Cognitive Medium Access: Exploration, Exploitation, and Competition

Lifeng Lai, *Member*, *IEEE*, Hesham El Gamal, *Fellow*, *IEEE*,
Hai Jiang, *Member*, *IEEE*, and H. Vincent Poor, *Fellow*, *IEEE*

Abstract—This paper considers the design of efficient strategies that allow cognitive users to choose frequency bands to sense and access among multiple bands with unknown parameters. First, the scenario in which a single cognitive user wishes to opportunistically exploit the availability of frequency bands is considered. By adopting tools from the classical bandit problem, optimal as well as low complexity asymptotically optimal solutions are developed. Next, the multiple cognitive user scenario is considered. The situation in which the availability probability of each channel is known is first considered. An optimal symmetric strategy that maximizes the total throughput of the cognitive users is developed. To avoid the possible selfish behavior of the cognitive users, a game-theoretic model is then developed. The performance of both models is characterized analytically. Then, the situation in which the availability probability of each channel is unknown a priori is considered. Low-complexity medium access protocols, which strike an optimal balance between exploration and exploitation in such competitive environments, are developed. The operating points of these low-complexity protocols are shown to converge to those of the scenario in which the availability probabilities are known. Finally, numerical results are provided to illustrate the impact of sensing errors and other practical considerations.

Index Terms—Bandit problem, cognitive radio, exploration, exploitation, medium access.

1 Introduction

Recently, the opportunistic spectrum access problem has been the focus of significant research activity [1], [2], [3], [4]. The underlying idea is to allow unlicensed users (i.e., cognitive users) to access the spectrum available when the licensed users (i.e., primary users) are not active. The presence of high-priority primary users and the requirement that the cognitive users should not interfere with them define a new medium access paradigm which we refer to as *cognitive medium access*. The overarching goal of our work is to design efficient and low complexity cognitive medium access protocols.

The spectral opportunities available to the cognitive users are expected to be time-varying on different time-scales. For example, on a small scale, multimedia data traffic of the primary users will tend to be bursty [5]. On a large scale, one would expect the activities of each user to vary throughout the day. Therefore, to avoid interfering with the primary network, the cognitive users must first probe to determine whether there are primary activities in each channel before transmission. Under the assumption that each cognitive user cannot access all of the available channels simultaneously, the main task of the medium access protocol is to distributively choose which channels each cognitive user should attempt to use in different time slots, in order to fully (or maximally) utilize the spectral opportunities. This decision process can be enhanced by taking into account any available statistical information about the primary traffic. For example, with a single cognitive user capable of accessing (sensing) only one channel at a time, the problem becomes trivial if the availability probability of each channel is known a priori. In this case, the optimal rule is for the cognitive user to access the channel with the highest probability of being free in all time slots. However, such time-varying traffic information is typically not available to the cognitive users a priori. The need to learn this information online creates a fundamental trade-off between exploitation and exploration. Exploitation refers to the short-term gain resulting from accessing the channel with the estimated highest probability of being free (based on the results of previous sensing decisions) whereas exploration is the process by which the cognitive user learns the statistical behavior of the primary traffic (by choosing possibly different channels to probe across time slots). In the presence of multiple cognitive users, the medium access algorithm must also account for the competition between different users over the same channel.

In this paper, we take a first step toward developing a framework for the design and analysis of cognitive medium access protocols in *uncertain* environments. As argued in the sequel, our approach allows for the construction of strategies that strike an optimal balance among exploration, exploitation, and competition under certain modeling assumptions. The key observation motivating our approach is the equivalence between our problem and the classical multiarmed bandit problem in the single-user case (see [6]

- *L. Lai is with the Department of Systems Engineering, University of Arkansas, Little Rock, 2801 S. University Ave., Little Rock, AR 72204. E-mail: lxlai@ualr.edu.*
- *H. El Gamal is with the Department of Electrical and Computer Engineering, Ohio State University, 205 Dreese Labs, 2015 Neil Ave., Columbus, OH 43210. E-mail: helgamal@ece.osu.edu.*
- *H. Jiang is with the Department of Electrical and Computer Engineering, University of Alberta, 9107 116 Street, Edmonton, Alberta, Canada T6G 2V4. E-mail: hai.jiang@ece.ualberta.ca.*
- *H.V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544. E-mail: poor@princeton.edu.*

Manuscript received 10 Sept. 2009; revised 8 Feb. 2010; accepted 28 Feb. 2010; published online 25 Aug. 2010.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2009-09-0372.
Digital Object Identifier no. 10.1109/TMC.2010.65.

and references therein). This equivalence allows for building a solid foundation for cognitive medium access using tools from reinforcement machine learning [7]. The connection between cognitive medium access and the multiarmed bandit problem has been independently observed in [8]. That work, however, is limited to special cases of the general approach presented here. In particular, in [8], the channels are assumed to be independent and the goal is to maximize the discounted sum of throughput. Related work also appears in [9] and [10], in which the availability of each channel is assumed to follow a Markov chain, whose transition matrix is known to the cognitive user. The only uncertainty faced by the cognitive user in that work is the particular realization of the channel, while in our work, the cognitive users also need to learn the statistics of the channel in real time. We will briefly discuss the Markovian model with unknown transition matrix in Section 5.

