Adaptive Channel Recommendation for Dynamic Spectrum Access

Xu Chen*, Jianwei Huang*, Husheng Li†
*Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong
†Department of Electrical Engineering and Computer Science, The University of Tennessee Knoxville, TN, USA
email:{cx008,jwhuang}@ie.cuhk.edu.hk,husheng@eecs.utk.edu

***Abstract*—We propose a dynamic spectrum access scheme where secondary users recommend "good" channels to each other and access accordingly. We formulate the problem as an average reward based Markov decision process. Since the action space of the Markov decision process is continuous (i.e., transmission probabilities), it is difficult to find the optimal policy by simply discretizing the action space and use the policy iteration, or value iteration. Instead, we propose a new algorithm based on the Model Reference Adaptive Search method, and prove its convergence to the optimal policy. Numerical results show that the proposed algorithm achieves up to 18% performance improvement than the static channel recommendation scheme and up to 63% performance improvement than the random access scheme, and is robust to channel dynamics.**

I. Introduction

Cognitive radio technology enables unlicensed secondary wireless users to opportunistically share the spectrum with licensed primary users, and thus offers a promising solution to address the spectrum under-utilization problem [1]. Designing an efficient spectrum access mechanism for cognitive radio networks, however, is challenging for several reasons: (1) *time-variation*: spectrum opportunities available for secondary users are often time-varying due to primary users' stochastic activities [1]; and (2) *limited observations*: each secondary user often has a limited view of the spectrum opportunities due to the limited spectrum sensing capability [2]. Several characteristics of the wireless channels, on the other hand, turn out to be useful for designing efficient spectrum access mechanisms: (1) *temporal correlations*: spectrum availabilities are correlated in time, and thus observations in the past can be useful in the near future [3]; and (2) *spatial correlation*: secondary users close to one another may experience similar spectrum availabilities [4]. In this paper, we shall explore the time and space correlations and propose a recommendation-based collaborative spectrum access algorithm, which achieves good communication performances for the secondary users.

Our algorithm design is directly inspired by the recommendation system in the electronic commerce industry. For example, existing owners of various products can provide recommendations (reviews) on Amazon.com, so that other potential customers can pick the products that best suit their needs. Motivated by this, Li in [5] proposed a static channel recommendation scheme, where secondary users recommend the channels they have successfully accessed to nearby secondary users. Since each secondary user originally only has a limited view of spectrum availability, such information exchange enables secondary users to take advantages of the correlations in time and space, make more informed decisions, and achieve a high total transmission rate.

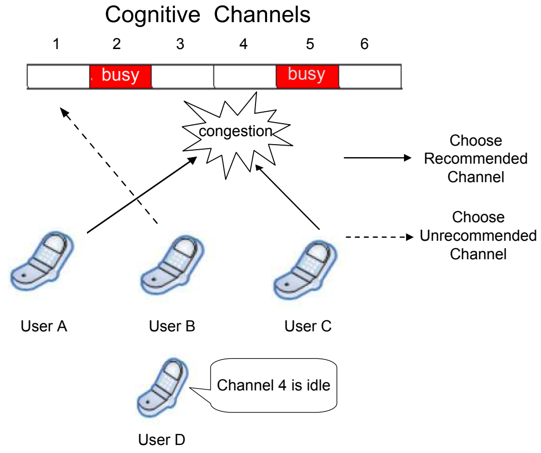

Fig. 1. Illustration of the channel recommendation scheme. User D recommends channel 4 to other users. As a result, both user A and user C access the same channel 4, and thus lead to congestion and a reduced rate for both users.

The recommendation scheme in [5], however, is rather static and does not dynamically change with network conditions. In particular, the static scheme ignores two important characteristics of cognitive radios. The first one is the *time variability* we mentioned before. The second one is the *congestion effect*. As depicted in Figure 1, too many users accessing the same good channel leads to congestion and a reduced rate for everyone.

To address the shortcomings of the static recommendation scheme, in this paper we propose an adaptive channel recommendation scheme, which adaptively changes the spectrum access probabilities based on users' latest channel recommendations. We formulate and analyze the system as a Markov decision process (MDP), and propose a numerical algorithm that always converges to the optimal spectrum access policy.

The main results and contributions of this paper include:

- *Markov decision process formulation*: we formulate and analyze the optimal recommendation-based spectrum ac-

This work is supported by the General Research Funds (Project Number 412509 and 412710) established under the University Grant Committee of the Hong Kong Special Administrative Region, China, and the US National Science Foundation under grants CCF-0830451 and ECCS-0901425.

cess as an average reward MDP.

- *Novel algorithm for finding the optimal policy*: we propose an algorithm based on the recently developed Model Reference Adaptive Search method [6] to find the optimal stationary spectrum access policy. The algorithm has a low complexity even when dealing with a continuous action space of the MDP. We also show that it always converges to the optimal stationary policy.
- *Superior Performance*: we show that the proposed algorithm achieves up to 18% performance improvement than the static channel recommendation scheme, and is also robust to channel dynamics.

The rest of the paper is organized as follows. We discuss the related work in Section II. We then introduce the system model and the static channel recommendation scheme in Sections III and IV, respectively. We next discuss the motivation for designing an adaptive channel recommendation scheme in Section V. The Markov decision process formulation is presented in Section VI, followed by the Model Reference Adaptive Search based algorithm in Section VII. We illustrate the performance of the algorithm through numerical results in Section VIII. Finally, we conclude in Section IX. *Due to space limitations, the details for several proofs are given in our online technical report [7].*

II. Related Work

The spectrum access by multiple secondary users can be either *uncoordinated* or *coordinated*. For the uncoordinated case, multiple secondary users compete with other for the resource. Huang *et al.* in [8] designed two auction mechanisms to allocate the interference budget among selfish users. Southwell and Huang in [9] studied the largest and smallest convergence time to an equilibrium when secondary users access multiple channels in a distributed fashion. Liu *et al.* in [10] modeled the interactions among spatially separated users as congestion games with resource reuse. Li and Han in [11] applied the graphic game theory to address the spectrum access problem with limited range of mutual interference. Anandkumar *et al.* in [12] proposed a learning-based approach for competitive spectrum access with incomplete spectrum information. Law *et al.* in [13] showed that uncoordinated spectrum access may lead to poor system performance.

For the coordinated spectrum access, Zhao *et al.* in [14] proposed a dynamic group formation algorithm to distribute secondary users' transmissions across multiple channels. Shu and Krunz proposed a multi-level spectrum opportunity framework in [15]. The above papers assumed that each secondary user knows the entire channel occupancy information. We consider the case where each secondary user only has a limited view of the system, and improve each other's information by recommendation.

