Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Transmission Control in Cognitive Radio Systems with Latency Constraints as a Switching Control Dynamic Game

Jane W. Huang, Vikram Krishnamurthy
Department of Electrical and Computer Engineering, The University of British Columbia
5500-2332 Main Mall, Vancouver, BC V6T 1Z4, Canada, Email: {janeh,vikramk}@ece.ubc.ca

Abstract— **This paper addresses the secondary user rate adaptation problem in cognitive radio networks. By modeling primary user activities and secondary user block fading channels as finite state Markov chains, we formulate the transmission rate adaptation problem of each secondary user as a zero-sum dynamic Markovian game with a delay constraint. The Nash equilibrium of the resulting game is available and all of the Nash equilibria have a unique value vector. Conditions are given so that the Nash equilibrium transmission policy of each user is a randomized mixture of pure threshold policies. Such threshold policies are easily implementable. Finally, we present a stochastic approximation algorithm which can adaptively estimate Nash equilibrium policies and track such policies for non-stationary problems where the statistics of the channel and user parameters evolve with time.**

I. Introduction

Cognitive radio systems [1], [2] present the opportunity to improve spectrum utilization by detecting unoccupied spectrum holes and assigning them to secondary users. Stochastic dynamic game theory is an essential tool for cognitive radio systems as it is able to exploit the correlated channels and primary user activities in the analysis of decentralized behavior of cognitive radios. In comparison, most games considered in wireless communication systems to date are static games. Such static game theoretic analyses have been applied in [3] in the context of power and rate control for uplink cellular systems. Furthermore, [4] and [5] adopted static game theoretic formulations to address the resource allocation problem in cognitive radio networks.

This paper considers the secondary user rate adaptation problem in cognitive radio systems where more than one user tries to access a spectrum hole. We assume a time division multiple access (TDMA) cognitive radio system model (as specified in IEEE 802.16 standard [6]) which schedules one user per spectrum hole at each time slot according to a predefined decentralized scheduling policy. The policy takes into account the behaviors of the primary users, as well as the channel quality and transmission delay of each secondary user. By modeling the primary user behaviors and transmission channel as a Markovian chain, the transmission rate adaptation problem of each user can be formulated as a zero-sum switching control Markovian dynamic game with a latency constraint.

A. Main Results

1. We first formulate the secondary user rate adaptation problem in cognitive radio networks as a constrained zero-sum switching controlled dynamic Markovian game in Section II. As we consider a TDMA system, the system state transmission probabilities only depend on the user who is transmitting. This feature fulfills the property of a special type of dynamic game which is called a switching control game [7], [8], which are games where the transition probabilities depend on only one player in each state. Such games can be solved by a finite sequence of Markov decision processes. In this paper, the problem is formulated to be a constrained switching control Markovian game [9], which is an extension of this type of game. Furthermore, both cost and constraint of the game we formulate are zero-sum.

2. It is shown that the Nash equilibria of the game are always available, and that all of them have the same value vector. Few realistic assumptions on the system are given in Section III.B. Under these assumptions, the optimal action policy is a randomization of two pure policies, each of the policies is monotone increasing on the buffer occupancy state.

3. Section III.D proposes a more computationally efficient algorithm to search for the optimal policy named stochastic approximation algorithm. Stochastic approximation algorithm greatly reduces the computational complexity of searching for the optimal policy. At the same time, it directly solves constrained zero-sum switching controlled Markovian games. The algorithm can do blind adaptation according to non-stationary channel and user statistics which evolve with time. Numerical results of the stochastic approximation algorithm are provided in Section IV.

II. Rate Adaptation Problem Formulation

This section describes the system model. The rate control problem of each secondary user can be formulated as a constrained dynamic Markovian game by modeling the primary user activities and channel quality as a Markovian chain. More specifically, under the predefined decentralized access rule, the problem presented is a special type of game namely a switching control Markovian dynamic game.

A. Markovian System Status Description

The channel quality of user k at time n is denoted as $h_k^{(n)}$. The channel model is assumed to be circularly symmetric complex Gaussian random variables which depend only on the previous time slot. After applying quantization for the channel quality, the channel state space is $h_k^{(n)} \in \{0, 1, 2, \dots\}$ where state 0 represents the case when the primary user is occupying the channel. The composition of channel states of all the K users can be written as $h^{(n)} = \{h_1^{(n)}, \dots, h_K^{(n)}\}$. Assume that the channel state $h^{(n)} \in \mathcal{H}, n = 1, 2, \dots, N$ is block fading, the block length equals to each time period. The channel states constitute a Markov process, its transition probability from time n to $(n+1)$ can be written as $\mathbb{P}(h^{(n+1)}|h^{(n)})$.