Our main contributions can be summarized as follows:

1. In a first scenario, we assume the existence of a single cognitive user capable of accessing only a single channel at any given time. In this setting, we derive an optimal sensing rule that maximizes the expected throughput obtained by the cognitive user. Compared with a genie-aided scheme, in which the cognitive user knows the primary network traffic information a priori, there is a throughput loss suffered by any medium access strategy. We obtain a lower bound on this loss and further construct a linear complexity single-index protocol that achieves this lower bound asymptotically (when the primary traffic behavior changes very slowly).
2. In a second scenario, we extend our work to the case in which the cognitive user is capable of accessing more than one channel simultaneously. The optimal solution as well as a low-complexity order-optimal solution are derived.
3. In a third scenario, we design distributed sensing rules by which the cognitive users take the competition from other cognitive users into consideration when making sensing decisions. We first characterize the optimal distributed sensing rule for the case in which the primary network statistics are available to the cognitive users. Under this idealistic assumption, we show that the throughput loss of the proposed distributed sensing rule, compared with a throughput-optimal centralized scheme, goes to zero exponentially as the number of cognitive users increases. To prevent any possible misbehavior by the cognitive users, we further design a game theoretically fair sensing rule, whose loss compared with the throughput-optimal centralized rule also goes to zero exponentially. Building on these results, we then devise distributed sensing rules that do not require prior knowledge about the traffic and converge to the optimal distributed rule and game theoretically fair rule, respectively.
4. Finally, the scenario with multiple cognitive users each having the ability to access more than one channel is considered. The optimal solution and low-complexity algorithms are developed.

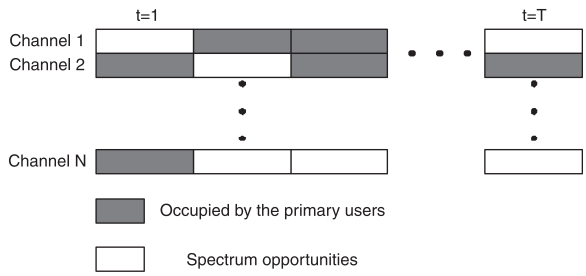


Fig. 1. Channel model.

The rest of the paper is organized as follows: Our network model is detailed in Section 2. Section 3 analyzes the scenario with a single cognitive user capable of sensing one channel or multiple channels at a time. The extension to the multiuser case is developed in Section 4. Section 5 discusses some practical issues and presents numerical results that support our theoretical claims. Finally, Section 6 summarizes our conclusions.

2 Network Model

Throughout this paper, uppercase letters (e.g., X) denote random variables, lowercase letters (e.g., x) denote realizations of the corresponding random variables, and calligraphic letters (e.g, $\mathcal{X}$) denote finite alphabet sets over which corresponding variables range. Also, uppercase boldface letters (e.g., $\mathbf{X}$) denote random vectors and lowercase boldface letters (e.g., $\mathbf{x}$) denote realizations of the corresponding random vectors.

Fig. 1 shows the channel model of interest. We consider a primary network consisting of N channels, $\mathcal{N} = \{1, \dots, N\}$, each with bandwidth B.[1] The users in the primary network are operated in a synchronous time-slotted fashion. We use i to refer to the channel index, j to refer to the time slot index, and k to refer to the index of the cognitive users. We assume that at each time slot, channel i is free with probability θ_i. Let $Z_i(j)$ be a random variable that equals 1 if channel i is free (i.e., available) at time slot j and equals 0 otherwise. Hence, given θ_i, $Z_i(j)$ is a Bernoulli random variable with probability density function (pdf)

$$h_{\theta_i}(z_i(j)) = \theta_i \delta(z_i(j) - 1) + (1 - \theta_i)\delta(z_i(j)),$$

where $\delta(\cdot)$ is the delta function. Furthermore, for a given $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$, $Z_i(j)$ are independent for each i and j. That is, for a given $\boldsymbol{\theta}$, whether a channel is free or not is independent between the time slots. The generalization to the Markovian model for multiple cognitive user case is briefly discussed in Section 5. We consider a block varying model in which the value of $\boldsymbol{\theta}$ is fixed for a block of T time slots and randomly changes at the beginning of the next block according to some joint pdf $f(\boldsymbol{\theta})$.