Our algorithm design is partially inspired by the recommendation systems in the electronic commerce industry, where analytical methods such as collaborative filtering [16] and multi-armed bandit process modeling [17] are useful. However, we cannot directly apply the existing methods to analyze cognitive radio networks due to the unique congestion effect in our model.

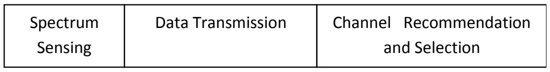

Fig. 2. Structure of each spectrum access time slot

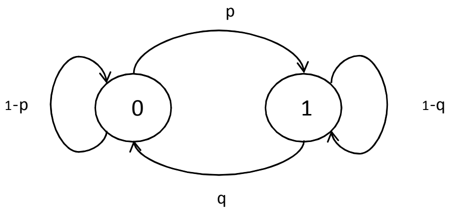

Fig. 3. Two states Markovian channel model

III. System Model

We consider a cognitive radio network with M independent and stochastically identical primary channels. N secondary users try to access these channels using a slotted transmission structure (see Figure 2). The secondary users can exchange information by broadcasting messages over a common control channel.[1] We assume that the secondary users are located close-by, thus they experience the same channel availability and can hear one another's broadcasting messages. To protect the primary transmissions, secondary users need to sense the channel states before the data transmission.

The system model is described as follows:

- *Channel State:* For each primary channel m, the channel state at time slot t is
$$S_m(t) = \begin{cases} 0, & \text{if channel } m \text{ is occupied by} \\ & \text{primary transmissions,} \\ 1, & \text{if channel } m \text{ is idle.} \end{cases}$$
- *Channel State Transition:* The channel states change according to independent and identical Markovian processes (see Figure 3). We denote the channel state probability vector of channel m at time t as
$$\boldsymbol{p}_m(t) \triangleq (Pr\{S_m(t)=0\}, Pr\{S_m(t)=1\}),$$
which follows a two-state Markov chain as
$$\boldsymbol{p}_m(t) = \boldsymbol{p}_m(t-1)\Gamma, \forall t \geq 1,$$
with the transition matrix
$$\Gamma = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$
Note that when $p=0$ or $q=0$, the channel state stays unchanged. In the rest of the paper, we will look at the more interesting and challenging cases where $0 < p \leq 1$ and $0 < q \leq 1$. The stationary distribution of the Markov chain is given as
$$\lim_{t\to\infty} Pr\{S_m(t)=0\} = \frac{q}{p+q}, \quad (1)$$
$$\lim_{t\to\infty} Pr\{S_m(t)=1\} = \frac{p}{p+q}. \quad (2)$$

[1] Please refer to [18] for the details on how to set up and maintain a reliable common control channel in cognitive radio networks.

- *Maximum Rate per Channel:* When a secondary user transmits successfully on an idle channel, it achieves a data rate of B. Here we assume that channels and users are homogeneous.
- *Congestion Effect:* When multiple secondary users try to access the same channel, each secondary user will execute the following two steps:
    - Randomly generate a continuous backoff timer value τ according to a common uniform distribution on $(0, \tau_{\max})$.
    - Once the timer expires, monitor the channel and transmit data only if the channel is clear.

Lemma 1. *If $k_m(t)$ secondary users compete to access the same channel m at time slot t, then the expected throughput of each user is $\frac{BS_m(t)}{k_m(t)}$.*

Lemma 1 shows that expected throughput of a secondary user decreases as the number of users accessing the same channel increases. However, the total expected rate of all $k_m(t)$ secondary users is $BS_m(t)$, i.e., there is no wasted resource due to users' competition.[2] Due to space limitations, we give the detailed proof of Lemma 1 in [7].

IV. Review of Static Channel Recommendation

The key idea of the static channel recommendation scheme in [5] is that secondary users inform each other about the available channels they have just accessed. More specifically, each secondary user executes the following three stages synchronously during each time slot (See Figure 2):

- *Spectrum sensing:* sense one of the channels based on channel selection result made at the end of the previous time slot.
- *Data transmission:* if the channel sensing result is idle, compete for the channel with the timer mechanism described in Section III. Then transmit data packets if the user successfully grabs the channel.
- *Channel recommendation and selection:*
    - *Announce recommendation:* if the user has successfully accessed an idle channel, broadcast this channel ID to all other secondary users.
    - *Collect recommendation:* collect recommendations from other secondary users and store them in a buffer. Typically, the correlation of channel availabilities between two slots diminishes as the time difference increases. Therefore, each secondary user will only keep the recommendations received from the most recent W slots and discard the out-of-date information. The user's own successful transmission history within W recent time slots is also stored in the buffer. W is a system design parameter and will be further discussed later.
    - *Select channel*: choose a channel to sense at the next time slot by putting more weights on the recommended channels according to a *static branching probability* P_{rec}. Suppose that the user has R different channel recommendations in the buffer, then the probability of accessing a channel m is

$$P_m = \begin{cases} \frac{P_{rec}}{R}, & \text{if channel m is recommended,} \\ \frac{1-P_{rec}}{M-R}, & \text{otherwise,} \end{cases} \tag{3}$$

A larger value of P_{rec} means that putting more weight on the recommended channels.

To illustrate the channel selection process, let us take the network in Figure 1 as an example. Suppose that the branching probability $P_{rec} = 0.4$. Since only $R = 1$ recommendation is available (i.e., channel 4), the probabilities of choosing the recommended channel 4 and any unrecommended channel are $\frac{0.4}{1} = 0.4$ and $\frac{1-0.4}{6-1} = 0.12$, respectively.

Numerical studies in [5] showed that the static channel recommendation scheme achieves a higher performance over the traditional random channel access scheme without information exchange. However, the fixed value of P_{rec} limits the performance of the static scheme, as explained next.

V. Motivations For Adaptive Channel Recommendation

The static channel recommendation mechanism is simple to implement due to a fixed value of P_{rec}. However, it may lead to significant congestions when the number of recommended channels is small. In the extreme case when only $R = 1$ channel is recommended, calculation (3) suggests that every user will access that channel with a probability P_{rec}. When the number of users N is large, the expected number of users accessing this channel NP_{rec} will be high. Thus heavy congestion happens and each secondary user will get a low expected throughput.

A better way is to adaptively change the value of P_{rec} based on the number of recommended channels. This is the key idea of our proposed algorithm. To illustrate the advantage of adaptive algorithms, let us first consider a simple heuristic adaptive algorithm. In this algorithm, we choose the branching probability such that the expected number of secondary users choosing a single recommended channel is one. To achieve this, we need to set P_{rec} as in Lemma 2.