Use $b_k^{(n)}$ to represent the buffer occupancy state of user k at time n and $b_k^{(n)} \in \{0, 1, \dots, L\}$, the composition of buffer states of all the K users is $b^{(n)} = \{b_1^{(n)}, \dots, b_K^{(n)}\}$ which is an element of secondary user buffer state space $\mathcal{B}$.

In the system model, there is new incoming traffic at the beginning of each time slot. Refer to the number of new incoming packets at the nth time slot of the kth user as $f_k^{(n)}$, then $f^{(n)} = \{f_1^{(n)}, \dots, f_K^{(n)}\}$ is an element of the incoming traffic space $\mathcal{F}$. For simplicity, the incoming traffic is assumed to be independent and identically distributed (i.i.d.) in terms of time index n and user index k.

Let $\mathcal{S}$ denote the finite system state space, which comprises channel state $\mathcal{H}$, secondary user buffer state $\mathcal{B}$ and incoming traffic state $\mathcal{F}$. That is, $\mathcal{S} = \mathcal{H} \times \mathcal{B} \times \mathcal{F}$. Here $\times$ denotes a Cartesian product. $\mathcal{S}_k$ is used to indicate the states where user k is scheduled for transmission. $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_K$ are disjoint subsets of $\mathcal{S}$ with the property of $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \dots \cup \mathcal{S}_K$.

B. Action and Costs

If the kth user is scheduled for transmission at the nth time slot, its action $a_k^{(n)}$ represents the bits/symbol rate of the transmission. Different $a_k^{(n)}$ lead to different transmission rates. Assuming the system uses uncoded M-ary quadrature amplitude modulation (QAM) modulation, the different bits/symbol rates determine the modulation schemes as follows: $M = 2^{a_k^{(n)}}$.

Transmission cost: Let $c_i(s^{(n)}, a_1^{(n)}, \dots, a_K^{(n)})$ denote the transmission cost of user i at time n. When the channel state in $h^{(n)} = \{h_1^{(n)}, \dots, h_K^{(n)}\}$ and the action taken at time n is $a_k^{(n)}$. Specifically, $c_k(s^{(n)}, a_k^{(n)})$ is chosen to be the transmission bit error rate (BER) introduced by user k during transmission. If the transmission and holding costs of each user is independent of every other user then the system reduces to K decoupled users. To allow for interaction among the users, we couple the transmission and holding costs of the users. In particular, we assume that the system has both zero-sum transmission costs and zero-sum constraint. The zero-sum transmission costs can be written as:

$$\sum_{i=1}^{K} c_i(s^{(n)}, a_1^{(n)}, \dots, a_K^{(n)}) = 0. \tag{1}$$

The zero-sum transmission costs can be viewed as the user k who is scheduled for transmission pays a cost to other users, which is equivalently the other users receive a reward. Due to the *"budget balance property"* [10], the transmission costs among all the users are zero-sum. In [10], a similar assumption has been used for the resource allocation in a wireless multimedia system.

With $s^{(n)} \in \mathcal{S}_k$ and user k being scheduled to transmit at time n, the performance of that user only depends on itself. Due to the equality among all the remaining $(K-1)$ users, the costs of all the users in the system are

$$\begin{aligned} c_k(s^{(n)}, a_1^{(n)}, \dots, a_K^{(n)}) &= c_k(s^{(n)}, a_k^{(n)}) \geq 0 \\ c_{i,i\neq k}(s^{(n)}, a_1^{(n)}, \dots, a_K^{(n)}) &= -\frac{1}{K-1} \cdot c_k(s^{(n)}, a_k^{(n)}). \end{aligned}$$

For notation convenience, in the following sections we will drop the subscript k by defining

$$c(s^{(n)}, a_k^{(n)}) := c_k(s^{(n)}, a_k^{(n)}).$$

Holding cost: Each user has an instantaneous Quality of Service (QoS) constraint denoted as $d_i(s^{(n)}, a_1^{(n)}, \dots, a_K^{(n)})$ where $i = 1, \dots, K$. If the QoS is chosen to be the delay (latency) then $d_i(s^{(n)}, a_1^{(n)}, \dots, a_K^{(n)})$ is a function of the buffer state $b_i^{(n)}$ of the current user i. The instantaneous holding costs will be subsequently included in an infinite horizon latency constraint. Since the transmission latency is independent of the actions of all the remaining users, it can be simplified as

$$d_k(s^{(n)}, a_1^{(n)}, \dots, a_K^{(n)}) = d_k(s^{(n)}, a_k^{(n)}) \geq 0, \quad s^{(n)} \in \mathcal{S}_k. \tag{2}$$

We assume zero-sum holding costs among all the users, which has the following property:

$$\sum_{i=1}^{K} d_i(s^{(n)}, a_1^{(n)}, \dots, a_K^{(n)}) = 0. \tag{3}$$

The zero-sum holding costs can be interpreted as another type of cost caused during transmission.