In our model, the cognitive users attempt to exploit the availability of free channels in the primary network by sensing the activity at the beginning of each time slot. Our work seeks to characterize efficient strategies for choosing

1. The case for different bandwidths B_i can be converted to an equivalent problem in which channel i is free with probability $\theta_{i,eq} = \theta_i B_i / B_{\max}$, and each channel has a bandwidth $B_{\max}$. Here, $B_{\max} = \max_{i \in \mathcal{N}} B_i$.

which channels to sense (access). The challenge here stems from the fact that the cognitive users are assumed to be unaware of $\boldsymbol{\theta}$ a priori. We consider two cases in which the cognitive user either has or does not have prior information about the pdf of $\boldsymbol{\theta}$, i.e., $f(\boldsymbol{\theta})$.

To further illustrate the point, let us consider a first scenario in which a single cognitive user is capable of sensing only one channel at a time. At time slot j, the cognitive user selects one channel $S(j) \in \mathcal{N}$ to access. If the sensing result shows that channel $S(j)$ is free, i.e., $Z_{S(j)}(j) = 1$, the cognitive user can send B bits over this channel; otherwise, the cognitive user will wait until the next time slot and pick a possibly different channel to access.[2] The outcome of the sensing algorithm is first assumed to be error-free. The effects of sensing errors are briefly discussed in Section 5. Therefore, the total number of bits that the cognitive user is able to send over one block (of T time slots) is

$$W = \sum_{j=1}^{T} B Z_{S(j)}(j). \tag{1}$$

It is now clear that W is a random variable that depends on the traffic in the primary network and, more importantly for us, on the medium access protocols employed by the cognitive user. Therefore, the overarching goal of Section 3.1 is to construct low-complexity medium access protocols that maximize

$$\mathbb{E}\{W\} = \mathbb{E}\left\{\sum_{j=1}^{T} B Z_{S(j)}(j)\right\}. \tag{2}$$

Intuitively, the cognitive user would like to select that channel with the highest probability of being free in order to obtain more transmission opportunities. If $\boldsymbol{\theta}$ is known then this problem is trivial: the cognitive user should choose the channel $i^* = \arg\max_{i\in\mathcal{N}} \theta_i$ to sense. The uncertainty in $\boldsymbol{\theta}$ imposes a fundamental trade-off between exploration, in order to learn $\boldsymbol{\theta}$, and exploitation, by accessing the channel with the highest estimated availability probability based on currently available information.

Similarly, for scenarios in which the cognitive user can access more than one channel at a time or there are multiple cognitive users, our goal is to design strategies that maximize the throughput of the cognitive users, as detailed in the following sections.

3 Single-User Analysis

3.1 Single Channel

We start by developing the optimal solution to the single user-single channel scenario under the idealized assumption that $f(\boldsymbol{\theta})$ is known a priori by the cognitive user. As argued below, the optimal medium access algorithm suffers from prohibitive computational complexity that grows exponentially with the block length T. This motivates the design of low-complexity asymptotically optimal approaches that are considered next. Interestingly, the proposed low-complexity technique does not require prior knowledge about $f(\boldsymbol{\theta})$.

2. The case in which the cognitive user can sense another channel immediately is considered in [11] and [12].

3.1.1 *Bayesian Approach*

Our single user-single channel cognitive medium access problem belongs to the class of bandit problems. In this setting, the decision maker must sequentially choose one process to observe from $N \geq 2$ stochastic processes. These processes usually have parameters that are unknown to the decision maker, and associated with each observation is a utility function. The objective of the decision maker is to maximize the sum or discounted sum of the utilities via a strategy that specifies which process to observe for every possible history of selections and observations. The following classical example illustrates the challenge facing our decision maker: A gambler enters a casino having N slot machines, the ith of which has winning probability $\theta_i, i \in \mathcal{N}$. The gambler does not know the values of the θ_is and must sequentially choose machines to play. The goal is to maximize the overall gain for a total of T plays. In this example, the stochastic processes are the outcomes of the slot machines, the utility function is the reward that the gambler gains each time, and the gambling strategy specifies which machine to play based on each possible past information pattern. A comprehensive treatment covering different variants of bandit problems can be found in [6].

