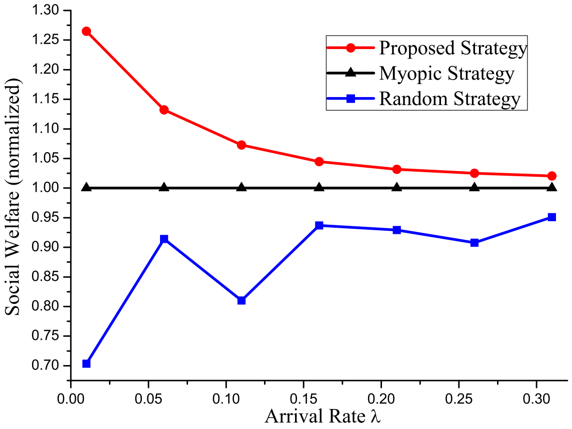

Fig. 4. Social welfare under 2-channel with $M = 5$ and $L = 5$.

B. Channel Access of Two Primary Channel Case

In this subsection, we evaluate the performance of the proposed Multi-dimensional MDP model, as well as the modified value iteration algorithm for the two-channel case. The parameters of the two primary channels are set to be: for channel 1, $r_0 = 55$s and $r_1 = 25$s; for channel 2, $r_0 = 25$s and $r_1 = 55$s, which means channel 1 is statistically better than channel 2. In the simulation, our proposed strategy is compared with centralized strategy, myopic strategy and random strategy in terms of social welfare. We first define the social welfare, W, when given a strategy profile $\pi = \{a_S, \forall S \in \mathcal{X}\}$ as

$$W = \sum_{S \in \mathcal{X}} \sigma^{\pi}(S)\Big(g_1 U(B_1, g_1) + g_2 U(B_2, g_2)\Big), \qquad (42)$$

where $S = (B_1, B_2, g_1, g_2)$ in the two-channel case, and $\sigma^{\pi}(S)$ is the stationary probability of state S. The four strategies we test are defined as follows.

- **Proposed strategy** is obtained by our proposed value iteration algorithm in Algorithm 1.
- **Centralized strategy** is obtained by exhaustively searching all possible $2^{|\mathcal{X}|}$ strategy profiles to maximize the social welfare, i.e., $\pi^c = \arg\max_{\pi} W^{\pi}$, where the superscript c means centralized strategy. We can see that the complexity of finding the centralized strategy is NP-hard.
- **Myopic strategy** is to maximize the immediate utility, i.e., to choose the channel with the largest immediate reward by $\pi^m = \{a_S = \underset{i \in [1,2]}{\text{argmax}}\, U(B_i, g_i), \forall S \in \mathcal{X}\}$, where the superscript m means myopic strategy.
- **Random strategy** is to randomly choose one channel with equal probability 0.5, i.e., $\pi^r = \{a_S = \text{rand}(1, 2), \forall S \in \mathcal{X}\}$, where the superscript r means random strategy.

In the simulation, we use the myopic strategy as the comparison baseline and show the results by normalizing the performance of each strategy by that of the myopic strategy.

In Fig. 3, we evaluate the social welfare performance of different methods. Due to the extremely high complexity of the centralized strategy, we consider the case with 2 belief levels and maximally 2 SUs in each channel, i.e., $M = 2$ and $L = 2$. Note that if $M = 2$ and $L = 3$, there are totally $2^{2^2 \cdot (3+1)^2} = 2^{64}$ possible strategy profiles to verify, which is computational intractable. Therefore, although slightly outperforming our proposed strategy as shown in Fig. 3, the centralized method is not applicable to the time-varying primary channels. Moreover, we also compare the proposed strategy with the myopic and random strategies under the case with $M = 5$ and $L = 5$ in Fig. 4. We can see that the proposed strategy performs the best among all the strategies.

We verify that the proposed strategy is a Nash equilibrium (NE) by simulating a new coming SU's expected utility in Fig. 5. The deviation probability in x-axis stands for the probability that a new coming SU deviates from the proposed strategy or centralized strategy. We can see that when there is no deviation, our proposed strategy performs better than the centralized strategy. Such a phenomenon is because the centralized strategy is to maximize the social welfare and thus sacrifices the new coming SU's expected utility. Moreover, we can see that the expected utility of a new coming SU decreases as the deviation probability increases, which verifies that the proposed strategy is a NE. On the other hand, by deviating from the centralized strategy, a new coming SU can

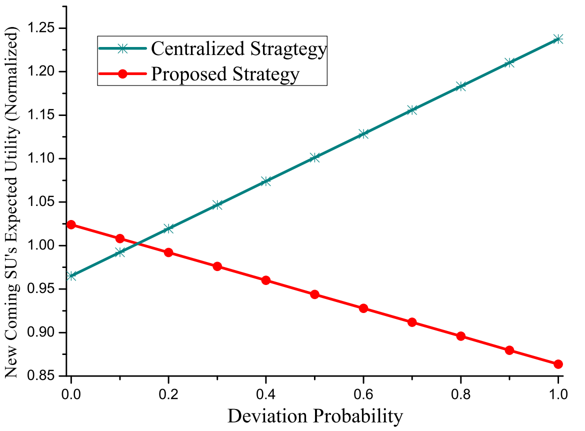

Fig. 5. NE verification under 2-channel with $M = 2$ and $L = 2$.

