Delay Sensitive Communications over Cognitive Radio Networks

Feng Wang, Jianwei Huang, *Senior Member, IEEE*, and Yuping Zhao

***Abstract*—Supporting the quality of service of unlicensed users in cognitive radio networks is very challenging, mainly due to the dynamic resource availability induced by the licensed users' activities. In this paper, we derive the optimal admission control and channel allocation decisions in cognitive overlay networks to support delay sensitive communications of unlicensed users. We formulate it as a Markov decision process problem, and solve it by transforming the original formulation into a stochastic shortest path problem. We then propose a simple heuristic control policy, which includes a threshold-based admission control scheme and and a largest-delay-first channel allocation scheme, and prove the optimality of the largest-delay-first channel allocation scheme. We further propose an improved policy using the rollout algorithm. By comparing the performance of both proposed policies with the upper-bound of the maximum revenue, we show that our policies achieve close-to-optimal performances with low complexities.**

***Index Terms*—Admission control, Markov decision process, Bellman's equation, rollout algorithm, cognitive radio networks, spectrum overlay.**

I. INTRODUCTION

COGNITIVE radio technology has the potential to significantly improve spectrum utilization and accommodate many more devices in the limited spectrum. Supporting Quality of Service (QoS), however, is challenging in cognitive radio networks due to the dynamically changing network resources. In this paper, we will design an admission control and channel allocation mechanism to support delay-sensitive real-time secondary unlicensed communications. Compared with the resource allocation in conventional communication networks, the unique challenge here is to incorporate the impact of primary licensed users on the availability of the communication resources.

Optimal channel selection of a *single* secondary unlicensed user has been well studied in the literature (*e.g.,* [2], [3]). Zhao *et al.* [2] considered the total expected reward maximization problem when the secondary user can only sense one channel at a time. Liu *et al.* [3] further considered the case where the secondary user can sense multiple channels simultaneously. The resource allocation problem becomes more complicated when there are multiple secondary users (*e.g.,* [4], [5]). Zhou *et al.* [4] jointly considered channel allocation with power control. Urgaonkar and Neely [5] developed opportunistic scheduling policies to provide performance guarantees.

Admission control is critical for supporting QoS when there are too many users that want to access the network simultaneously. In traditional cellular networks, many results have shown that the optimal admission control policy has a threshold structure (*e.g.,* [6]–[8]). In cognitive radio networks, researchers have studied admission control for both underlay networks (*e.g.*, [9]–[11]) and overlay networks (*e.g.,* [12], [13]). In cognitive overlay networks, admission control is often jointly pursued with channel allocation, as the secondary users can only access idle channels not occupied by primary users. Admission control also can be jointly considered with other mechanisms, *e.g.*, Kim and Shin [12] considered joint optimal admission and eviction control using semi-Markov decision process and linear programming. Mutlu *et al.* [13] investigated the problem of optimal spot pricing of spectrum for maximizing the profit from the admission of secondary users.

In this paper, we consider the joint admission control and channel allocation problem for cognitive overlay networks. Our problem is very different from the throughput maximization for elastic data traffic studied in most previous literature [4], [5]. We want to support the secondary users' real-time applications (*e.g.,* VoIP and video streaming) with stringent delay constraints.

The rest of the paper is organized as follows. We describe the system model in Section II, and formulate the admission control and channel allocation problem as a Markov Decision Process (MDP) in Section III. In Section IV, we transform the problem into a stochastic shortest path problem and prove the convergence of the Bellman's equation. Section V proposes a heuristic control policy and an improved rollout policy, together with the corresponding theoretical analysis and simulation results. We finally conclude in Section VI.

II. SYSTEM MODEL

This paper studies a cognitive radio network as shown in

Manuscript received January 20, 2011; revised June 30 and November 27, 2011; accepted December 15, 2011. The associate editor coordinating the review of this paper and approving it for publication was C.-F. Chiasserini.

F. Wang is with the Beijing Space Technology Development and Test Center, China Academy of Space Technology, Beijing 100094, China. He was with the State Key Laboratory of Advanced Optical Communication Systems & Networks, Peking University, Beijing 100871, China (e-mail: fengwangpku@gmail.com).

J. Huang (corresponding author) is with the Department of Information Engineering, the Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: jwhuang@ie.cuhk.edu.hk).

Y. Zhao is with the State Key Laboratory of Advanced Optical Communication Systems & Networks, Peking University, Beijing 100871, China (e-mail: yuping.zhao@pku.edu.cn).

Part of the work was done when F. Wang visited the Chinese University of Hong Kong between July and December, 2009. Part of the results was presented at IEEE GLOBECOM 2010 [1]. The authors acknowledge the contribution of J. Zhu in [1].

This work is supported by the General Research Funds (Project Number 412710 and 412511) established under the University Grant Committee of the Hong Kong Special Administrative Region, China, and by Important National Science and Technology Specific Projects of China (Project Number 2009ZX03003-011-01).

Digital Object Identifier 10.1109/TWC.2012.020812.110133

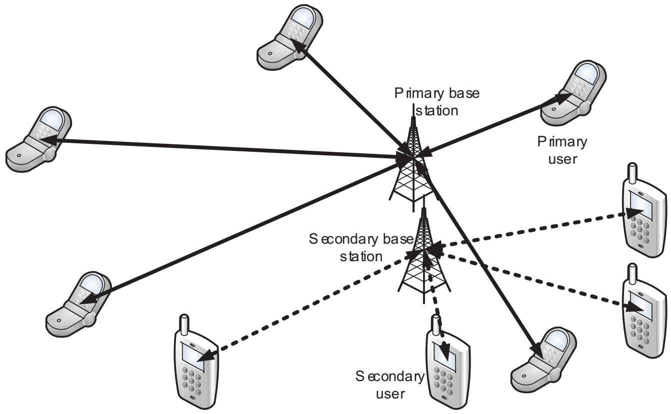

Fig. 1. A cognitive radio network scenario. In the secondary network, the dotted arrows denote the channels between the secondary base station and the secondary users.

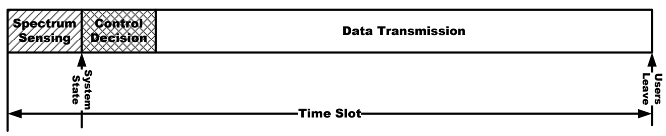

Fig. 2. The components of a time slot.

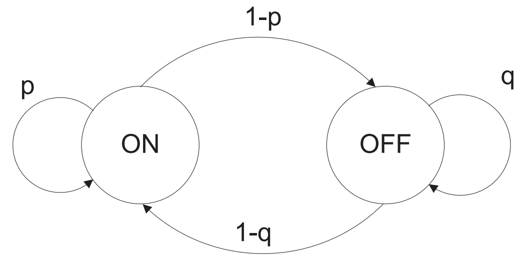

Fig. 3. Markovian ON/OFF model of channel activities.

Fig. 1. We consider an infrastructure-based secondary unlicensed network, where a secondary network operator senses the channel availabilities (*i.e.,* primary licensed users' activities) and decides the admission control and channel allocation for the secondary users. A similar network architecture has been considered in several recent literature (*e.g.,* [14]–[17]). Comparing with the distributed network architecture where end users need to perform spectrum sensing individually, the network architecture considered in this paper has the advantage of reducing the complexity of the secondary user devices and providing better QoS support. Such infrastructure-based network without user sensing requirement is also consistent with the recent ruling of FCC (Federal Communications Commission) on the TV white space sharing [18].

One way to realize network-based spectrum sensing is to construct a sensor network that is dedicated to sensing the radio environment in space and time [19]. The secondary base station will collect the sensing information from the sensor network and provide it to the unlicensed users, which is called "sensing as service". There has been significant current research efforts along this direction in the context of an European project SENDORA [20], which aims at developing techniques based on sensor networks for supporting coexistence of licensed and unlicensed wireless users in a same area.

In our model, the time is divided into equal length slots. Primary users' activities remain roughly unchanged within a single time slot. This means that it is enough for the operator to sense once at the beginning of each time slot (see Fig. 2). For readers who are interested in the optimization of the time slot length to balance sensing and data transmission, see [21].

The network has a set $\mathcal{J} = \{1, \ldots, J\}$ of orthogonal primary licensed channels. The state of each channel follows a Markovian ON/OFF process as in Fig. 3. If a channel is "ON", then it means that the primary user is not active on the channel and the channel condition is good enough to support the transmission rate requirement of a secondary user. Here we assume that all secondary users want to achieve the same target transmission rate (*e.g.,* that of a same type of video streaming application). If a channel is "OFF", then either a primary user is active on this channel, or the channel condition is not good enough to achieve the secondary user's target rate. In the time slotted system, the channel state changes from "ON" to "OFF" ("OFF" to "ON", respectively) between adjacent time slots with a probability p (q, respectively). When a channel is "ON", it can be used by a secondary unlicensed user.