Due to the equality among all the remaining $(K-1)$ users, their holding costs can be expressed as

$$d_{i,i\neq k}(s^{(n)}, a_k^{(n)}) = -\frac{1}{K-1} \cdot d_k(s^{(n)}, a_k^{(n)}). \tag{4}$$

Drop the subscribe k of the user being scheduled for transmission, and its holding cost can be rewritten in the following way:

$$d(s^{(n)}, a_k^{(n)}) := d_k(s^{(n)}, a_k^{(n)}).$$

C. Switching Control Game and Transition Probabilities

With the above setup, the decentralized transmission control problem in a cognitive radio system is formulated as a switching control game [8]. In such a game, the transition probabilities depend only on the action of the kth user when the state $s \in \mathcal{S}_k$. The transition probabilities of the switching control game can be mathematically written as follows:

$$\begin{aligned} &\mathbb{P}(s^{(n+1)}|s^{(n)}, a_1, \dots, a_K) \\ &= \begin{cases} \mathbb{P}(s^{(n+1)}|s^{(n)}, a_1) & \text{if } s^{(n)} \in \mathcal{S}_1 \\ \mathbb{P}(s^{(n+1)}|s^{(n)}, a_2) & \text{if } s^{(n)} \in \mathcal{S}_2 \\ \dots & \\ \mathbb{P}(s^{(n+1)}|s^{(n)}, a_K) & \text{if } s^{(n)} \in \mathcal{S}_K \end{cases}. \end{aligned}$$

According to the property of the switching control game, when the kth user is scheduled for transmission, the transition probability between the current composite state $s = [h, b, f]$ and the next state $s' = [h', b', f']$ depends only on the action of the kth user a_k. The transition probability function of our problem can now be mathematically expressed by the following equation.

$$\begin{aligned} & \mathbb{P}(s'|s, a_1, a_2, \dots, a_K) \\ = & \mathbb{P}(s'|s, a_k) \\ = & \prod_{i=1}^{K} \mathbb{P}(h_i'|h_i) \cdot \prod_{i=1, i \neq k}^{K} \mathbb{P}(b_i'|b_i) \cdot \mathbb{P}(b_k'|b_k, a_k). \end{aligned}$$

As each user is equipped with a size L buffer, the buffer occupancy of user k evolves according to Lindley's equation [11]

$$b_k^{(n+1)} = \min\left([b_k^{(n)} - a_k^{(n)}]^+ + f_k^{(n)}, L\right). \tag{5}$$

The evolution of the buffer state of user $i = 1, 2, \dots, K$ when $i \neq k$ follows the following rule:

$$b_i^{(n+1)} = \min(b_i^{(n)} + f_i^{(n)}, L).$$

For user k, the buffer state transition probability depends on the distribution of its incoming traffic and its action. Its mathematical expression is

$$\mathbb{P}(b_k'|b_k, a_k) = \begin{cases} \mathbb{P}(f_k = b_k' - [b_k^{(n)} - a_k^{(n)}]^+) & b_k' < L \\ \sum_{x=L-[b_k^{(n)}-a_k^{(n)}]^+}^{\infty} \mathbb{P}(f_k = x) & b_k' = L \end{cases}. \tag{6}$$

For those users who are not scheduled for transmission, the buffer state transition probability depends only on the distribution of incoming traffic, which can be written as

$$\mathbb{P}(b_i'|b_i) = \begin{cases} \mathbb{P}(f_i = b_i' - b_i) & b_i' < L \\ \sum_{x=L-b_i}^{\infty} \mathbb{P}(f_i = x) & b_i' = L \end{cases}. \tag{7}$$