Lemma 2. *If we choose the branching probability $P_{rec} = \frac{R}{N}$, then the expected number of secondary users choosing any one of the R recommended channels is one.*

Due to space limitations, we give the detailed proof of Lemma 2 in [7].

Without going through detailed analysis, it is straightforward to show the benefit for such adaptive approach through simple numerical examples. Let us consider a network with $M = 10$ channels and $N = 5$ secondary users. For each channel m, the initial channel state probability vector is $\boldsymbol{p}_m(0) = (0, 1)$ and the transition matrix is

$$\Gamma = \begin{bmatrix} 1 - 0.01\epsilon & 0.01\epsilon \\ 0.01\epsilon & 1 - 0.01\epsilon \end{bmatrix},$$

where ϵ is called the dynamic factor. A larger value of ϵ implies that the channels are more dynamic over time. We

[2] This may not be true for other random MAC mechanisms such as the slotted Aloha.

are interested in the time average system throughput

$$U = \frac{\sum_{t=1}^{T}\sum_{n=1}^{N} u_n(t)}{T},$$

where $u_n(t)$ is the throughput of user n at time slot t. In the simulation, we set the total number of time slots $T = 2000$.

We implement the following three channel access schemes:

- Random access scheme: each secondary user selects a channel randomly.
- Static channel recommendation scheme as in [5] with the *optimal* constant branching probability $P_{rec} = 0.7$.
- Heuristic adaptive channel recommendation scheme with the variable branching probability $P_{rec} = \frac{R}{N}$.

Figure 4 shows that the heuristic adaptive channel recommendation scheme outperforms the static channel recommendation scheme, which in turn outperforms the random access scheme. Moreover, the heuristic adaptive scheme is more robust to the dynamic channel environment, as it decreases slower than the static scheme when ϵ increases.

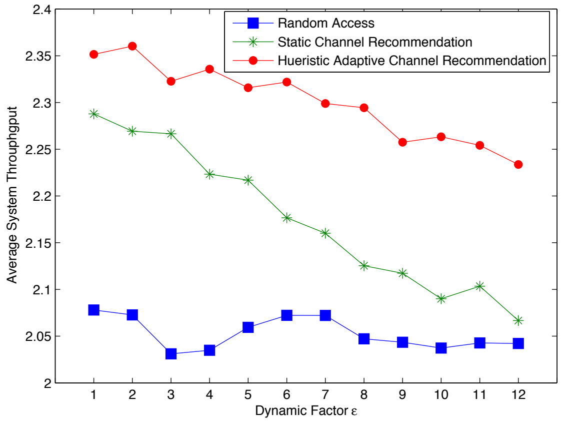

Fig. 4. Comparison of three channel access schemes

We can imagine that an optimal adaptive scheme (by setting the right $P_{rec}(t)$ over time) can further increase the network performance. However, computing the optimal branching probability in closed-form is very difficult. In the rest of the paper, we will focus on designing an efficient algorithm to achieve the optimum.

VI. Adaptive Channel Recommendation Scheme

To find the optimal adaptive spectrum access strategy, we formulate the system as a Markov Decision Process (MDP). For the sake of simplicity, we assume that the recommendation buffer size $W = 1$, i.e., users only consider the recommendations received in the last time slot. Our method also applies to the case when $W > 1$ by using a high-order MDP formulation, although the analysis is more involved.

A. MDP Formulation For Adaptive Channel Recommendation

We model the system as a MDP as follows:

- *System state*: $R \in \mathcal{R} \triangleq \{0, 1, ..., \min\{M, N\}\}$ denotes the number of recommended channels at the end of time slot t. Since we assume that all channels are statistically identical, then there is no need to keep track of the recommended channel IDs[3].
- *Action*: $P_{rec} \in \mathcal{P} \triangleq (0, 1)$ denotes the branching probability of choosing the set of recommended channels.
- *Transition probability*: The probability that action P_{rec} in system state R in time slot t will lead to system state R' in the next time slot is
$$P_{R,R'}^{P_{rec}} = Pr\{R(t+1) = R' | R(t) = R, P_{rec}(t) = P_{rec}\}.$$
We can compute this probability as in (4), with detailed derivations given in [7].
- *Reward*: $U(R, P_{rec})$ is the expected system throughput in the next time slot when the action P_{rec} is taken under the current system state R, i.e.,
$$U(R, P_{rec}) = \sum_{R \in \mathcal{R}'} P_{R,R'}^{P_{rec}} U_{R'},$$
where $U_{R'}$ is the system throughput in state R'. If R' idle channels are utilized by the secondary users in a time slot, then these R' channels will be recommended at the end of the time slot. Thus, we have
$$U_{R'} = R'B.$$
Recall that B is the data rate that a single user can obtain on an idle channel.
- *Stationary Policy:* $\pi \in \Omega \triangleq \mathcal{P}^{|\mathcal{R}|}$ maps each state R to an action P_{rec}, i.e., $\pi(R)$ is the action P_{rec} taken when the system is in state R. The mapping is stationary and does not depend on time t.

Given a stationary policy π and the initial state $R_0 \in \mathcal{R}$, we define the network's value function as the time average system throughput, i.e.

$$\Phi_\pi(R_0) = \lim_{T\to\infty} \frac{1}{T} E_\pi \left[\sum_{t=0}^{T-1} U(R(t), \pi(R(t)))\right].$$

We want to find an optimal stationary policy π^* that maximizes the value function $\Phi_\pi(R_0)$ for any initial state R_0, i.e.

$$\pi^* = \arg\max_\pi \Phi_\pi(R_0), \forall R_0 \in \mathcal{R}.$$

Notice that this is a system wide optimization, although the optimal solution can be implemented in a distributed fashion. This is because every user knows the number of recommended channels R, and it can determine the same optimal access probability locally.