We consider an infinitely backlog case, where there are many secondary users who want to access the idle channels. Each idle channel can be used by at most one secondary user at any given time. A secondary user represents an unlicensed user communicating with the secondary base station as shown in Fig. 1. The secondary users are interested in real-time applications such as video streaming and VoIP, which require steady data rates with stringent delay constraints. The key QoS parameter is the *accumulative delay*, which is the total delay that a secondary user experiences after it is admitted into the system. Once a secondary user is admitted into the network, it may finish the session *normally* with a certain probability. However, if the user experiences an accumulative delay larger than a threshold, then its QoS significantly drops (*e.g.,* freezing happens for video streaming) and the user will be *forced to terminate*.

To make the analysis tractable, we make several assumptions. First, we assume that the availabilities of all channels follow the same Markovian model. This is reasonable if the traffic types of different primary users are similar (*e.g.,* all primary users are voice users). Second, we assume that all secondary users experience the same channel availability independent of their locations. This is reasonable when the secondary users are close-by. Third, we assume the spectrum sensing is error-free. This can be well approximated by having enough sensors performing collaborating sensing. Furthermore, we assume that all channels are homogeneous and can provide the same data rate to any single secondary user using any channel. Finally, we assume that all secondary users are homogeneous (*i.e.,* interested in the same application such as video streaming). Each secondary user only requires one available channel to satisfy its rate requirement. Several of the above assumptions can be relaxed by increasing the state space of the MDP formulation. As we will see shortly, the admission control and channel allocation issue in this homogeneous case is already quite complicated and admits no closed-form solutions. The analysis and insights of this paper will enable us to further consider heterogeneous channels and secondary users in the future.

III. Problem Formulation

We formulate the admission control and channel allocation problem as an MDP [22]. In an infinite-horizon MDP with a set of finite states $\mathcal{S}$, the state evolves through time according to a transition probability matrix $\{P_{x_k x_{k+1}}\}$, which depends on both the current state and the control decision from a set $\mathcal{U}$. More specifically, if the network is in state x_k in time slot k and selects a decision $u(x_k) \in \mathcal{U}(x_k)$, then the network obtains a revenue $g(x_k, u(x_k))$ in time slot k and moves to state x_{k+1} in time slot $k+1$ with probability $P_{x_k x_{k+1}}(u(x_k))$. We want to maximize the long-term time average revenue, *i.e.,*

$$\lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{k=0}^{T-1} g(x_k, u(x_k))\right\}. \tag{1}$$

A. The State Space

The system state describes system information after the network performs spectrum sensing at the beginning of the time slot (see Fig. 2). It consists of two components:

- A *channel state component*, $m = \boldsymbol{a}^T \cdot \boldsymbol{a}$, describes the number of available channels. Here $\boldsymbol{a} = (a_j, \forall j \in \mathcal{J})$ is the channel availability vector, where $a_j = 1$ (or 0) when channel j is available (or not).
- A *user state component*, $\boldsymbol{\omega_e} = (\omega_{e,i}, \forall i \in \mathcal{D})$, describes the numbers of secondary users with different accumulative delays. Here $\mathcal{D} = \{0, 1, \ldots, D_{\max}\}$ is the set of possible delays, and $\omega_{e,i}$ denotes the number of secondary users whose accumulative delay is i.

We let $\mathcal{M}$ denote the feasible set of the channel state component, and Ω denote the feasible set of the user state component. The state space is given by $\mathcal{S} = \{(m, \boldsymbol{\omega_e}) | m \in \mathcal{M}, \boldsymbol{\omega_e} \in \Omega\}$.

State θ is said to be *accessible* from state η if and only if it is possible to reach state θ from η, *i.e.,* $P\{reach\ \theta | start\ in\ \eta\} > 0$ [23]. Two states that are accessible to each other are said to be able to *communicate* with each other. In our formulation, all the states in space $\mathcal{S}$ are accessible from state $\mathbf{0}$, which is defined as a state where there is no available channel and no single admitted secondary user in the system. Since it is possible to have $m = 0$ in several consecutive time slots (when primary traffic is heavy and occupies all channels), thus state $\mathbf{0}$ is accessible from any state in the state space $\mathcal{S}$. Hence, all the states communicate with each other and the Markov chain is *irreducible*. Finally, the state space is finite, so all the states are *positive recurrent* [23]. This property turns out to be critical for the analysis in Section IV.

B. The Control Space

For the state $x_k = \{m, \boldsymbol{\omega_e}\} \in \mathcal{S}$ in each time slot k, the set of available control choices $\mathcal{U}(x_k)$ depends on the relationship between the channel state and the user state. The control vector $u(x_k) = \{u_a, \boldsymbol{u_e}\}$ consists of two parts: scalar u_a denotes the number of admitted new secondary users, and vector $\boldsymbol{u_e} = \{u_{e,i}, \forall i \in \mathcal{D}\}$ denotes the numbers of secondary users who are allocated channels and have accumulative delays of $i \in \mathcal{D}$ at the beginning of the current time slot. Without loss of generality, we assume $0 \leq u_a \leq J$, *i.e.,* we will never admit more secondary users than the total number of channels. This leads to $0 \leq u_{e,0} \leq \omega_{e,0} + u_a$, $0 \leq u_{e,i} \leq \omega_{e,i}$ for all $i \in [1, D_{\max}]$, and $0 \leq \sum_{i=0}^{D_{\max}} u_{e,i} \leq m$. Since $m \leq J$, the cardinality of the control space $\mathcal{U}$ is $J^{D_{\max}+2}$.

C. The State Transition

Current state $x_k = \{m, \boldsymbol{\omega_e}\} \in \mathcal{S}$ together with the control $u(x_k) \in \mathcal{U}(x_k)$ determine the probability of reaching the next state $x_{k+1} = \{m', \boldsymbol{\omega'_e}\}$.

First, the transition of channel state component from m to m' depends on the underlying primary traffic. We can divide m' available channels into two groups: one group contains m'_1 channels which are available in the (current) time slot k, the other group contains m'_2 channels which are not available in time slot k. Let us define the set $\mathcal{Z} = \{(m'_1, m'_2) | m' = m'_1 + m'_2, 0 \leq m'_1 \leq m, 0 \leq m'_2 \leq J - m\}$. Then we can calculate the probability based on the i.i.d. ON/OFF model in Section II:

$$P_{mm'} = \sum_{(m'_1, m'_2) \in \mathcal{Z}} \left\{ \binom{m}{m'_1} p^{m'_1} (1-p)^{m-m'_1} \binom{J-m}{m'_2} (1-q)^{m'_2} q^{J-m-m'_2} \right\}. \tag{2}$$

Thus the channel transition function is $f_s(m) = m'$ with probability $P_{mm'}$ for all $m' \in \mathcal{M}$.

Let us define $\boldsymbol{\omega_c} = \{\omega_{c,i}, \forall i \in \mathcal{D}\}$ as the number of secondary users who normally complete their connections (not due to delay violation) in time slot k. For example, a user may terminate a video streaming session after the movie finishes, or terminate a VoIP session when the conversation is over. If we assume that all users have the same completion probability P_f per slot when they are actively served, then the event of having ρ out of τ users completing their connections (denoted as $f_c(\tau) = \rho$) happens with probability $\binom{\tau}{\rho} P_f^{\rho} (1 - P_f)^{\tau-\rho}$.

Finally, define ω_q as the number of secondary users who are forced to terminate their connections during time slot k. The state transition can be written as

$$\begin{cases} m' = f_s(m), \\ \omega_{c,i} = f_c(u_{e,i}), \forall i \in \mathcal{D}, \\ \omega_q = \omega_{e,D_{\max}} - u_{e,D_{\max}}, \\ \omega'_{e,0} = u_{e,0} - \omega_{c,0}, \\ \omega'_{e,1} = u_{e,1} + (\omega_{e,0} + u_a - u_{e,0}) - \omega_{c,1}, \\ \omega'_{e,i} = u_{e,i} + (\omega_{e,i-1} - u_{e,i-1}) - \omega_{c,i}, \forall i \in [2, D_{\max}]. \end{cases} \tag{3}$$

Let us take a network with $J = 10$ and $D_{max} = 2$ as a numerical example. In a particular time slot, assume that there are $m = 7$ channels available and a total of 6 secondary users admitted in the system: 1 user with zero accumulative delay, 3 users with 1 time slot of accumulative delay, and 2 users with 2 time slots of accumulative delay. Then the state vector is $\{m, \boldsymbol{\omega_e}\} = \{7, (1, 3, 2)\}$. Assume the control decision is to admit 2 new users and to allocate available channels to the users except one of the new users, *i.e.,* $u = \{u_a, \boldsymbol{u_e}\} = \{2, (2, 3, 2)\}$. Thus if there is no user completing a connection in the current time slot and $m' = 4$ available channels in the next time slot, the system state becomes $\{m', \boldsymbol{\omega_e'}\} = \{4, (2, 4, 2)\}$.