D. TDMA Channel Access Rule

This paper adopts a TDMA cognitive radio system model (IEEE 802.16 [6]). A decentralized channel access algorithm is designed for the TDMA channel. The mechanism of the algorithm is described as follows. In the beginning of every time slot, each user tries to access the channel after a certain time delay t. The value of t is given in equation (8) by applying an opportunistic scheduling algorithm [12]. After the first user accessing the channel, all remaining users will detect the channel occupancy and stop accessing. However, when the spectrum is occupied by a primary user (assume the system has perfect information concerning primary user activities), the channel states of all users $h_k^{(n)}$ will be set to zero, which in turn will lead the value of t^* to be infinity, meaning no user will access the channel during the nth period. The parameters involved in this channel access algorithm are specified as follows:

$$\begin{aligned} t^* &= \arg\min_{k \in \{1,2,\dots,K\}} \frac{\gamma}{b_k^{(n)} h_k^{(n)}} \\ k^* &= \arg\min_{k \in \{1,2,\dots,K\}} \frac{\gamma}{b_k^{(n)} h_k^{(n)}}, \quad \text{if } h_k^{(n)} \neq 0 \\ k^* &= 0, \quad \text{if } h_1^{(n)} =, \dots, = h_k^{(n)} = 0, \end{aligned} \tag{8}$$

where k^* is the user scheduled for transmission, and t^* is the time delay of that user. γ is a system parameter which is determined by the system. $b_k^{(n)}$ and $h_k^{(n)}$ are two elements affecting the scheduling algorithm. $b_k^{(n)}$ evaluates the package delay of each user at time n, and users with longer delay are preferred to be selected to transmit. $h_k^{(n)}$ is the channel quality parameter, and user with better channel quality has better a chance of being scheduled for transmission.

E. Switching Controlled Markovian Game Formulation

At time instant n, assume user k is scheduled for transmission according to the system access rule specified in (8). We define $c(s^{(n)}, a_k^{(n)}) \geq 0$ to be the instantaneous cost of user k when the system is in state $s^{(n)}$. We use $\Phi_1, \Phi_2, \dots, \Phi_K$ to represent the set of all pure policies of each user, respectively. The infinite horizon expected total discounted cost[1] of any user i given his transmission policy π_i ($\pi_i \in \Phi_i$), can be written as

$$C_i(\pi_i) = \mathbb{E}_{\pi_i}\left[\sum_{n=1}^{\infty} \beta^{n-1} \cdot c_i(s^{(n)}, a_k^{(n)})\right] \tag{9}$$

where $0 \leq \beta < 1$ is the discount factor, the state $s^{(n)} \in \mathcal{S}_k$ and the expectation is taken over action $a_k^{(n)}$ as well as system state $s^{(n)}$ evolution for $n = 1, 2, \dots$. We aim to find an optimal policy π_i^* that can minimize the overall expected discounted cost of the ith user subject to its constraint. We can mathematically formulate the design of π_i^* as the following optimization problem.

Optimization Problem 1: The ith user has a discounted cost defined in (9) and a latency constraint shown in (11). Assume at time slot n, the system schedules user k to transmit. The holding cost of user i at that time is $d_i(s^{(n)}, a_k^{(n)})$. $\widetilde{D}_i$ is a system parameter depending on the system requirement of user i. Consider an infinite time horizon, the optimal policy of user i is chosen to optimize the discounted cost subject to its constraint, as follows:

$$\min_{\pi_i \in \Phi_i} \quad C_i(\pi_i) \tag{10}$$

$$s.t. \quad D(\pi_i) = \mathbb{E}_{\pi_i}\left[\sum_{n=1}^{\infty} \beta^{n-1} \cdot d_i(s^{(n)}, a_k^{(n)})\right] \leq \widetilde{D}_i \tag{11}$$

□

In this game, every user tries to minimize their cost by choosing the optimal action when that user accesses the channel. In the next section, we will prove the uniqueness of the Nash equilibrium value vector in this game.

III. Randomized Threshold Nash Equilibrium for Dynamic Markovian Game

This section first state the Nash equilibrium in a zero-sum dynamic Markovian switching control game is available and all the Nash equilibria have a unique value vector. Then a structural result on the Nash equilibrium policy is presented. Finally, a computationally efficient stochastic approximation algorithm is proposed to search for the Nash equilibrium.

[1] There are two criteria for evaluating the cost of a Markov decision process (MDP): expected average cost criterion and expected total discounted reward criterion. The reason we choose discounted cost criterion is because it is mathematically simpler than the average cost criterion. And the average cost criterion has several technicalities with obtaining the stationary optimal policies. In addition, if the stationary policy exists, when the discounted factor $\beta \to 1$, the policy obtained under the average cost criterion is the same from that of the discounted cost criterion, $C_{avg} = \lim_{\beta \to 1}(1-\beta) \cdot C_{dis}$. Here, C_{avg}, C_{dis} denote the average cost and discounted cost, respectively.

A. The Existence and Optimality of The Nash Equilibrium

Shapley's Theorem says a discounted, zero-sum, stochastic game possesses a value vector that is the unique solution for the game [8]. By applying this result to our system, we can introduce the following theorem.