VII. Model Reference Adaptive Search For Optimal Spectrum Access Policy

Next we will design an algorithm that can converge to the optimal policy under general system parameters. Since the action space of the adaptive channel recommendation MDP is continuous (i.e., choosing a probability P_{rec} in $(0, 1)$), the traditional method of discretizing the action space followed by the policy, or value iteration cannot guarantee to converge to the optimal policy in our problem. To overcome this difficulty, we propose a new algorithm developed from the

[3] Users need to know the IDs of the recommended channels in order to access them. However, the IDs are not important in terms of MDP analysis.

$$
\begin{aligned}
P_{R,R'}^{P_{rec}} = & \sum_{m_r+m_u=R'} \sum_{R\geq \bar{m}_r \geq m_r, M-R\geq \bar{m}_u \geq m_u} \sum_{n_r+n_u=N, n_r\geq \bar{m}_r, n_u \geq \bar{m}_u} \binom{N}{n_r} P_{rec}^{n_r}(1-P_{rec})^{n_u} \\
& \cdot \binom{\bar{m}_r}{m_r} (1-q)^{m_r} q^{\bar{m}_r - m_r} \frac{R!}{(R-\bar{m}_r)!} \binom{n_r-1}{\bar{m}_r-1} R^{-n_r} \\
& \cdot \binom{\bar{m}_u}{m_u} (\frac{p}{p+q})^{m_u} (\frac{q}{p+q})^{\bar{m}_u - m_u} \frac{(M-R)!}{(M-R-\bar{m}_u)!} \binom{n_u-1}{\bar{m}_u-1} (M-R)^{-n_u}. \qquad (4)
\end{aligned}
$$

Model Reference Adaptive Search method, which was recently developed in the Operations Research community [6]. We will show that the proposed algorithm is easy to implement and is provably convergent to the optimal policy.

A. Model Reference Adaptive Search Method

We first introduce the basic idea of the Model Reference Adaptive Search (MRAS) method. Later on, we will show how the method can be used to obtain optimal spectrum access policy for our problem.

The MRAS method is a new randomized method for global optimization [6]. The key idea is to randomize the original optimization problem over the feasible region according to a specified probabilistic model. The method then generates candidate solutions and updates the probabilistic model on the basis of elite solutions and a reference model, so that to guide the future search toward better solutions.

Formally, let $J(x)$ be the objective function to maximize. The MRAS method is an iterative algorithm, and it includes three phases in each iteration k:

- *Random solution generation*: generate a set of random solutions $\{x\}$ in the feasible set χ according to a parameterized probabilistic model $f(x, v_k)$, which is a probability density function (pdf) with parameter v_k. The number of solutions to generate is a fixed system parameter.
- *Reference distribution construction*: select *elite* solutions among the randomly generated set in the previous phase, such that the chosen ones satisfy $J(x) \geq \gamma$. Construct a reference probability distribution as
$$g_k(x)= \begin{cases} \frac{I_{\{J(x)\geq\gamma\}}}{E_{f(x,v_0)}[\frac{I_{\{J(x)\geq\gamma\}}}{f(x,v_0)}]} & k=1, \\ \frac{e^{J(x)}I_{\{J(x)\geq\gamma\}}g_{k-1}(x)}{E_{g_{k-1}}[e^{J(x)}I_{\{J(x)\geq\gamma\}}]} & k\geq 2. \end{cases} \qquad (5)$$
Here $I_{\{\varpi\}}$ is an indicator function, which equals 1 if the event ϖ is true and zero otherwise. Parameter v_0 is the initial parameter for the probabilistic model (used during the first iteration, i.e., $k=1$), and $g_{k-1}(x)$ is the reference distribution in the previous iteration (used when $k \geq 2$).
- *Probabilistic model update*: update the parameter v of the probabilistic model $f(x,v)$ by minimizing the Kullback-Leibler divergence between $g_k(x)$ and $f(x,v)$, i.e.
$$v_{k+1} = \arg\min_{v} E_{g_k}\left[\ln\frac{g_k(x)}{f(x,v)}\right]. \qquad (6)$$

By constructing the reference distribution according to (5), the expected performance of random elite solutions can be improved under the new reference distribution, i.e.,

$$
\begin{aligned}
E_{g_k}[e^{J(x)}I_{\{J(x)\geq\gamma\}}]= & \frac{\int_{x\in\chi} e^{2J(x)}I_{\{J(x)\geq\gamma\}}g_{k-1}(x)dx}{E_{g_{k-1}}[e^{J(x)}I_{\{J(x)\geq\gamma\}}]} \\
= & \frac{E_{g_{k-1}}[e^{2J(x)}I_{\{J(x)\geq\gamma\}}]}{E_{g_{k-1}}[e^{J(x)}I_{\{J(x)\geq\gamma\}}]} \\
\geq & E_{g_{k-1}}[e^{J(x)}I_{\{J(x)\geq\gamma\}}]. \qquad (7)
\end{aligned}
$$

To find a better solution to the optimization problem, it is natural to update the probabilistic model (from which random solution are generated in the first stage) as close to the new reference probability as possible, as done in the third stage.

B. Model Reference Adaptive Search For Optimal Spectrum Access Policy

In this section, we design an algorithm based on the MRAS method to find the optimal spectrum access policy. Here we treat the adaptive channel recommendation MDP as a global optimization problem over the policy space. The key challenge is the choice of proper probabilistic model $f(\cdot)$, which is crucial for the convergence of the MRAS algorithm.

1) Random Policy Generation: To apply the MRAS method, we first need to set up a random policy generation mechanism. Since the action space of the channel recommendation MDP is continuous, we use the Gaussian distributions. Specifically, we generate sample actions $\pi(R)$ from a Gaussian distribution for each system state $R \in \mathcal{R}$ independently, i.e. $\pi(R) \sim \mathcal{N}(\mu_R, \sigma_R^2)$.[4] In this case, a candidate policy π can be generated from the joint distribution of $|\mathcal{R}|$ independent Gaussian distributions, i.e.,

$$
\begin{aligned}
(\pi(0), ..., \pi(\min\{M,N\})) \sim & \mathcal{N}(\mu_0, \sigma_0^2) \times \cdots \\
& \times \mathcal{N}(\mu_{\min\{M,N\}}, \sigma^2_{\min\{M,N\}}).
\end{aligned}
$$

As shown later, Gaussian distribution has nice analytical and convergent properties for the MRAS method.

For the sake of brevity, we denote $f(\pi(R), \mu_R, \sigma_R)$ as the pdf of the Gaussian distribution $\mathcal{N}(\mu_R, \sigma_R^2)$, and denote $f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma})$ as random policy generation mechanism with parameters $\boldsymbol{\mu} \triangleq (\mu_0, ..., \mu_{\min\{M,N\}})$ and $\boldsymbol{\sigma} \triangleq (\sigma_0, ..., \sigma_{\min\{M,N\}})$, i.e.,

$$
\begin{aligned}
f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma}) = & \prod_{R=0}^{\min\{M,N\}} f(\pi(R), \mu_R, \sigma_R) \\
= & \prod_{R=0}^{\min\{M,N\}} \frac{1}{\sqrt{2\varphi\sigma_R^2}} e^{-\frac{(\pi(R)-\mu_R)^2}{2\sigma_R^2}},
\end{aligned}
$$

[4] Note that the Gaussian distribution has a support of $(-\infty, +\infty)$, which is larger than the feasible region of $\pi(R)$. This issue will be handled in Section VII-B2.

where φ is the circumference-to-diameter ratio.