D. The Objective Function

Our system optimization objective is to choose the optimal control decision for each possible state to maximize the expected average revenue per time slot (also called stage), *i.e.,*

$$\max \lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{k=0}^{T-1} g(x_k, u(x_k))\right\}. \tag{4}$$

Here the revenue function is computed at the end of each time slot k as follows:

$$g(x_k, u(x_k)) = R_c \sum_{i=0}^{D_{\max}} \omega_{c,i}(k) + R_t \sum_{i=0}^{D_{\max}} \omega_{e,i}(k) - C_q \omega_q(k), \tag{5}$$

where $R_c \geq 0$ is the reward of completing the connection of a secondary user normally (without violating the maximum delay constraints), $R_t \geq 0$ is the reward of maintaining the connection of a secondary user, and $C_q \geq 0$ is the penalty of forcing to terminate a connection. By choosing different values of R_c, R_t, and C_q, a network designer can achieve different objective functions. In this paper, we assume that the values of R_c, R_t, and C_q are given parameters.

IV. Analysis of the MDP Problem

We define a sequence of control actions as a policy, $\boldsymbol{\mu} = \{u(x_0), u(x_1), \cdots\}$, where $u(x_k) \in \mathcal{U}(x_k)$ for all k. A policy is stationary if the choice of decision only depends on the state and is independent of the time. Let

$$V_{\boldsymbol{\mu}}(\theta) = \lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{k=0}^{T-1} g(x_k, u(x_k)) | x_0 = \theta\right\}$$

be the expected revenue in state θ under policy $\boldsymbol{\mu}$. Our objective is to find the best policy $\boldsymbol{\mu^*}$ to optimize the average revenue per stage starting from an initial state θ.

As shown in our prior preliminary results [1], the average revenue per stage under any stationary policy is independent of the initial state, and the *average revenue maximization problem* could be transformed into the *stochastic shortest path problem*. More specifically, we pick a state n as the start state of the stochastic shortest path problem, and define an *artificial* termination state t from the state n. The transition probability from an arbitrary state θ to the termination state t satisfies $P_{\theta t}(\boldsymbol{\mu}) = P_{\theta n}(\boldsymbol{\mu})$, as show in Fig. 4.

In the stochastic shortest path problem, we define $-\hat{g}(n, \boldsymbol{\mu})$ as the expected stage *cost* incurred at state n under policy $\boldsymbol{\mu}$. Let A^* be the optimal average revenue per stage starting from the state n to the terminal state t, and let $A^* - \hat{g}(n, \boldsymbol{\mu})$ be the normalized expected stage cost. Then the normalized expected terminal cost from the state $x_0 = n$ under the policy $\boldsymbol{\mu}$, $h^{\boldsymbol{\mu}}(n) = \lim_{N\to\infty} E\left\{\sum_{k=0}^{N-1}\{A^* - g(x_k, u(x_k))\}\right\}$, is zero when the policy $\boldsymbol{\mu}$ is optimal. The cost minimization in the stochastic shortest path problem is equivalent to the original

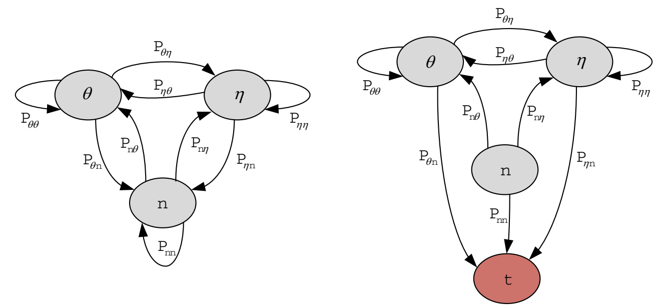

Fig. 4. Transition probability of the shortest path problem.

average revenue per stage maximization problem. Let $h^*(\theta)$ denote the optimal cost of the stochastic shortest path starting at state $\theta \in \mathcal{S}$, then we get the corresponding Bellman's equation as follows [22]:

$$h^{\boldsymbol{\mu}}(\theta) = \min_{\boldsymbol{\mu}} \left\{A^* - \hat{g}(\theta, \boldsymbol{\mu}) + \sum_{\eta\in\mathcal{S}} p_{\theta\eta}(\boldsymbol{\mu}) h^{\boldsymbol{\mu}}(\eta)\right\}, \ \theta \in \mathcal{S}. \tag{6}$$

If $\boldsymbol{\mu^*}$ is a stationary policy that maximizes the cycle revenue, we have the following equations:

$$h^*(\theta) = A^* - \hat{g}(\theta, \boldsymbol{\mu}^*) + \sum_{\eta\in\mathcal{S}} p_{\theta\eta}(\boldsymbol{\mu}^*) h^*(\eta), \ \theta \in \mathcal{S}. \tag{7}$$

The Bellman's equation is an iterative way to solve MDP problems. Next we show that solving the Bellman's equation (8) in the stochastic shortest path problem leads to the optimal solution.

Proposition 1: For the stochastic shortest path problem, given any initial values of terminal costs $h_0(\theta)$ for all states $\theta \in \mathcal{S}$, the sequence $\{h_l(\theta), l = 1, 2, \ldots\}$ generated by the iteration

$$h_{l+1}(\theta) = \min_{\boldsymbol{\mu}} \left\{A^* - \hat{g}(\theta, \boldsymbol{\mu}) + \sum_{\eta\in\mathcal{S}} P_{\theta\eta}(\boldsymbol{\mu}) h_l(\eta)\right\}, \theta \in \mathcal{S}, \tag{8}$$

converges to the optimal terminal cost $h^*(\theta)$ for each state θ.

Proof: For an arbitrary state θ and an admissible policy $\boldsymbol{\mu}$, there exists an integer γ satisfying $P\{x_\gamma \neq t | x_0 = \theta, \boldsymbol{\mu}\} < 1$ [24]. Let $\rho = \max_{(\theta,\boldsymbol{\mu})} P\{x_\gamma \neq t | x_0 = \theta, \boldsymbol{\mu}\}$, then $\rho < 1$ and $P\{x_{2\gamma} \neq t | x_0 = \theta, \boldsymbol{\mu}\} = P\{x_{2\gamma} \neq t | x_\gamma \neq t, x_0 = \theta, \boldsymbol{\mu}\} \cdot P\{x_\gamma \neq t | x_0 = \theta, \boldsymbol{\mu}\} \leq \rho^2$. Therefore, we get $P\{x_{\phi\gamma} \neq t | x_0 = \theta, \boldsymbol{\mu}\} \leq \rho^\phi$.

We break down the cost $h^{\boldsymbol{\mu}}(x_0)$ into the portions incurred over the first $K\gamma$ time slots (K is a positive integer) and over the remaining time slots, *i.e.,*

$$\begin{aligned} h^{\boldsymbol{\mu}}(x_0) &= \lim_{N\to\infty} E\left\{\sum_{k=0}^{N-1}\{A^* - g(x_k, u(x_k))\}\right\} \\ &= E\left\{\sum_{k=0}^{K\gamma-1}\{A^* - g(x_k, u(x_k))\}\right\} \\ &\quad + \lim_{N\to\infty} E\left\{\sum_{k=K\gamma}^{N-1}\{A^* - g(x_k, u(x_k))\}\right\}. \end{aligned} \tag{9}$$

Define $\Gamma = \gamma \max_{(\theta,\boldsymbol{\mu})} |A^* - \hat{g}(\theta, \boldsymbol{\mu})|$, which denotes the upper bound on the cost of an γ-slot cycle when termination does not occur during the cycle. Then, the expected cost during the K-th γ-slot cycle (time slots $K\gamma$ to $(K+1)\gamma-1$) is upper bounded by $\rho^K\Gamma$, so that

$$\begin{aligned} & E\left\{\left|h^{\boldsymbol{\mu}}(x_0)-\sum_{k=0}^{K\gamma-1}\{A^*-g(x_k,u(x_k))\}\right|\right\} \\ &=\left|\lim_{N\to\infty}E\left\{\sum_{k=K\gamma}^{N-1}\{A^*-g(x_k,u(x_k))\}\right\}\right| \\ &\le \Gamma\sum_{\phi=K}^{\infty}\rho^{\phi}=\frac{\rho^K\Gamma}{1-\rho}. \end{aligned} \tag{10}$$