Theorem 1: The Nash equilibrium of the constrained switching control Markovian game with zero-sum transmission cost (1) and zero-sum latency constraint (3) is able to be obtained by a value iterative optimization algorithm [8]. All the Nash equilibria obtained converge to a unique value vector. This ensures the global optimality of the Nash equilibrium policy obtained. □

B. Structural Result on Randomized Threshold Policy

First, we list three assumptions. Based on these three assumptions, Theorem 2 is introduced.

- *A1:* The set of policies that satisfy constraint (11) is non-empty, to ensure the delay constraint of the system is valid.
- *A2:* Transmission cost $c(s, a_k)$ and holding cost $d(s, a_k)$ are submodular[2] functions of b_k, a_k for any channel quality h_k of the current user, and are independent of the incoming traffic f_k. And they are also nondecreasing functions on b_k for any h_k, f_k and a_k.
- *A3:* $\sum_{b'_k=l}^{L} \mathbb{P}(b'_k|b_k, a_k)$ is a submodular function on b_k, a_k and nondecreasing in b_k for any l and a_k .

Theorem 2: Represent the optimal policy to Optimization Problem 1 with constraint (11) is as $\pi_k^*(s)$. If assumptions A1-A3 hold, then the optimal policy $\pi_k^*(s)$ is a randomization of two pure policies $\pi_k^1(s)$ and $\pi_k^2(s)$. Each of these two pure policies is a nondecreasing function with respect to buffer occupancy state b_k (defined in Section II.A).
The proof of Theorem 2 is available in [13]. □

C. Verification of Assumptions in Theorem 2

This subsection will verify the assumptions listed in Section III.B. The model we consider is based on a TDMA cognitive radio system. The cost of the switching control dynamic game is evaluated by transmission BER and the constraint is defined to be the transmission delay. The channel model is assumed to be complex Gaussian random variables which has zero mean.

First, we define the holding cost caused by user k when the state is $s \in \mathcal{S}_k$,

$$d(s, a_k) = \frac{b_k}{\overline{f}}. \tag{12}$$

Here $\overline{f}$ is the average number of incoming packets, which is parameter of the system. Assumption A1 holds if there exist an action such that $a_k > \overline{f}$.

In our system, we choose the transmission cost to be the BER which is function of the channel quality. Assume the channel states are quantized by quantization threshold parameters $\Gamma(h)_1, \Gamma(h)_2, \ldots$ which are selected by the system. System transmission cost $BER(\gamma, a_k)$ depends on the random channel gain $\gamma \in [\Gamma(h)_{i-1}, \Gamma(h)_i)$. Therefore, the transmission cost is

$$BER_k^i(h_k, a_k) = \frac{\int_{\Gamma(h)_{i-1}}^{\Gamma(h)_i} BER(\gamma, a_k) g(\gamma) d\gamma}{\int_{\Gamma(h)_{i-1}}^{\Gamma(h)_i} g(\gamma) d\gamma} \tag{13}$$

$$BER_k(\gamma, a_k) = 0.2 \times \exp\left[\frac{-1.6\gamma}{(2^{a_k} - 1)}\right], \tag{14}$$

where $g(\gamma)$ denotes the probability distribution of signal-to-noise ratio (SNR), and the expectation is taken over the γ when the channel state h_k belongs to quantization region $[\Gamma(h)_{i-1}, \Gamma(h)_i)$. The BER approximation expression in (14) is from [14]. When the system uses uncoded M-ary quadrature modulation (QAM), where $M = 2^{a_k}$.

By substituting (14) into (13), it can be seen that the averaged transmission cost is independent of the buffer occupancy b_k. It is obvious from (12) that the holding cost function $d(s, a_k)$ is increasing in b_k and independent of a_k. Thus, assumption A2 holds.

The buffer state occupancy evolves according to Lindley's recursion equation (5). Given the current state buffer occupancy, and actions the user takes, the transition probability to the next state buffer occupancy depends on the probability of incoming traffic, as shown in (6). Then the buffer state transition probability can be rewritten as $\mathbb{P}(b'_k|b_k, a_k) = \mathbb{P}(b'_k|(b_k - a_k))$. Assume the incoming traffic is evenly distributed within the range of 0 to the buffer size L. This can be mathematically written as $\mathbb{P}(f_k < 0 \ or \ f_k > L) = 0$ and $\mathbb{P}(0 \leq f_k \leq L) = \frac{1}{L+1}$. The buffer state transition probability can now be given by

$$\mathbb{P}(b'_k|b_k, a_k) = \begin{cases} \frac{1}{L+1} & b'_k < L \\ \frac{1+[b_k^n - a_k^n]^+}{L+1} & b'_k = L \end{cases}. \tag{15}$$

Therefore, the buffer state occupancy transition probability is independent from b_k and a_k when $b'_k < L$ while it is first order stochastically increasing in $b_k - a_k$ when $b'_k = L$. This result verifies $\sum_{b'_k=l}^{L} \mathbb{P}(b'_k|b_k, a_k)$ is nondecreasing in b_k in A3. According to (15) we can see that $\mathbb{P}(b'_k = L|b_k, a_k)$ is submodular in b_k, a_k, thus A3 holds.