2) System Throughput Evaluation: Given a candidate policy π randomly generated based on $f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma})$, we need to evaluate the expected system throughput Φ_π. From (4), we obtain the transition probabilities $P_{R,R'}^{\pi(R)}$ between any system state $R, R' \in \mathcal{R}$. Since a policy π leads to a finitely irreducible Markov chain, we can obtain its stationary distribution. Let us denote the transition matrix of the Markov chain as $Q \triangleq [P_{R,R'}^{\pi(R)}]_{|\mathcal{R}|\times|\mathcal{R}|}$ and the stationary distribution as $\boldsymbol{p} = (Pr(0), ..., Pr(\min\{M, N\}))$. Obviously, the stationary distribution can be obtained by solving the equation

$$\boldsymbol{p}Q = \boldsymbol{p}.$$

We then calculate the expected system throughput Φ_π by

$$\Phi_\pi = \sum_{R\in\mathcal{R}} Pr(R)U_R.$$

Note that in the discussion above, we assume that $\pi \in \Omega$ implicitly, where Ω is the feasible policy space. Since Gaussian distribution has a support over $(-\infty, +\infty)$, we thus extend the definition of expected system throughput Φ_π over $(-\infty, +\infty)^{|\mathcal{R}|}$ as

$$\Phi_\pi = \begin{cases} \sum_{R\in\mathcal{R}} Pr(R)U_R & \pi \in \Omega, \\ -\infty & \text{Otherwise.} \end{cases}$$

In this case, whenever any generated policy π is not feasible, we have $\Phi_\pi = -\infty$. As a result, such policy π will not be selected as an elite sample (discussed next) and will not used for probability updating. Hence the search of MRAS algorithm will not bias towards any unfeasible policy space.

3) Reference Distribution Construction: To construct the reference distribution, we first need to select the elite policies. Suppose L candidate policies, $\pi_1, \pi_2, ..., \pi_L$, are generated at each iteration. We order them based on an increasing order of the expected system throughputs Φ_π, i.e., $\Phi_{\hat{\pi}_1} \leq \Phi_{\hat{\pi}_2} \leq ... \leq \Phi_{\hat{\pi}_L}$, and set the elite threshold as

$$\gamma = \Phi_{\hat{\pi}_{\lceil(1-\rho)L\rceil}},$$

where $0 < \rho < 1$ is the elite ratio. For example, when $L = 100$ and $\rho = 0.4$, then $\gamma = \Phi_{\hat{\pi}_{60}}$ and the last 40 samples in the sequence will be selected as elite samples. Note that as long as L is sufficiently large, we shall have $\gamma < \infty$ and hence only feasible policies π are selected. According to (5), we then construct the reference distribution as

$$g_k(\pi) = \begin{cases} \frac{I_{\{\Phi_\pi \geq \gamma\}}}{E_{f(\pi,\boldsymbol{\mu}_0,\boldsymbol{\sigma}_0)}[\frac{I_{\{\Phi_\pi \geq \gamma\}}}{f(\pi,\boldsymbol{\mu}_0,\boldsymbol{\sigma}_0)}]} & k = 1, \\ \frac{e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} g_{k-1}(\pi)}{E_{g_{k-1}}[e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}}]} & k \geq 2. \end{cases} \tag{8}$$

4) Policy Generation Update: For the MRAS algorithm, the critical issue is the updating of random policy generation mechanism $f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma})$, or solving the problem in (6). The optimal update rule is described as follow.

Theorem 1. *The optimal parameter $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ that minimizes the Kullback-Leibler divergence between the reference distribution $g_k(\pi)$ in (8) and the new policy generation mechanism $f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma})$ is*

$$\mu_R = \frac{\int_{\pi\in\Omega} e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} \pi(R) d\pi}{\int_{\pi\in\Omega} e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} d\pi}, \forall R \in \mathcal{R}, \tag{9}$$

$$\sigma_R^2 = \frac{\int_{\pi\in\Omega} e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} [\pi(R) - \mu_R]^2 d\pi}{\int_{\pi\in\Omega} e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} d\pi}, \forall R \in \mathcal{R}. \tag{10}$$

Proof: First, from (8), we have

$$g_1(\pi) = \frac{I_{\{\Phi_\pi \geq \gamma\}}}{E_{f(\pi,\boldsymbol{\mu}_0,\boldsymbol{\sigma}_0)}[\frac{I_{\{\Phi_\pi \geq \gamma\}}}{f(\pi,\boldsymbol{\mu}_0,\boldsymbol{\sigma}_0)}]} = \frac{I_{\{\Phi_\pi \geq \gamma\}}}{\int_{\pi\in\Omega} I_{\{\Phi_\pi \geq \gamma\}} d\pi},$$

and

$$\begin{aligned} g_2(\pi) &= \frac{e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} g_1(\pi)}{E_{g_1}[e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}}]} \\ &= \frac{e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} I_{\{\Phi_\pi \geq \gamma\}}}{E_{g_1}[e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}}] \int_{\pi\in\Omega} I_{\{\Phi_\pi \geq \gamma\}} d\pi} \\ &= \frac{e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} I_{\{\Phi_\pi \geq \gamma\}}}{\int_{\pi\in\Omega} e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} \frac{I_{\{\Phi_\pi \geq \gamma\}}}{\int_{\pi\in\Omega} I_{\{\Phi_\pi \geq \gamma\}} d\pi} d\pi \int_{\pi\in\Omega} I_{\{\Phi_\pi \geq \gamma\}} d\pi} \\ &= \frac{e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}}}{\int_{\pi\in\Omega} e^{\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} d\pi}. \end{aligned}$$

Repeat the above computation iteratively, we have

$$g_k(\pi) = \frac{e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}}}{\int_{\pi\in\Omega} e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} d\pi}, k \geq 1. \tag{11}$$

Then the problem in (6) is equivalent to solving

$$\begin{aligned} &\max_{\boldsymbol{\mu},\boldsymbol{\sigma}} \int_{\pi\in\Omega} g_k(\pi) \ln f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma}) d\pi, \\ &\text{subject to} \quad \boldsymbol{\mu}, \boldsymbol{\sigma} \succeq 0. \end{aligned} \tag{12}$$