Let $h_0(x_0)$ be a terminal cost function as defined in the proposition, and then its expected value under $\boldsymbol{\mu}$ after $K\gamma$ time slots is bounded by

$$\begin{aligned} |E\{h_0(x_{K\gamma})\}| &= \left|\sum_{\theta\in\mathcal{S}}P(x_{K\gamma}=\theta|x_0,\boldsymbol{\mu})h_0(\theta)\right| \\ &\le \left(\sum_{\theta\in\mathcal{S}}P(x_{K\gamma}=\theta|x_0,\boldsymbol{\mu})\right)\max_{\theta\in\mathcal{S}}|h_0(\theta)|. \end{aligned} \tag{11}$$

Since the probability that $x_{K\gamma}\neq t$ is less than or equal to ρ^K for any policy, we have $|E\{h_0(x_{K\gamma})\}|\le\rho^K\max_{\theta\in\mathcal{S}}|h_0(\theta)|$. Therefore, we can get

$$\begin{aligned} & -\rho^K\max_{\theta\in\mathcal{S}}|h_0(\theta)|+h^{\boldsymbol{\mu}}(x_0)-\frac{\rho^K\Gamma}{1-\rho} \\ &\le E\left\{h_0(x_{K\gamma})+\sum_{k=0}^{K\gamma-1}\{A^*-g(x_k,u(x_k))\}\right\} \\ &\le \rho^K\max_{\theta\in\mathcal{S}}|h_0(\theta)|+h^{\boldsymbol{\mu}}(x_0)+\frac{\rho^K\Gamma}{1-\rho}. \end{aligned} \tag{12}$$

The expected value in the middle term of the above inequalities is the $K\gamma$-slot cost of policy $\boldsymbol{\mu}$ starting from state x_0 with a terminal cost $h_0(x_{K\gamma})$. The minimum of this cost over all $\boldsymbol{\mu}$ is equal to the value $h_{K\gamma}(x_0)$, which is generated by the dynamic programming recursion (8) after $K\gamma$ iterations. Thus, by taking the minimum over $\boldsymbol{\mu}$ in (12), we obtain for all x_0 and K,

$$\begin{aligned} -\rho^K\max_{\theta\in\mathcal{S}}|h_0(\theta)|+h^*(x_0)-\frac{\rho^K\Gamma}{1-\rho}\le h_{K\gamma}(x_0) \\ \le\rho^K\max_{\theta\in\mathcal{S}}|h_0(\theta)|+h^*(x_0)+\frac{\rho^K\Gamma}{1-\rho}. \end{aligned} \tag{13}$$

And by taking the limit when $K\to\infty$, the terms involving ρ^K will go to zero, and we obtain $\lim_{K\to\infty}h_{K\gamma}(x_0)=h^*(x_0)$ for all x_0. In addition, since $|h_{K\gamma+q}(x_0)-h_{K\gamma}(x_0)|\le\rho^K\Gamma,\ q=1,2,\cdots,\gamma-1$, we have $\lim_{K\to\infty}h_{K\gamma+q}(x_0)=\lim_{K\to\infty}h_{K\gamma}(x_0)=h^*(x_0)$ for all $q=1,\cdots,\gamma-1$. ■

Proposition 1 shows that solving the Bellman's equation leads to the optimal average revenue A^* and the optimal differential cost h^*. The Bellman's equation can often be solved using value iteration or policy iteration algorithms; details can be found in [24] and [25]. Once having A^* and h^*, we can compute the optimal control decision $u^*(\theta)$ that minimizes the immediate differential cost of the current stage plus the remaining expected differential cost for state θ, *i.e.,*

$$u^*(\theta)=\arg\min_{\boldsymbol{\mu}}\left\{A^*-\hat{g}(\theta,\boldsymbol{\mu})+\sum_{\eta\in\mathcal{S}}p_{\theta\eta}(\boldsymbol{\mu})h^*(\eta)\right\}. \tag{14}$$

V. Suboptimal Control and Dynamic Programming

Solving the Bellman's equation does not lead to a closed-form optimal control policy, and the iterative computation is time-consuming for our problem with a large state space. To resolve this issue, a broad class of suboptimal control methods referred as *approximate dynamic programming* (ADP) have been proposed in [22]. Next we first propose a simple heuristic control policy in Section V-A. Then in Section V-B, we will improve the performance of the heuristic algorithm by using the idea of *rollout algorithm* (which is a class of ADP algorithms). It is known that the suboptimal policy based on the rollout algorithm is identical to the policy obtained by a *single policy improvement step* of the classical policy iteration method [24], [25].

A. Heuristic Control Policy

Several observations can help us with the suboptimal algorithm design. First, the channel state transitions are determined by the underlying primary traffic and are not affected by any control policy. Second, all secondary users experience the same channel availability independent of their locations, and all channels are homogenous and provide the same data rates. This means that we are interested in *how many* users to admit rather than *who* to admit, and we only care *how many* channels are available instead of *which* are available. This motivates us to first consider admission control and channel allocation separately.

For the admission control, we first consider a simple *threshold-based* strategy, where a new user will be admitted if and only if the total number of admitted users is smaller than the threshold. Given a fixed admission control threshold T_h, there are many ways of performing the channel allocation. To resolve this issue, we propose the *largest-delay-first* strategy, which allocates available channels to admitted users with the largest accumulated delay first.

Proposition 2: The largest-delay-first channel allocation policy is optimal under any fixed threshold-based admission control policy.

Proof: Under a threshold-based admission control policy, the number of admitted users in the system is constant in any time slot. The objective function in (4) is equal to the maximization problem $\max E\{g(x,u(x))\}$ due to ergodicity of the instant revenue $g(x_k,u(x_k))$. Let $\Omega_c=E\left\{\sum_{i=0}^{D_{max}}\omega_{c,i}\right\}$ be the expected number of normally completed users at the end of each time slot, $\Omega_e=E\left\{\sum_{i=0}^{D_{max}}\omega_{e,i}\right\}$ be the expected number of users in the network at the end of each time slot, and $\Omega_q=E\{\omega_q\}$ be the expected number of forcefully terminated users at the end of each time slot. Then

$$\max E\{g(x,u(x))\}=\max\{R_c\Omega_c+R_t\Omega_e-C_q\Omega_q\}. \tag{15}$$

Under the threshold-based admission control policy, $\Omega_e = \sum_{i=0}^{D_{max}} \omega_{e,i}$ in all time slot k and equals to the threshold.

In the largest-delay-first policy, let L_c be the expected length of a normally completed session, D_c the expected delay of a normally completed session, and L_q the expected length of a forcefully terminated session. Now let us consider an arbitrary channel allocation policy as the benchmark, and we use the superscript (g) to denote all parameters corresponding to this particular channel allocation policy, *i.e.,* $\Omega_c^{(g)}$, $\Omega_e^{(g)}$, $\Omega_q^{(g)}$, $L_c^{(g)}$, $D_c^{(g)}$, and $L_q^{(g)}$. We will show that the largest-delay-first policy is no worse than this benchmark policy, which will prove the proposition.

Because all actively served users have the same completion probability P_f independent of the channel allocation decisions, we can show that $\Omega_c = \Omega_c^{(g)}$, $\Omega_e = \Omega_e^{(g)}$, and $L_c = L_c^{(g)}$. Since the largest-delay-first policy always allocates available channels to the secondary users with the largest delay, we have $D_c \geq D_c^{(g)}$.

Here comes the critical proof step. We consider Ω_e virtual channels, one for each user in the network. If the secondary user is allocated an available *physical* channel, then its virtual channel is "idle" in that time slot; otherwise its virtual channel is "busy" and causes a delay. In the long run (when $T \to \infty$), we have the following:

$$\begin{aligned}\Omega_e \cdot T &= \Omega_c T(L_c + D_c) + \Omega_q T(L_q + D_{max}) \\ &= \Omega_c^{(g)} T(L_c^{(g)} + D_c^{(g)}) + \Omega_q^{(g)} T(L_q^{(g)} + D_{max}).\end{aligned} \tag{16}$$

Based on the relationships we just derived in the previous paragraph, we have

$$\Omega_q(L_q + D_{max}) \leq \Omega_q^{(g)}(L_q^{(g)} + D_{max}). \tag{17}$$

Since the number of available channels is the same under the two channel allocation policies in any time slot , we have

$$\Omega_c T L_c + \Omega_q T L_q = \Omega_c^{(g)} T L_c^{(g)} + \Omega_q^{(g)} T L_q^{(g)}. \tag{18}$$

Since $\Omega_c = \Omega_c^{(g)}$ and $L_c = L_c^{(g)}$, (18) implies that $\Omega_q L_q = \Omega_q^{(g)} L_q^{(g)}$. Together with inequality (17), we have $\Omega_q \leq \Omega_q^{(g)}$.