D. Stochastic Approximation Algorithm

Based on the structural result on the optimal policy stated in Theorem 2, we will introduce a *stochastic approximation algorithm* in this section. This algorithm can efficiently reduce the complexity of finding the optimal policies. The search domain is the set of optimal policies $\pi_k^*(s)$ described in Theorem 2.

When there are two actions available with action set $a_k = \{0, 1\}$. An optimal policy $\pi_k^*(s)$ can be defined by three parameters $b_1(s), b_2(s)$ and p. $\pi_k^*(s)$ is chosen to be 0 when $0 \leq b_k < b_1(s)$, it is p when $b_1(s) \leq b_k < b_2(s)$ and 1 when $b_2(s) \leq b_k$. Here, p is the randomization factor, $b_1(s)$ and

[2]A function $f : \mathcal{A} \times \mathcal{B} \times \mathcal{C} \rightarrow \mathcal{R}$ is submodular in (a, b) for any fixed $c \in \mathcal{C}$. Then for all $a' \geq a$ and $b' \geq b$, $f(a', b'; c) - f(a, b'; c) \leq f(a', b; c) - f(a, b; c)$ holds.

$b_2(s)$ are the lower buffer state threshold and higher buffer state threshold, respectively. The search for each optimal policy problem is now converted to the estimation of these three parameters. The total number of parameters to be estimated for the whole system is $3 \times |h_k| \times K$. $|\cdot|$ denotes the cardinality of a set. We first compute the continuous optimal values of $b_1(s)$ and $b_2(s)$, then they are rounded off to the nearest discrete values. This is a relaxation of the original discrete stochastic optimization problem as the buffer state in the problem setup is discrete.

The Simultaneous perturbation stochastic approximation (SPSA) method [15] is adopted to estimate the parameters. SPSA is especially efficient in high-dimensional problems in terms of providing a good solution for a relatively small number of measurements for the objective function. The essential feature of SPSA is the underlying gradient approximation, which requires only two objective function measurements per iteration regardless of the dimension of the optimization problem. These two measurements are made by simultaneously varying, in a properly random fashion all of the variables in the problem. The detailed algorithm is described in Algorithm 1.

The first part of the algorithm initializes system variables.

Algorithm 1 Stochastic Approximation Algorithm

1: **Initialization**: $\theta^{(0)}$, λ^0; $n = 0$; $\rho = 4$;
2: Initialize constant perturbation step size β and gradient step size α;
3: **Main Iteration**
4: **if** $s^{(n)} \in \mathcal{S}_k$ **then**
5: $m_k = 3 \times |h_k|$;
6: Generate $\Delta^{(n)} = [\Delta_1^{(n)}, \Delta_2^{(n)}, \ldots, \Delta_{m_k}^{(n)}]^T$; $\Delta_i^{(n)}$ are Bernoulli random variables with $p = \frac{1}{2}$.
7: $\theta_{k+}^{(n)} = \theta_k^{(n)} + \beta \times \Delta^{(n)}$;
8: $\theta_{k-}^{(n)} = \theta_k^{(n)} - \beta \times \Delta^{(n)}$;
9: $\Delta C_k^{(n)} = \frac{c(s^{(n)},\theta_{k+}^{(n)}) - c(s^{(n)},\theta_{k-}^{(n)})}{2\beta}[(\Delta_1^{(n)})^{-1}, \ldots, (\Delta_{m_k}^{(n)})^{-1}]^T$;
10: $\Delta D_k^{(n)} = \frac{d(s^{(n)},\theta_{k+}^{(n)}) - d(s^{(n)},\theta_{k-}^{(n)})}{2\beta}[(\Delta_1^{(n)})^{-1}, \ldots, (\Delta_{m_k}^{(n)})^{-1}]^T$;
11: $\theta_k^{(n+1)} = \theta_k^{(n)} - \alpha \times \left(\Delta C_k^{(n)} + \Delta D_k^{(n)} \cdot \max\left[0, \lambda_k^{(n)} + \rho \cdot \left(D(s^{(n)}, \theta_k^{(n)}) - \widetilde{D_k}\right)\right]\right)$;
11: $\lambda_k^{(n+1)} = \max\left[(1 - \frac{\alpha}{\rho} \cdot \lambda_k^{(n)}), \lambda_k^{(n)} + \alpha \cdot \left(D(s^{(n)}, \theta_k^{(n)}) - \widetilde{D_k}\right)\right]$;
12: **end if**
13: The parameters of other users remain unchanged;
14: $n = n + 1$;
15: The iteration terminates when the values of the parameters $\theta^{(n)}$ converge; else return back to Step 3.