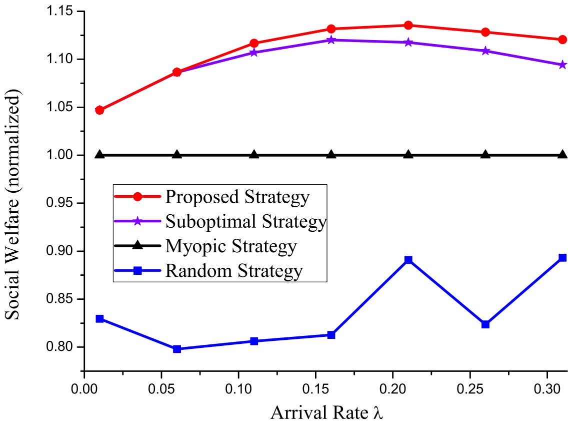

Fig. 6. Social welfare under 3-channel with $M = 5$ and $L = 5$.

obtain higher expected utility, which means that the centralized strategy is not a NE and SUs have incentive to deviate.

C. Fast Algorithm for Multiple Channel Access

In this simulation, we evaluate the performance of the proposed fast algorithm for multi-channel case, which is denoted as suboptimal strategy hereafter. In Fig. 6, the suboptimal strategy is compared with the proposed strategy, myopic strategy and random strategy in terms of social welfare under 3-channel case, where the channel parameters are set to be: for channel 1, $r_0 = 55$s and $r_1 = 25$s; for channel 2, $r_0 = 45$s and $r_1 = 40$s; for channel 3, $r_0 = 25$s and $r_1 = 55$s. We can see that the suboptimal strategy achieves the social welfare very close to that of the optimal one, i.e., the proposed strategy using modified value iteration, and is still better than the myopic and random strategies. Therefore, considering the low complexity of the suboptimal strategy, it is more practical to use the suboptimal strategy for the multi-channel case.

VII. Conclusion

In this paper, we extended the previous Chinese Restaurant Game work [7] into a dynamic population setting and proposed a Dynamic Chinese Restaurant Game for cognitive radio networks. Based on the Bayesian learning rule, we introduced a channel state learning method for SUs to estimate the primary channel state. We modeled the channel access problem as a Multi-dimensional MDP and designed a modified value iteration algorithm to find the optimal strategy. The simulation results show that compared with the centralized approach that maximizes the social welfare with an intractable computational complexity, the proposed scheme achieves comparable social welfare performance with much lower complexity, while compared with random strategy and myopic strategy, the proposed scheme achieves much better social welfare performance. Moreover, the proposed scheme maximizes a new coming SU's expected utility and thus achieves Nash equilibrium where no user has the incentive to deviate.

References

[1] B. Wang and K. J. R. Liu, "Advances in cognitive radios: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 5–23, 2011.

[2] K. J. R. Liu and B. Wang, *Cognitive Radio Networking and Security: A Game Theoretical View*. Cambridge University Press, 2010.

[3] B. Wang, Y. Wu, and K. J. R. Liu, "Game theory for cognitive radio networks: An overview," *Computer Networks*, vol. 54, no. 14, pp. 2537–2561, 2010.

[4] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 3, pp. 40–62, 2011.

[5] W. H. Sandholm, "Negative externalities and evolutionary implementation," *The Review of Economic Studies*, vol. 72, no. 3, pp. 885–915, 2005.

[6] 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. 9, pp. 2127–2159, 2006.

[7] C.-Y. Wang, Y. Chen, and K. J. R. Liu., "Chinese restaurant game - part I: Theory of learning with negative network externality," http://arxiv.org/abs/1112.2188.

[8] H. V. Zhao, W. S. Lin, and K. J. R. Liu, *Behavior Dynamics in Media-Sharing Social Networks*. Cambridge University Press, 2011.

[9] Y. Chen and K. J. R. Liu, "Understanding microeconomic behaviors in social networking: An engineering view," *IEEE Signal Process. Mag.*, vol. 29, no. 2, pp. 53–64, 2012.

[10] D. Aldous, I. Ibragimov, J. Jacod, and D. Aldous., "Exchangeability and related topics," *Lecture Notes in Mathematics*, vol. 1117, 1985.

[11] C.-Y. Wang, Y. Chen, and K. J. R. Liu., "Chinese restaurant game - part II: Applications to wireless networking, cloud computing, and online social networking," http://arxiv.org/abs/1112.2187.

[12] B. Zhang, Y. Chen, C.-Y. Wang, and K. J. R. Liu, "Learning and decision making with negative externality for opportunistic spectrum access," in *Proc. IEEE Globecom*, 2012, pp. 1–6.

[13] C. Jiang, Y. Chen, K. J. R. Liu, and Y. Ren, "Analysis of interference in cognitive radio networks with unknown primary behavior," in *Proc. IEEE ICC*, 2012, pp. 1746–1750.

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

[15] D. R. Cox, *Renewal Theory*. Butler and Tanner, 1967.

[16] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 11, pp. 329–346, 2003.

[17] L. G. Epstein, J. Noor, and A. Sandroni, "Non-bayesian learning," *The B.E. Journal of Theoretical Economics*, vol. 10, no. 1, pp. 1–16, 2010.

[18] C. Jiang, Y. Chen, and K. J. R. Liu, "A renewal-theoretical framework for dynamic spectrum access with unknown primary behavior," in *Proc. IEEE Globecom*, 2012, pp. 1–6.

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

[20] C. Jiang, Y. Chen, Y.-H. Yang, C.-Y. Wang, and K. J. R. Liu, "Supplementary information for dynamic chinese restaurant game in cognitive radio networks," [Online]. Available: http://www.sig.umd.edu/jcx/INFOCOM2013SuppInfo.pdf.