Because $\Omega_c = \Omega_c^{(g)}$, $\Omega_e = \Omega_e^{(g)}$, and $\Omega_q \leq \Omega_q^{(g)}$, we have $R_c\Omega_c + R_t\Omega_e - C_q\Omega_q \geq R_c\Omega_c^{(g)} + R_t\Omega_e^{(g)} - C_q\Omega_q^{(g)}$, *i.e.,* $\max E\left\{g(x_k, u(x_k))\right\} \geq \max E\left\{g^{(g)}(x_k, u(x_k))\right\}$. This shows that our proposed largest-delay-first channel allocation policy is no worse than any channel allocation algorithm, and thus is optimal with a threshold-based admission control. ■

For performance comparison, we further define two benchmark channel allocation strategies.

- **Strategy 1:** allocate the available channels to the admitted users with the smallest accumulated delays. If there is a tie, break it randomly.
- **Strategy 2:** allocate the available channels to the admitted users randomly.

In Fig. 5 and Fig. 6, we compare the proposed channel allocation policy and the two benchmark policies with different total number of channels. All three policies follow the same threshold-based admission control policies. From these figures, we observe that the proposed largest-delay-first policy is no worse than the other two under all choices of parameters.

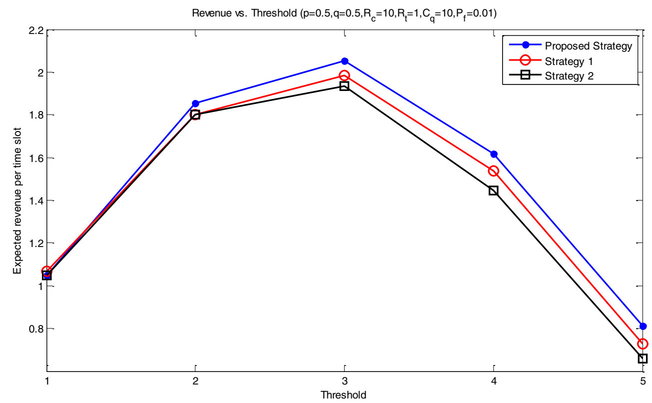

Fig. 5. Revenue versus threshold of three different strategies ($J = 5, D_{\max} = 5$).

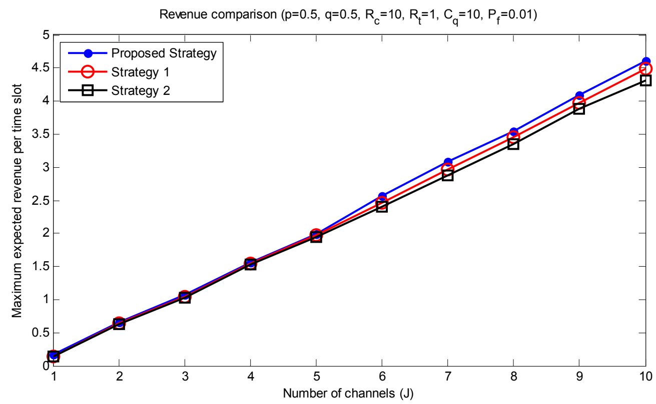

Fig. 6. Maximum expected revenue comparison with different channels.

B. Rollout Control Policy

The heuristic algorithm proposed in Section V-A can be further improved by the rollout algorithm. The general background of the rollout algorithm is in [24]–[26]. In this subsection, based on the analysis of the heuristic control policy, we propose a simplified rollout algorithm (*rollout control policy*) to further improve the performance.

Consider two different user states $\boldsymbol{\omega}_e^{(1)}$ and $\boldsymbol{\omega}_e^{(2)}$ that have the same number of secondary users. If it is possible for transit from state $\boldsymbol{\omega}_e^{(1)}$ to state $\boldsymbol{\omega}_e^{(2)}$ under a particular channel condition without admitting any new user, then obviously the total time delay of $\boldsymbol{\omega}_e^{(1)}$ summed over all users must be less than that of $\boldsymbol{\omega}_e^{(2)}$ (as each user either has the same delay or a larger delay during the transition). We give the following definitions:

Definition 1 (User State Comparison): Consider two different user states $\boldsymbol{\omega}_e^{(1)}$ and $\boldsymbol{\omega}_e^{(2)}$ that have the same number of secondary users. If it is possible to transit from state $\boldsymbol{\omega}_e^{(1)}$ to state $\boldsymbol{\omega}_e^{(2)}$ under a particular channel condition without admitting any new user, then $\boldsymbol{\omega}_e^{(1)}$ is *better* than $\boldsymbol{\omega}_e^{(2)}$, denoted $\boldsymbol{\omega}_e^{(1)} \gtrdot \boldsymbol{\omega}_e^{(2)}$.

Definition 2 (Quality of Channel State): Consider a user state $\boldsymbol{\omega}_e^{(1)}$ and a channel state m. The channel state m is $\boldsymbol{B}$ (Bad) for the user state $\boldsymbol{\omega}_e^{(1)}$ if and only if m is less than the total number of users in $\boldsymbol{\omega}_e^{(1)}$. Otherwise, the channel state m is $\boldsymbol{G}$ (Good) for the user state $\boldsymbol{\omega}_e^{(1)}$.

Now consider a heuristic control policy with the admission control threshold N_{th} and largest-delay-first channel allocation mechanism. Under this policy, we can divide the infinite-horizon process into infinite number of segments separated by the time slots in which there is at least one user leaving the system (normal completion or forced termination). Then we can define a new average revenue $\bar{g}(N_{th}, \theta)$ and its expectation $\bar{G}(N_{th}, \theta))$ over each segment. Due to the threshold-based admission control, we will only admit new users in the first slot of a segment.

Definition A (Average Revenue and Expected Average Revenue) If the network state is θ at the beginning of time slot k, and at least one user leaves the system for the first time (normal completion or forced termination) in time slot $k+\delta$, we define the average revenue over the period $[k, k+\delta]$ as

$$\bar{g}(N_{th}, \theta) = \frac{n_c(N_{th}, \theta)}{\delta+1} R_c - \frac{n_d(N_{th}, \theta)}{\delta+1} C_q + N_{th} R_t, \quad (19)$$

where $n_c(N_{th}, \theta)$ is number of users completing connections normally in time slot $k+\delta$, and $n_d(N_{th}, \theta)$ is number of users being forced to terminate in time slot $k+\delta$. The expected average revenue is denoted as

$$\begin{aligned} \bar{G}(N_{th}, \theta) &= E\{\bar{g}(N_{th}, \theta)\} \\ &= N_c(N_{th}, \theta) R_c - N_d(N_{th}, \theta) C_q + N_{th} R_t, \end{aligned} \quad (20)$$

where $N_c(N_{th}, \theta) = E\left\{\frac{n_c(N_{th},\theta)}{\delta+1}\right\}$ and $N_d(N_{th}, \theta) = E\left\{\frac{n_d(N_{th},\theta)}{\delta+1}\right\}$.

The *expected revenue* in Definition V-B is different from the *instant revenue* in (5). The expected revenue is defined under a very special case, where no new users are admitted except in the first time slot and no users leave the network except in the last time slot of the interval. The instant revenue defined in (5) is the revenue for a generic time slot. Furthermore, $\bar{G}(N_{th}, \theta)$ represents the expected average revenue per time slot when maintaining a fixed number of users until someone leaves. Although the precise value of $\bar{G}(N_{th}, \theta)$ is hard to compute explicitly, we have the following result as a corollary of Proposition 2.

Proposition 3: Given any fixed N_{th} and θ, the largest-delay-first channel allocation policy achieves the maximum $\bar{G}(N_{th}, \theta)$.

Based on Proposition 3, we will still use the largest-delay-first strategy channel allocation. The key remaining issue is how to improve the admission control policy. Next we characterize the properties of the largest-delay-first channel allocation policy (the expected average revenue $\bar{G}(N_{th}, \theta)$ in the heuristic control policy) in several lemmas, which enable us to design a better heuristic algorithm for the admission control part.

According to the lemmas given in Appendix A, we can characterize $\bar{G}(N_{th}, \theta)$ as follows.

Proposition 4: $\bar{G}(N_{th}, \theta)$ is a concave function of N_{th}.