$\theta^{(n)}$ represents the union of all the parameters we search for at the nth time slot and $\theta_k^{(n)}$ indicates parameters of the kth user. β and α denote the constant perturbation step size and constant gradient step size, respectively. In the main part of the algorithm, SPSA algorithm is applied to iteratively update system parameters. When the kth user is scheduled to transmit at time slot n, parameters $\theta_k^{(n)}$ and the Lagrangian constant $\lambda_k^{(n)}$ can be updated after introducing a random perturbation vector $\Delta^{(n)}$. In the meanwhile, the parameters of the other users remain unchanged. The algorithm terminates when the parameters $\theta^{(n)}$ converge.

Theorem 3: $\{\theta^n(\alpha), \lambda^n(\alpha)\}$ are system parameters generated by the stochastic approximation algorithm 2. Define the piecewise constant interpolated continuous-time processes of $\{\theta^n(\alpha), \lambda^n(\alpha)\}$ to be $\{\theta^t(\alpha), \lambda^t(\alpha)\}$. When t is within the range of $[n\alpha, (n+1)\alpha)$, the value of $\theta^t(\alpha)$ is set to be $\theta^n(\alpha)$ and $\lambda^t(\alpha)$ is set to be $\lambda^n(\alpha)$. The mathematical expressions are given as follows:

$$\begin{aligned} \theta^t(\alpha) &= \theta^n(\alpha) \qquad t \in [n\alpha, (n+1)\alpha), \\ \lambda^t(\alpha) &= \lambda^n(\alpha) \qquad t \in [n\alpha, (n+1)\alpha). \end{aligned}$$

[16], [17] For sufficient large ρ, as $\alpha \to 0$ and $t \to \infty$, $\{\theta^t(\alpha), \lambda^t(\alpha)\}$ converge in probability to the KT pair of (10,11) which is specified in (16). □

The optimal policies of (10,11) which satisfy the Kuhn Tucker (KT) condition can be defined as follows. π_i^* belongs to the KT set when

$$\begin{aligned} &KT = \{\pi_i^* : \exists \lambda_i > 0, such\ that \\ &\nabla_{\pi_i} C_i + \nabla_{\pi_i} \lambda_i (D_i - \widetilde{D}_i) = 0, i = 1, \ldots, K\}, \quad (16) \end{aligned}$$

where C_i and D_i are the optimization objective (9) and delay constraint (11) respectively. Moreover, π_i^* satisfies the second order sufficiency conditions: $\nabla^2_{\pi_i} C_i + \nabla^2_{\pi_i} (D_i - \widetilde{D}_i) \geq 0$ is positive definite for all the i, and $(D_i - \widetilde{D}_i) = 0$, $\lambda_i > 0$, $i = 1, \ldots, K$.

Note here that in the stochastic approximation algorithm, we first compute the continuous values of $b_1(s)$ and $b_2(s)$, then they are rounded off to the nearest discrete values. This relaxation leads to the continuous value of π_i during calculation, thus, (16) is differentiable on π_i.

IV. Numerical Examples

This section presents a numerical example of the Nash equilibrium transmission policy using the stochastic approximation algorithms proposed in Section III. The channel quality measurements are quantized into two different states, $\{1, 2\}$. In the models used, each user has a size 10 buffer. In the system configuration, the transmission costs, the holding costs and buffer transition probability matrices are chosen to ensure A2-A3 specified in Section III.B. The channel transition probability matrices are generated randomly.