Substituting (11) into (12), we have

$$\begin{aligned} &\max_{\boldsymbol{\mu},\boldsymbol{\sigma}} \int_{\pi\in\Omega} e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} \ln f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma}) d\pi, \\ &\text{subject to} \quad \boldsymbol{\mu}, \boldsymbol{\sigma} \succeq 0. \end{aligned} \tag{13}$$

Function $f(\pi(R), \mu_R, \sigma_R)$ is log-concave, since it is the pdf of the Gaussian distribution. Since the log-concavity is closed under multiplication, then $f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \prod_{R=0}^{\min\{M,N\}} f(\pi(R), \mu_R, \sigma_R)$ is also log-concave. It implies the problem in (12) is a concave optimization problem. The corresponding first order conditions are

$$\frac{\partial \int_{\pi\in\Omega} e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} \ln f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma}) d\pi}{\partial \mu_R} = 0, \forall R \in \mathcal{R},$$

$$\frac{\partial \int_{\pi\in\Omega} e^{(k-1)\Phi_\pi} I_{\{\Phi_\pi \geq \gamma\}} \ln f(\pi, \boldsymbol{\mu}, \boldsymbol{\sigma}) d\pi}{\partial \sigma_R} = 0, \forall R \in \mathcal{R},$$

which leads to (9) and (10). Due to the concavity of the optimization problem in (12), the solution is also the global optimum for the random policy generation updating. ■

5) MARS Algorithm For Optimal Spectrum Access Policy: Based on the MARS algorithm, we generate L candidate polices at each iteration. Then the updates in (9) and (10) are replaced by the sample average version in (16) and (17), respectively. As a summary, we describe the MARS-based algorithm for finding the optimal spectrum access policy of adaptive channel recommendation MDP in Algorithm 1.

Algorithm 1 MRAS-based Algorithm For Adaptive Recommendation Based Optimal Spectrum Access

```
Initialize parameters for Gaussian distributions $(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0)$, the elite ratio $\rho$, and the stopping criterion $\xi$. Set initial elite threshold $\gamma_0 = 0$ and iteration index $k = 0$.
repeat:
    Increase iteration index $k$ by 1.
    Generate $L$ candidate policies $\pi_1, ..., \pi_L$ from the random policy generation mechanism $f(\pi, \boldsymbol{\mu}_{k-1}, \boldsymbol{\sigma}_{k-1})$.
    Select elite policies by setting the elite threshold $\gamma_k = \max\{\Phi_{\hat{\pi}_{\lceil (1-\rho)L \rceil}}, \gamma_{k-1}\}$.
    Update the random policy generation mechanism by
        $\mu_{R,k} = \frac{\sum_{i=1}^{L} e^{(k-1)\Phi_\pi} I_{\{\Phi_{\pi_i} \geq \gamma_k\}} \pi_i(R)}{\sum_{i=1}^{L} e^{(k-1)\Phi_\pi} I_{\{\Phi_{\pi_i} \geq \gamma_k\}}}, \quad \forall R \in \mathcal{R}, \quad (16)$
        $\sigma_{R,k}^2 = \frac{\sum_{i=1}^{L} e^{(k-1)\Phi_\pi} I_{\{\Phi_{\pi_i} \geq \gamma_k\}} [\pi_i(R) - \mu_R]^2}{\sum_{i=1}^{L} e^{(k-1)\Phi_\pi} I_{\{\Phi_{\pi_i} \geq \gamma_k\}}}, \quad \forall R \in \mathcal{R}. \quad (17)$
until $\max_{R \in \mathcal{R}, \|} \sigma_R < \xi$.
```

C. Convergence of Model Reference Adaptive Search

In this part, we discuss the convergence property of the MRAS-based optimal spectrum access policy. For ease of exposition, we assume that the adaptive channel recommendation MDP has a unique global optimal policy. Numerical studies in [6] show that the MRAS method also converges where there are multiple global optimal solutions. We shall show that the random policy generation mechanism $f(\pi, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)$ will eventually generate the optimal policy.

Theorem 2. *For the MRAS algorithm, the limiting point of the policy sequence $\{\pi_k\}$ generated by the sequence of random policy generation mechanism $\{f(\pi, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)\}$ converges pointwisely to the optimal spectrum access policy π^* for the adaptive channel recommendation MDP, i.e.,*

$$\lim_{k\to\infty} E_{f(\pi,\boldsymbol{\mu}_k,\boldsymbol{\sigma}_k)}[\pi(R)] = \pi^*(R), \forall R \in \mathcal{R}, \quad (14)$$

$$\lim_{k\to\infty} Var_{f(\pi,\boldsymbol{\mu}_k,\boldsymbol{\sigma}_k)}[\pi(R)] = 0, \forall R \in \mathcal{R}. \quad (15)$$

The proof is given in [7].

From Theorem 2, we see that parameter $(\mu_{R,k}, \sigma_{R,k})$ for updating in (16) and (17) also converge, i.e.,

$$\lim_{k\to\infty} \mu_{R,k} = \pi^*(R), \forall R \in \mathcal{R},$$

$$\lim_{k\to\infty} \sigma_{R,k} = 0, \forall R \in \mathcal{R}.$$

Thus, we can use $\max_{R \in \mathcal{R}} \sigma_{R,k} < \xi$ as the stopping criterion in Algorithm 1.

VIII. Simulation Results

In this section, we investigate the proposed adaptive channel recommendation scheme by simulations. The results show that the adaptive channel recommendation scheme not only achieves a higher performance over the static channel recommendation scheme and random access scheme, but also is more robust to the dynamic change of the channel environments.

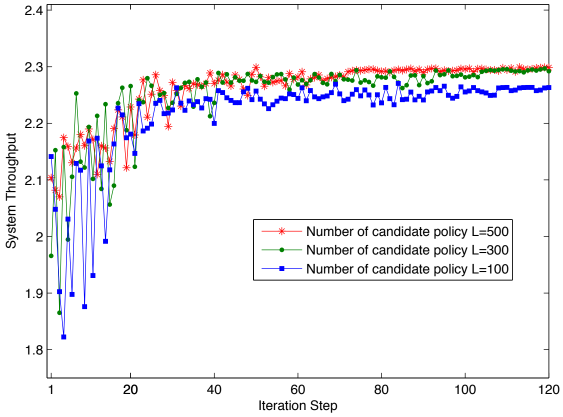

Fig. 5. The convergence of MRAS-based algorithm with different number of candidate policies per iteration

A. Simulation Setup

We consider a cognitive radio network consisting of multiple independent and stochastically identical primary channels. In order to take the impact of primary user's long run behavior into account, we consider the following two types of channel state transition matrices:

$$\text{Type 1: } \Gamma_1 = \begin{bmatrix} 1 - 0.005\epsilon & 0.005\epsilon \\ 0.025\epsilon & 1 - 0.025\epsilon \end{bmatrix}, \quad (18)$$

$$\text{Type 2: } \Gamma_2 = \begin{bmatrix} 1 - 0.01\epsilon & 0.01\epsilon \\ 0.01\epsilon & 1 - 0.01\epsilon \end{bmatrix}, \quad (19)$$

where ϵ is the dynamic factor. Recall that a larger ϵ means that the channels are more dynamic over time. Using (2), we know that channel models Γ_1 and Γ_2 have the stationary channel idle probabilities of $1/6$ and $1/2$, respectively. In other words, the primary activity level is much higher with the Type 1 channel than with the Type 2 channel.