Proof: The second order derivative of $\bar{G}(N_{th}, \theta)$ in terms of N_{th} is

$$\bar{G}''(N_{th}, \theta) = N_c''(N_{th}, \theta) R_c - N_d''(N_{th}, \theta) C_q, \quad (21)$$

where $N_c''(N_{th}, \theta) < 0$ and $N_d''(N_{th}, \theta) > 0$ based on Lemma 1 and Lemma 2 in Appendix A. Thus we have $\bar{G}''(N_{th}, \theta) < 0$, *i.e.,* $\bar{G}(N_{th}, \theta)$ is a concave function of N_{th}. ■

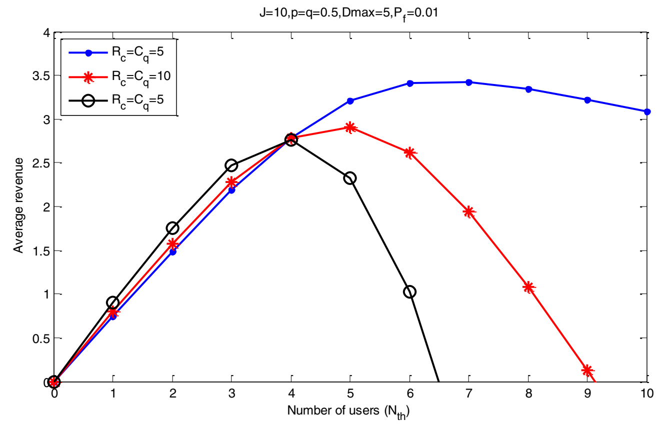

Fig. 7. The values of $\bar{G}(N_{th}, \theta)$ versus N_{th} corresponding to different values of R_c and C_q when $J = 10$, $p = 0.5$, $q = 0.5$, $D_{max} = 5$, $P_f = 0.01$, $R_t = 0.7$ and $\theta = \{m, [0, 0, 0, 0, 0, 0]\}$.

Figure 7 plots $\bar{G}(N_{th}, \theta)$ versus N_{th} with fixed $\theta = \{m, [0, 0, 0, 0, 0, 0]\}$ and different values of R_c and C_q.

Now we are ready to discuss the heuristic admission control policy. Given a state $\theta = (m, \boldsymbol{\omega_e})$, the admission control decision can be either *maintaining* or *searching*, depending on the relationship between the channel state component m and user state component $\boldsymbol{\omega_e}$. More precisely, if m is $\boldsymbol{B}$ (Bad) for $\boldsymbol{\omega_e}$, the network coordinator will maintain the current user population and do not admit any new user (*i.e.,* maintaining). This is because the network resource is not enough to support the current users, and admitting new users will make the situation worse. If m is $\boldsymbol{G}$ (Good) for $\boldsymbol{\omega_e}$, the network coordinator first searches for the value of N_{th}^* that achieves $\max_{N_{th}} \bar{G}(N_{th}, \theta)$ (*i.e.,* searching), and then admits the number of users equal to the difference between N_{th}^* and the current users in the network. Proposition 4 shows that $\bar{G}(N_{th}, \theta)$ has a unique maximizer N_{th}^* (with a fixed state θ), and implies a simple stopping rule for the numerical search. If we have $\bar{G}(N_{th}' - 1, \theta) \le \bar{G}(N_{th}', \theta)$ and $\bar{G}(N_{th}', \theta) \ge \bar{G}(N_{th}' + 1, \theta)$, then $N_{th}^* = N_{th}'$.

The heuristic admission control introduced above is a rollout control policy based on the theory in [24]–[26]. More specifically, the value of $\max_{N_{th}} \bar{G}(N_{th}, \theta)$ computed in the searching step is the *cost-to-go* starting from a state θ. As Proposition 4 shows that this is a concave maximization problem, we can use several well-known numerical methods to achieve this. One possibility is the gradient decent method, which has a linear convergence rate as shown in [27]. More precisely, the maximum number of convergence of the gradient decent method is proportional to $\log(\bar{G}(N_{th}^{initial}, \theta) - \bar{G}(N_{th}^{optimal}, \theta))/\epsilon$, where ϵ is the stopping criterion. Since the precise value of $\bar{G}(N_{th}, \theta)$ is hard to compute with a low complexity, we will use an approximation $\tilde{G}(N_{th}, \theta)$ instead in the searching step. In this paper, we use an on-line computation (simulation) to get $\tilde{G}(N_{th}, \theta)$. Mover specifically, for each choice of (N_{th}, θ), we can obtain the value of $\bar{g}((N_{th}, \theta))$ as in (19) for each particular simulation, and take the average over many simulations to obtain an approximation $\tilde{G}(N_{th}, \theta)$. The memory requirements are proportional to the expected length of the segments separated by the time slots in which there is at least one user leaving the system (normal completion or forced termination)

C. Revenue Boundary

In this subsection, we will compare the performance of two heuristic policies that we have proposed. Before that, we will establish an upper-bound of the revenue achievable under any control policy (heuristic or optimal). We call the bound the *revenue boundary*.

We first prove the following property of the expected average revenue $\bar{G}(N_{th}, \theta)$.

Proposition 5: For a fixed number of users N_{th}, if there are two states $\theta_1 = \{m, \boldsymbol{\omega}_{\boldsymbol{e}}^{(\mathbf{1})}\}$ and $\theta_2 = \{m, \boldsymbol{\omega}_{\boldsymbol{e}}^{(\mathbf{2})}\}$ such that $\boldsymbol{\omega}_{\boldsymbol{e}}^{(\mathbf{1})} > \boldsymbol{\omega}_{\boldsymbol{e}}^{(\mathbf{2})}$, we have $\bar{G}(N_{th}, \theta_1) > \bar{G}(N_{th}, \theta_2)$.

Proof: According to Lemma 3 in Appendix A, we have $N_c(N_{th}, \theta_1) > N_c(N_{th}, \theta_2)$ and $N_d(N_{th}, \theta_1) < N_d(N_{th}, \theta_2)$. By substituting them into (20), we get $\bar{G}(N_{th}, \theta_1) > \bar{G}(N_{th}, \theta_2)$. ■

Then we can characterize the revenue boundary.

Proposition 6: Consider a network state $\bar{\theta} = \{m, [0, 0, 0, \cdots]\}$, where there are m available channels. The maximum expected revenue per time slot achieved by any policy, denoted by G_{max}, satisfies $G_{max} < \max_{m,N_{th}} \{\bar{G}(N_{th}, \bar{\theta})\}$, where N_{th} is an admission control threshold.

Proof: Assume $\hat{\theta} = \{m, \boldsymbol{\omega}_{\boldsymbol{e}}^{\boldsymbol{\theta}}\}$ and $\eta = \{m, \boldsymbol{\omega}_{\boldsymbol{e}}^{\boldsymbol{\eta}}\}$ are two network states with m available channels and N_{th} users, where $\boldsymbol{\omega}_{\boldsymbol{e}}^{\boldsymbol{\theta}} = [N_{th}, 0, 0, \cdots]$. If $\boldsymbol{\omega}_{\boldsymbol{e}}^{\boldsymbol{\eta}} \neq \boldsymbol{\omega}_{\boldsymbol{e}}^{\boldsymbol{\theta}}$, we have $\boldsymbol{\omega}_{\boldsymbol{e}}^{\boldsymbol{\theta}} > \boldsymbol{\omega}_{\boldsymbol{e}}^{\boldsymbol{\eta}}$. From Proposition 5, we get $\bar{G}(N_{th}, \hat{\theta}) > \bar{G}(N_{th}, \eta)$. In addition, after the control decision in the first time slot, N_{th} new secondary users are admitted in the case of $\overline{\theta}$ (since there are originally no users in the system), and no new secondary user is admitted in the case of $\hat{\theta}$ (since there are already N_{th} users with zero accumulative delay in the system). Thus, after the first time slot, we achieve the same state in both cases. In the following time slots, the expected changes of the two cases are thus the same. Therefore, according to the definition of $\overline{G}$ in Definition V-B, we have $\bar{G}(N_{th}, \bar{\theta}) = \bar{G}(N_{th}, \hat{\theta})$. Therefore, the cost-to-go we compute in the search step is never larger than $\max_{m,N_{th}} \{\bar{G}(N_{th}, \bar{\theta})\}$. As the optimal policy can be viewed as a special case of the rollout policy by using the optimal policy as the *base policy*, it follows that the expected revenue per time slot of any policy (including the optimal one) is less than $\max_{m,N_{th}} \{\bar{G}(N_{th}, \bar{\theta})\}$. ■

In Section II, we have assumed perfect spectrum sensing. Under this assumption, the control policy of the throughput maximization problem studied in [4], [5] can be simplified into admitting secondary users to make full use of the available channels in each time slot, which we call *greedy admission control* in this paper. Such greedy admission control policy will admit new users whenever possible such that the total active users in a time slot equals to the number of available channels. Comparing with our proposed policy, this greedy policy is more aggressive and does not consider channel availabilities in the future, and thus will lead to a larger number of forced dropped users. We have plotted the expected revenue of the greedy admission control in Fig. 8, with the comparison with our proposed admission control and the revenue boundary. We can see that even the performance of our proposed heuristic control policy is better than that of the greedy control policy. The heuristic control policy (with the threshold-based admission control) is simple but effective, while the rollout algorithm achieves a slightly better performance but with a much higher computational complexity. The actual performance gap between the proposed algorithms and the optimal policy could be even smaller, as the revenue boundary in Proposition 6 may not be very tight.