It considers a system with 2 different secondary users, with each user having 2 different action choices $\{0, 1\}$. As it is a constrained switching controlled Markovian game, the Nash equilibrium policy is a randomization of two pure policies. Each optimal transmit policy can be determined by three parameters, namely, lower threshold $b_l(s)$, upper threshold $b_h(s)$ and randomization factor p. The stochastic approximation algorithm is applied to find the Nash equilibrium policy, the simulation results of user 1 with $h_2 = 1, b_2 = 1$ are shown in Fig. 4. The figure shows that the optimal

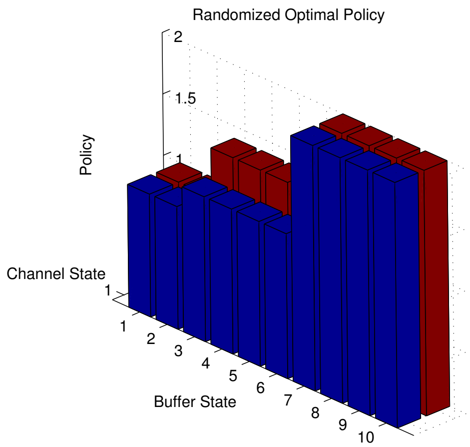

Fig. 1. The Nash equilibrium transmission control policy obtained via stochastic approximation algorithm. A 2 users system is considered where each has a size 10 buffer.

transmission policies are no longer deterministic but are a randomization of two pure policies, and that each Nash equilibrium policy is monotone increasing on the buffer state.

V. Conclusions

We formulate the secondary users rate adaptation problem in a cognitive radio system as a constrained zero-sum switching control dynamic Markovian game. It is shown that the Nash equilibria of the game are always available and have a unique value vector. Based on few assumptions, the optimal policy is a randomization of two pure monotone policies. This allows us to propose a more computationally efficient stochastic approximation algorithm. Numerical example is provided to verify these results.

References

[1] J. Mitola III, “Cognitive radio for flexible mobile multimedia communications,” in *proceedings of IEEE Mobile Multimedia conference*, November 1999, pp. 33-10.

[2] S. Haykin, “Cognitive radio: Brain-empowered wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201-220, Feburary, 2005.

[3] W. Yu, G. Ginis, and J. M. Cioffi, “Distributed multiuser power control for digital subscriber lines,” *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1105-1115, June 2002.

[4] N. Nie and C. Comaniciu, “Adaptive channel allocation spectrum etiquette for cognitive radio networks,” in *Proceedings of IEEE DySPAN*, November 2005, pp. 269-278.

[5] M. Felegyhazi, M. Cagalj, S. S. Bidokhti, and J. Hubaux, “Noncooperative multi-radio channel allocation in wireless networks,” in *Proceedings of IEEE INFOCOM*, May 2007, pp. 1442-1450.

[6] IEEE, “IEEE std 802.16-2004.” IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, 2004.

[7] S. R. Mohan and T. E. S. Raghavan, “An algorithm for discounted switching control stochastic games,” *OR Spektrum*, vol. 9, no. 1, pp. 41-45, March 1987.

[8] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. New York: Springer-Verlag, 1997.

[9] E. Altman, “Constrained Markov Decision Processes”. London: Chapman and Hall, 1999.

[10] F. Fu, T. M. Stoenescu, and M. van der Schaar, “A pricing mechanism for resource allocation in wireless multimedia applications,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 2, pp. 264-279, August 2007.

[11] D. Djonin and V. Krishnamurthy, “MIMO transmission control in fading channels - a constrained markov decision process formulation with monotone randomized policies” *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 5069-5083, October 2007.

[12] A. Farrokh and V. Krishnamurthy, “Opportunistic scheduling for streaming users in HSDPA multimedia systems,” *IEEE Transactions on Multimedia Systems*, vol. 8, no. 4, pp. 844-855, August 2006.

[13] J. W. Huang and V. Krishnamurthy, “Transmission control in cognitive radio as a markovian dynamic game - structural result on randomized threshold policies,” *submitted to IEEE Transactions on Communications*, revised in July 2008.

[14] S. T. Chung and A. J. Goldsmith, “Degrees of freedom in adaptive modulation: a unified view,” *IEEE Transaction on Communications*, vol. 49, no. 9, pp. 1561-1571, September 2001.

[15] J. C. Spall, *Introduction to Stochastic Search and Optimization*, ser. Wiley-Interscience series in Discrete Mathematics and Optimization. Wiley-Interscience, 2003.

[16] F. Vazquez Abad and V. Krishnamurthy, “Constrained stochastic approximation algorithms for adaptive control of constrained Markov decision processes,” in*Proceedings of 42nd IEEE Conference on Decision and Control*, December 2003, pp. 2823-2828.

[17] V. Krishnamurthy, K. Martin, and F. Vazquez Abad,“Implementation of gradient estimation to a constrained Markov decision problem,” in *Proceedings of 42nd IEEE Conference on Decision and Control*, 2003, pp. 4841-4846.