We initialize the parameters of MRAS algorithm as follows. We set $\mu_R = 0.5$ and $\sigma_R = 0.5$ for the Gaussian distribution, which has 68.2% support over the feasible region $(0, 1)$. We found that the performance of the MRAS algorithm is insensitive to the elite ratio ρ when $\rho \leq 0.3$. We thus choose $\rho = 0.1$.

When using the MRAS-based algorithm, we need to determine how many (feasible) candidate policies to generate in each iteration. Figure 5 shows the convergence of MRAS algorithm with 100, 300, and 500 candidate policies per iteration, respectively. We have two observations. First, the number of iterations to achieve convergence reduces as the number of candidate policies increases. Second, the convergence speed is insignificant when the number changes from 300 to 500. We thus choose $L = 500$ for the experiments in the sequel.

B. Simulation Results

We consider two simulation scenarios: (1) the number of channels is greater than the number of secondary users, and (2) the number of channels is smaller than the number of

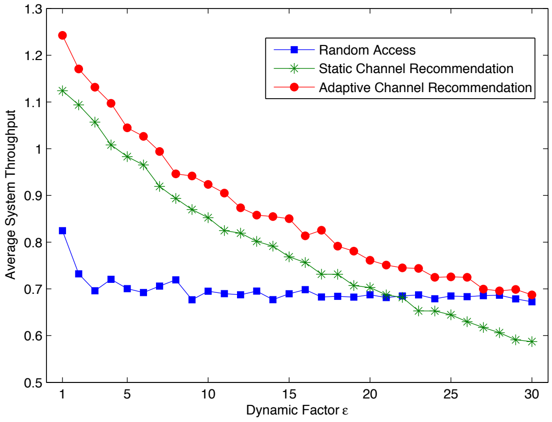

Fig. 6. System throughput with $M = 10$ channels and $N = 5$ users under the Type 1 channel state transition matrix

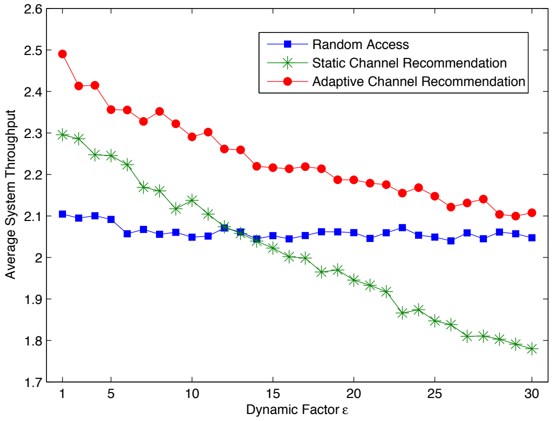

Fig. 7. System throughput with $M = 10$ channels and $N = 5$ users under the Type 2 channel state transition matrix

secondary users. For each case, we compare the adaptive channel recommendation scheme with the static channel recommendation scheme in [5] and a random access scheme.

1) More Channels, Fewer Users: We implement three spectrum access schemes with $M = 10$ channels and $N = 5$ secondary users . As there are enough channels to choose from, congestion is not a major issue in this setting. We choose the dynamic factor ϵ within a wide range to investigate the robustness of the schemes to the channel dynamics. The results are shown in Figures 6 – 9. From these figures, we see that

- *Superior performance of adaptive channel recommendation scheme (Figures 6 and 7)*: the adaptive channel recommendation scheme performs better than the random access scheme and static channel recommendation scheme. Typically, it offers 5%$\sim$18% performance gain over the static channel recommendation scheme.
- *Impact of channel dynamics (Figures 6 and 7)*: the performances of both adaptive and static channel recommendation schemes degrade as the dynamic factor ϵ increases. The reason is that both two schemes rely on the recommendation information from previous time slots to make decisions. When channel states change rapidly, the value of recommendation information diminishes. However, the adaptive channel recommendation is much more robust to the dynamic channel environment changing (See Figure 9). This is because the optimal adaptive policy takes the channel dynamics into account while the static one does not.
- *Impact of channel idleness level (Figures 8 and 9)*: Figure 8 shows the performance gain of the adaptive channel recommendation scheme over the random access scheme under two different types of transition matrix scenarios. We see that the performance gain decreases with the idle probability of the channel. This shows that the information of channel recommendations can enhance the spectrum access more efficiently when the primary activity level increases (i.e., when the channel idle probability is low as in Type I channel). Interestingly, Figure 9 shows that the performance gain of the adaptive channel recommendation scheme over the static channel recommendation scheme tends to increase with the channel idleness probability. This illustrates that the adaptive channel recommendation scheme can better utilize the channel opportunities than the static

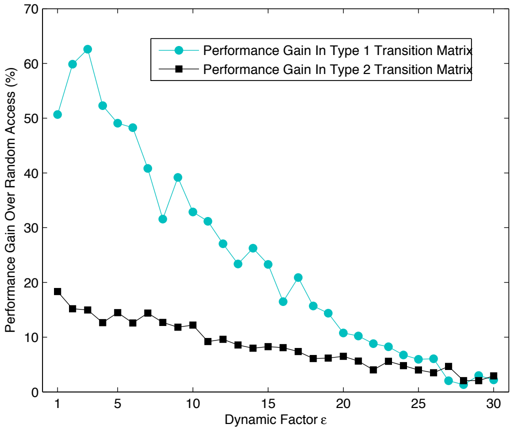

Fig. 8. Performance gain over random access scheme. The Type 1 and Type 2 channels have the stationary channel idle probabilities of $1/6$ and $1/2$, respectively.