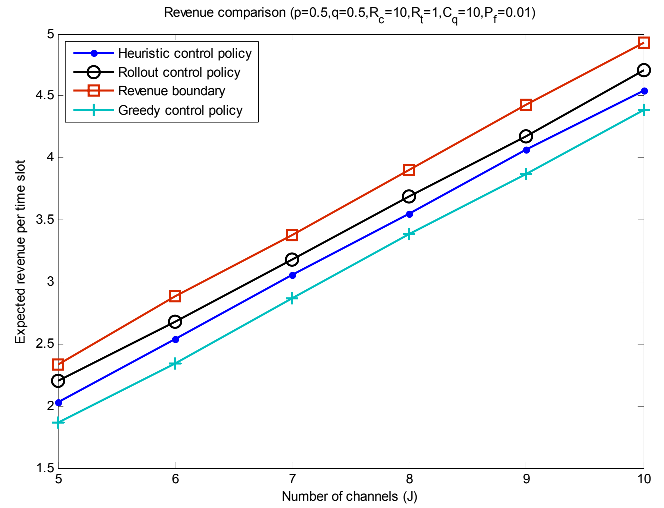

Fig. 8. Expected revenue comparison between the greedy control policy, the heuristic control policy, the rollout control policy, and the revenue boundary when $D_{max} = 5$, $p = 0.5$, $q = 0.5$, $R_c = 10$, $R_t = 1$, $C_q = 10$, $P_f = 0.01$ and $J \in [5, 10]$.

VI. Conclusions

Supporting QoS over cognitive radio networks is very challenging, mainly due to the uncertainty of available communication resources. As one further step towards understanding this under-explored yet practically important research area, we considered supporting delay sensitive traffic in cognitive radio networks. The key is to jointly optimize admission control and channel allocation, in order to balance the number of concurrent sessions and the QoS of each session. We formulated the problem as an infinite-horizon Markov decision process problem, and proved that the optimal average revenue is independent of the initial system state. Then we transformed the original problem into a stochastic shortest path problem, and proved that the Bellman's equation converged to the optimal policy. Furthermore, we proposed a heuristic control policy and proved that the largest-delay-first strategy is optimal given threshold-based admission control. We further proposed a rollout algorithm that improves upon the heuristic algorithm by doing dynamic admission control. By comparing with a revenue bound, we show that both of our proposed algorithms achieve close-to-optimal performance.

Appendix

A. Several Lemmas for Proving Propositions 4 and 5

After defining the expected revenue $\bar{g}(N_{th}, \theta)$ and the expected average revenue $\bar{G}(N_{th}, \theta)$ in Definition V-B, we give the following intermediate lemmas to help to illustrate the properties of $\bar{G}(N_{th}, \theta)$ in terms of the first and second order derivatives.

Lemma 1: For a fixed state θ, $N_c(N_{th}, \theta)$ is a non-decreasing and concave function of the number of users N_{th}.

Proof: Recall that all the users have the same completion probability P_f when they are actively served. Thus we have $N_c(N_{th}+1, \theta) \geq N_c(N_{th}, \theta)$, as having one more user means that it is possible to actively serve one more user and thus have one more normal session completion. Furthermore, we assume that under the same channel condition and over a period of time slots, the incremental number of served users per time slot is Δ_1 when the number of users changes from $N_{th}-1$ to N_{th}. Then Δ_2, the incremental number of served users per time slot from N_{th} to $N_{th}+1$, should be no bigger than Δ_1. This is because if $N_{th}+1$ users can be allocated available channels, N_{th} users could be allocated available channels in the same time slot. Therefore, we have $N_c(N_{th}+1, \theta) - N_c(N_{th}, \theta) \leq N_c(N_{th}, \theta) - N_c(N_{th}-1, \theta)$, which means $N_c(N_{th}, \theta)$ is a non-decreasing and concave function of N_{th}. ■

Lemma 2: For a fixed state θ, $N_d(N_{th}, \theta)$ is a non-decreasing and convex function of the number of users N_{th}.

Proof: Having one more admitted user means that a higher probability of a forced termination, *i.e.,* $N_d(N_{th}+1, \theta) \geq N_d(N_{th}, \theta)$. Under the largest-delay-first channel allocation policy, define $\Delta_1 = N_d(N_{th}, \theta) - N_d(N_{th}-1, \theta)$ and the additional user as U_{ser}, and $\Delta_2 = N_d(N_{th}+1, \theta) - N_d(N_{th}, \theta)$. For discussion convenience, we call the system with $N_{th}-1$ users as *Case 1*, the system with N_{th} users as *Case 2*, and the system with $N_{th}+1$ users as *Case 3*. In Case 2, we divide users into two parts: U_{ser} and other $N_{th}-1$ users. In Case 3, we also divide users into two parts: U_{ser} and other N_{th} users. Then we define $N_d(N_{th}+1, \theta) = N'_d(N_{th}, \theta) + N_d^3(U_{ser}, \theta)$ and $N_d(N_{th}, \theta) = N'_d(N_{th}-1, \theta) + N_d^2(U_{ser}, \theta)$. Here $N_d^3(U_{ser}, \theta)$ and $N'_d(N_{th}, \theta)$ represent the corresponding parts of $N_d(N_{th}+1, \theta)$ caused by the forced termination of U_{ser} and other users in Case 3, respectively; $N_d^2(U_{ser}, \theta)$ and $N'_d(N_{th}-1, \theta)$ represent the corresponding parts of $N_d(N_{th}, \theta)$ caused by the forced termination of U_{ser} and other users in Case 2, respectively. On this basis, we further define $\Delta_2 = \Delta'_2 + \Delta''_2$, where $\Delta'_2 = N'_d(N_{th}, \theta) - N'_d(N_{th}-1, \theta)$ and $\Delta''_2 = N_d^3(U_{ser}, \theta) - N_d^2(U_{ser}, \theta)$.

In Case 2 and Case 3, we now exclude the user U_{ser} from the system and assume the channels allocated to U_{ser} are occupied by primary users. Then we can have the above expression of Δ'_2 to illustrate the effect of the increased user U_{ser} from $N_{th}-1$ to N_{th}. Comparing $\Delta'_2 = N'_d(N_{th}, \theta) - N'_d(N_{th}-1, \theta)$ with $\Delta_1 = N_d(N_{th}, \theta) - N_d(N_{th}-1)$, the difference is that in any time slot (on any sample path), the channel state of Δ'_2 is always no better than that of the Δ_1 case (as the extra user U_{ser} may occupy an available channel). Therefore, in terms of the expected number of users forced to leave the system per time slot, the effect of the increased user to Δ'_2 is larger than that to Δ_1. This leads to $\Delta'_2 \geq \Delta_1$. Moreover, considering U_{ser} from Case 2 to Case 3, we have $\Delta''_2 \geq 0$ under the largest-delay-first policy. From the above analysis, we get $\Delta_2 \geq \Delta_1$, *i.e.,* $N_d(N_{th}+1, \theta) - N_d(N_{th}, \theta) \geq N_d(N_{th}, \theta) - N_d(N_{th}-1, \theta)$, which means $N_d(N_{th}, \theta)$ is a non-decreasing and convex function of N_{th} [28]. ■

Lemma 3: For a fixed number of users N_{th}, if there are two states $\theta_1 = \{m, \boldsymbol{\omega}_e^{(1)}\}$ and $\theta_2 = \{m, \boldsymbol{\omega}_e^{(2)}\}$ such that $\boldsymbol{\omega}_e^{(1)} \gg \boldsymbol{\omega}_e^{(2)}$, we have $N_c(N_{th}, \theta_1) > N_c(N_{th}, \theta_2)$ and $N_d(N_{th}, \theta_1) < N_d(N_{th}, \theta_2)$.

Proof: The lemma directly follows the definitions of $\boldsymbol{\omega}_e^{(1)} \gg \boldsymbol{\omega}_e^{(2)}$ in Definition 1 and $N_c(N_{th}, \theta)$, $N_d(N_{th}, \theta)$ in Definition V-B. If $\boldsymbol{\omega}_e^{(1)} \gg \boldsymbol{\omega}_e^{(2)}$, the user state $\boldsymbol{\omega}_e^{(1)}$ can reach the user state $\boldsymbol{\omega}_e^{(2)}$ under a proper channel condition and a control policy. Consider two systems with the initial states θ_1 and θ_2, respectively, and follow the same channel conditions over time and the same control policy. When a user is forced to leave the system (completes the connection, respectively) with θ_1, in the system with θ_2, there must be a user that is forced to leave (completes the connection or is forced to leave, respectively) in the current or an earlier time slot. Therefore, we get $N_c(N_{th}, \theta_1) > N_c(N_{th}, \theta_2)$ and $N_d(N_{th}, \theta_1) < N_d(N_{th}, \theta_2)$ based on the definitions of $N_c(N_{th}, \theta)$ and $N_d(N_{th}, \theta)$. ■

REFERENCES

[1] F. Wang, J. Zhu, J. Huang, and Y. Zhao, "Admission control and channel allocation for supporting real-time applications in cognitive radio networks," in *Proc. 2010 IEEE GLOBECOM*, pp. 1–6.

[2] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: structure, optimality, and performance," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5431–5440, Dec. 2008.

[3] X. Liu, B. Krishnamachari, and H. Liu, "Channel selection in multi-channel opportunistic spectrum access networks with perfect sensing," in *Proc. 2010 IEEE DySPAN*, pp. 1–8.

[4] X. Zhou, G. Li, D. Li, D. Wang, and A. Soong, "Probabilistic resource allocation for opportunistic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2870–2879, Sep. 2010.

[5] R. Urgaonkar and M. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 766–777, June 2009.

[6] T.-C. Chau, K. Wong, and B. Li, "Optimal call admission control with QoS guarantee in a voice/data integrated cellular network," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1133–1141, May 2006.

[7] J. Hou, J. Yang, and S. Papavassiliou, "Integration of pricing with call admission control to meet QoS requirements in cellular networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 13, no. 9, pp. 898–910, Sep. 2002.

[8] D. K. Kim, D. Griffith, and N. Golmie, "A new call admission control scheme for heterogeneous wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3000–3005, Oct. 2010.

[9] X. Kang, Y.-C. Liang, and H. K. Garg, "Fading cognitive multiple access channels: outage capacity regions and optimal power allocation," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2382–2391, July 2010.

[10] J. Xiang, Y. Zhang, T. Skeie, and J. He, "QoS aware admission and power control for cognitive radio cellular networks," *Wireless Commun. and Mobile Comput.*, vol. 9, no. 11, pp. 1520–1531, Nov. 2009.

[11] L. B. Le and E. Hossain, "Resource allocation for spectrum underlay in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5306–5315, Dec. 2008.

[12] H. Kim and K. Shin, "Optimal admission and eviction control of secondary users at cognitive radio hotspots," in *Proc. 2009 IEEE Sensor, Mesh and Ad Hoc Communications and Networks*, pp. 1–9.

[13] H. Mutlu, M. Alanyali, and D. Starobinski, "Spot pricing of secondary spectrum access in wireless cellular networks," *IEEE/ACM Trans. Networking*, vol. 17, no. 6, pp. 1794–1804, Dec. 2009.

[14] J. Chapin and W. Lehr, "Cognitive radios for dynamic spectrum access—the path to market success for dynamic spectrum access technology," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 96–103, May 2007.

[15] J. Peha, "Sharing spectrum through spectrum policy reform and cognitive radio," *Proc. IEEE*, vol. 97, no. 4, pp. 708–719, Apr. 2009.

[16] L. Duan, J. Huang, and B. Shou, "Competition with dynamic spectrum leasing," in *Proc. 2010 IEEE DySPAN*, pp. 1–11.

[17] ——, "Investment and pricing with spectrum uncertainty: a cognitive operators perspective," *IEEE Trans. Mobile Computing*, 2011.

[18] Available: http://www.fcc.gov/

[19] M. Weiss, S. Delaere, and W. Lehr, "Sensing as a service: an exploration into practical implementations of DSA," in *Proc. 2010 IEEE DySPAN*, pp. 1–8.
[20] "Scenario descriptions and system requirements," European Union, Project number ICT-2007-216076, 2008.
[21] S. Huang, X. Liu, and Z. Ding, "Optimal sensing-transmission structure for dynamic spectrum access," in *Proc. 2009 IEEE INFOCOM*, pp. 2295–2303.
[22] D. Bertsekas, "Dynamic programming and suboptimal control: a survey from ADP to MPC," *European J. Control*, vol. 44, no. 4-5, pp. 310–334, 2005.
[23] S. Ross, *Introduction to Probability Models*, 9th edition. Academic Press, 2007.
[24] D. Bertsekas, *Dynamic Programming and Optimal Control, Vol. 1*, 3rd edition. Athena Scientific, 2005.
[25] ——, *Dynamic Programming and Optimal Control, Vol. 11*, 3rd edition. Athena Scientific, 2007.
[26] F. Wang, J. Huang, and Y. Zhao, "Admission control and channel allocation for supporting real-time applications in cognitive radio networks," Tech. Rep., 2011. Available: http://arxiv.org/abs/1201.3059v1
[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[28] B. Fox, "Discrete optimization via marginal analysis," *Management Science*, vol. 13, no. 3, pp. 210–216, Nov. 1966.

Feng Wang received B.S. in Electronic Information Engineering from Shandong University (Jinan, Shandong, P.R.China) in 2005 and Ph.D. in Communication and Information System from Peking University (Beijing, P.R.China) in 2011. He visited the Chinese University of Hong Kong as a Research Assistant between July to December, 2009. He is currently an engineer in the Beijing Space Technology Development and Test Center, China Academy of Space Technology, Beijing, P.R.China. His current research interests include resource allocation, cognitive radio and wireless sensor networks.

Jianwei Huang (S'01-M'06-SM'11) is an Assistant Professor in the Department of Information Engineering at the Chinese University of Hong Kong. He received B.S. in Electrical Engineering from Southeast University (Nanjing, Jiangsu, China) in 2000, M.S. and Ph.D. in Electrical and Computer Engineering from Northwestern University (Evanston, IL, USA) in 2003 and 2005, respectively. He worked as a Postdoc Research Associate in the Department of Electrical Engineering at Princeton University during 2005-2007. He was a visiting scholar in the School of Computer and Communication Sciences at École Polytechnique Fédérale De Lausanne (EPFL) during the Summer Research Institute in June 2009, and a visiting scholar in the Department of Electrical Engineering and Computer Sciences at University of California-Berkeley in August 2010.

Dr. Huang currently leads the Network Communications and Economics Lab (ncel.ie.cuhk.edu.hk), with the main research focus on nonlinear optimization and game theoretical analysis of communication networks, especially on network economics, cognitive radio networks, and smart grid. He is the recipient of the IEEE Marconi Prize Paper Award in Wireless Communications in 2011, the International Conference on Wireless Internet Best Paper Award 2011, the IEEE GLOBECOM Best Paper Award in 2010, the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2009, Asia-Pacific Conference on Communications Best Paper Award in 2009, and Walter P. Murphy Fellowship at Northwestern University in 2001.

Dr. Huang has served as Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS - Cognitive Radio Series, Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, Guest Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on "Economics of Communication Networks and Systems," Lead Guest Editor of IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS special issue on "Game Theory in Communication Systems," Lead Guest Editor of *IEEE Communications Magazine* Feature Topic on "Communications Network Economics," and Guest Editor of several other journals including (Wiley) *Wireless Communications and Mobile Computing*, *Journal of Advances in Multimedia*, and *Journal of Communications*.

Dr. Huang has served as Vice Chair of IEEE MMTC (Multimedia Communications Technical Committee) (2010-2012), Director of IEEE MMTC E-letter (2010), the TPC Co-Chair of IEEE WiOpt (International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks) 2012, the Publicity Co-Chair of IEEE Communications Theory Workshop 2012, the TPC Co-Chair of IEEE ICCC Communication Theory and Security Symposium 2012, the Student Activities Co-Chair of IEEE WiOpt 2011, the TPC Co-Chair of IEEE GlOBECOM Wireless Communications Symposium 2010, the TPC Co-Chair of IWCMC (the International Wireless Communications and Mobile Computing) Mobile Computing Symposium 2010, and the TPC Co-Chair of GameNets (the International Conference on Game Theory for Networks) 2009. He is also TPC member of leading conferences such as INFOCOM, MobiHoc, ICC, GLBOECOM, DySPAN, WiOpt, NetEcon, and WCNC. He is a senior member of the IEEE.

Yuping Zhao received the B.S. and M.S. degrees in electrical engineering from Northern Jiaotong University, Beijing, P.R.China, in 1983 and 1986, respectively. She received the Ph.D. and Doctor of Science degrees in wireless communications from Helsinki University of Technology, Helsinki, Finland, in 1997 and 1999, respectively. She was a System Engineer for telecommunication companies in China and Japan. She worked as a research engineer at the Helsinki University of Technology, Helsinki, Finland, and at the Nokia Research Center in the field of radio resource management for wireless mobile communication networks. Currently, she is a professor in the State Key Laboratory of Advanced Optical Communication Systems & Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing, P.R.China. Her research interests include the areas of wireless communications and corresponding signal processing, especially for OFDM, UWB and MIMO systems, cooperative networks, cognitive radio, and wireless sensor networks.