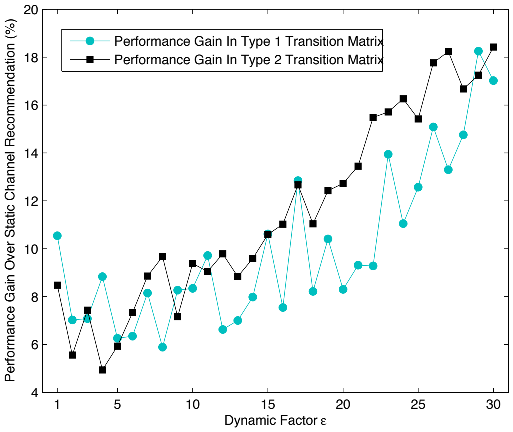

Fig. 9. Performance gain over static channel recommendation scheme. The Type 1 and Type 2 channels have the stationary channel idle probabilities of $1/6$ and $1/2$, respectively.

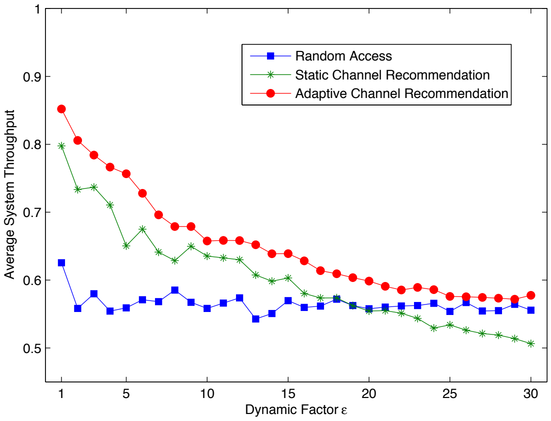

Fig. 10. System throughput with $M = 5$ channels and $N = 10$ users under Type 1 channel state transition matrix

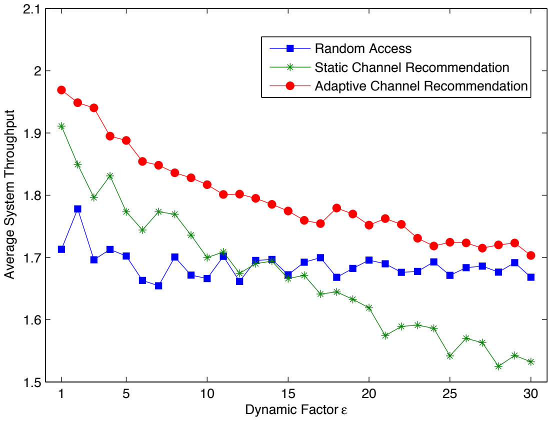

Fig. 11. System throughput with $M = 5$ channels and $N = 10$ users under Type 2 channel state transition matrix

recommendation scheme given the information of channel recommendations.

2) Fewer Channels, More Users: We next consider the case of $M = 5$ channels and $N = 10$ users, and show the simulation results in Figures 10 and 11. We can check that the observations in Section VIII-B1 still hold. In other words, the adaptive channel recommendation scheme still has a better performance than static one and random access scheme when the cognitive radio network suffers serve congestion effect.

IX. Conclusion

In this paper, we propose an adaptive channel recommendation scheme for efficient spectrum sharing. We formulate the problem as an average reward based Markov decision process. We then propose a novel MRAS-based algorithm that is provably convergent to the optimal policy. Numerical results show that our proposed algorithm outperforms the static approach in the literature by up to 18% and the random access scheme by up to 63% in terms of system throughput. Our algorithm is also more robust to the channel dynamics compared to the static counterpart.

In terms of future work, we will further characterize the theoretical properties of the optimal policy (some preliminary results can be found in [7]). We also consider to extend the analysis by taking the heterogeneity of channels into consideration. We also plan to consider the case where the secondary users are selfish. Designing an incentive-compatible channel recommendation mechanism for that case will be very interesting and challenging.

References

[1] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, Royal Institute of Technology (KTH) Stockholm, Sweden, 2000.

[2] 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, pp. 589–600, 2007.

[3] M. Wellens, J. Riihijarvi, and P. Mahonen, "Empirical time and frequency domain models of spectrum use," *Elsevier Physical Communications*, vol. 2, pp. 10–32, 2009.

[4] M. Wellens, J. Riihijarvi, M. Gordziel, and P. Mahonen, "Spatial statistics of spectrum usage: From measurements to spectrum models," in *IEEE International Conference on Communications*, 2009.

[5] H. Li, "Customer reviews in spectrum: recommendation system in cognitive radio networks," in *IEEE Symposia on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2010.

[6] J. Hu, M. Fu, and S. Marcus, "A model reference adaptive search algorithm for global optimization," *Operations Research*, vol. 55, pp. 549–568, 2007.

[7] X. Chen, J. Huang, and H. Li, "Adaptive channel recommendation for dynamic spectrum access," Department of Information Engineering, The Chinese University of Hong Kong, Tech. Rep., 2010. [Online]. Available: http://arxiv.org/abs/1102.4728

[8] J. Huang, R. Berry, and M. L. Honig, "Auction-based spectrum sharing," *ACM/Springer Mobile Networks and Applications Journal*, 2006.

[9] R. Southwell and J. Huang, "Convergence dynamics of resource-homogeneous congestion games," in *International Conference on Game Theory for Networks*, Shanghai, China, April 2011.

[10] M. Liu, S. Ahmad, and Y. Wu, "Congestion games with resource reuse and applications in spectrum sharing," in *International Conference on Game Theory for Networks*, 2009.

[11] H. Li and Z. Han, "Competitive spectrum access in cognitive radio networks: graphical game and learning," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2010.

[12] A. Anandkumar, N. Michael, and A. Tang, "Opportunistic spectrum access with multiple users: learning under competition," in *The IEEE International Conference on Computer Communications (Infocom)*, 2010.

[13] L. M. Law, J. Huang, M. Liu, and S. Li, "Price of anarchy of cognitive mac games," in *IEEE Global Communications Conference*, 2009.

[14] J. Zhao, H. Zheng, and G. Yang, "Distributed coordination in dynamic spectrum allocation networks," in *IEEE Symposia on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2005.

[15] T. Shu and M. Krunz, "Coordinated channel access in cognitive radio networks: a multi-level spectrum opportunity perspective," in *The IEEE International Conference on Computer Communications (Infocom)*, 2009.

[16] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, pp. 61–70, 1992.

[17] B. Awerbuch and R. Kleinberg, "Competitive collaborative learning," *Journal of Computer and System Sciences*, vol. 74, pp. 1271–1288, 2008.

[18] C. Cormio, Kaushik, and R. Chowdhury, "Common control channel design for cognitive radio wireless ad hoc networks using adaptive frequency hopping," *Elsevier Journal of Ad Hoc Networks*, vol. 8, pp. 430–438, 2010.