We are now ready to rigorously formulate our problem. The cognitive user employs a medium access strategy Γ, which will select channel $S(j) \in \mathcal{N}$ to sense at time slot j for any possible causal information pattern obtained through the previous $j-1$ observations:

$$\Psi(j) = \{s(1), z_{s(1)}(1), \ldots, s(j-1), z_{s(j-1)}(j-1)\}, \quad j \geq 2,$$

i.e., $s(j) = \Gamma(f, \Psi(j))$. Notice that $z_{s(j)}(j)$ is the sensing outcome of the jth time slot, in which $s(j)$ is the channel being accessed. If $j = 1$, there is no accumulated information, thus $\Psi(1) = \phi$ and $s(1) = \Gamma(f)$. Γ could be stochastic, i.e., for certain $\Psi(j)$, the cognitive user may randomly choose channel i from a set $\mathcal{A} \subseteq \mathcal{N}$ with probability p_i, such that $\sum_{i\in\mathcal{A}} p_i = 1$. The utility that a cognitive user obtains by making decision $S(j)$ at time slot j is the number of bits it can transmit at time slot j, which is $BZ_{S(j)}(j)$. We denote the expected value of the payoff obtained by a cognitive user using strategy Γ as

$$W_\Gamma = \mathbb{E}_f\left\{\sum_{j=1}^{T} B Z_{S(j)}(j)\right\}. \tag{3}$$

We denote $V^*(f, T) = \sup_\Gamma W_\Gamma$, which is the largest throughput that the cognitive user could obtain when the spectral opportunities are governed by $f(\boldsymbol{\theta})$ and the exact value of each realization of $\boldsymbol{\theta}$ is not known by the user.

Each medium access decision made by the cognitive user has two effects. The first one is the short-term gain, i.e., an immediate transmission opportunity if the chosen channel is found to be free. The second one is the long-term gain, i.e., the updated statistical information about $\boldsymbol{\theta}$. This information will help the cognitive user in making better decisions in the future stages. There is an interesting trade-off between the short and long-term gains. If we only want to maximize the short-term gain, we can choose the channel with the highest availability probability to sense, based on the current information. This myopic strategy maximally exploits the existing information. On the other hand, by choosing other channels to sense, we gain valuable statistical information about $\boldsymbol{\theta}$ which can effectively guide future decisions. This process is typically referred to as exploration.

More specifically, let $f^j(\boldsymbol{\theta})$ be the updated pdf of $\boldsymbol{\theta}$ after making $j-1$ observations. We begin with $f^1(\boldsymbol{\theta}) = f(\boldsymbol{\theta})$. After observing $z_{s(j)}(j)$, we update the pdf using the following Bayesian formula:

1. If $z_{s(j)}(j) = 1$,

$$f^{j+1}(\boldsymbol{\theta}) = \frac{\theta_{s(j)} f^j(\boldsymbol{\theta})}{\int \zeta_{s(j)} f^j(\boldsymbol{\zeta}) d\boldsymbol{\zeta}}. \tag{4}$$

2. If $z_{s(j)}(j) = 0$,

$$f^{j+1}(\boldsymbol{\theta}) = \frac{(1-\theta_{s(j)}) f^j(\boldsymbol{\theta})}{\int (1-\zeta_{s(j)}) f^j(\boldsymbol{\zeta}) d\boldsymbol{\zeta}}. \tag{5}$$

Now, [6, Lemma 2.3.1] proves that every bandit problem with finite horizon has an optimal solution. Applying this result to our setup, we obtain the following:

Lemma 1. *For any prior pdf f, there exists an optimal strategy Γ^* to the channel selection problem (3), and $V^*(f,T)$ is achievable. Moreover, V^* satisfies the following condition:*

$$V^*(f,T) = \max_{s(1)\in\mathcal{N}} \mathbb{E}_f\{BZ_{s(1)} + V^*(f_{Z_{s(1)}}, T-1)\}, \tag{6}$$

where $f_{Z_{s(1)}}$ is the conditional pdf updated using (4) and (5) as if the cognitive user chooses $s(1)$ and observes $Z_{s(1)}$. Also, $V^*(f_{Z_{s(1)}}, T-1)$ is the value of a bandit problem with prior information $f_{Z_{s(1)}}$ and $T-1$ sequential observations.

In principle, Lemma 1 provides the solution to problem (3). Effectively, it decouples the calculation at each stage, and hence, allows the use of dynamic programming to solve the problem. The idea is to solve the channel selection problem with a smaller dimension first and then use backward induction to obtain the optimal solution for a problem with a larger dimension. Starting with $T=1$, the second term inside the expectation in (6) is 0, since $T-1=0$. Hence, the optimal solution is to choose channel i having the largest value of $\mathbb{E}_f\{BZ_i\}$, which can be calculated as

$$\mathbb{E}_f\{BZ_i\} = B\int \theta_i f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

and $V^*(f,1) = \max_{i\in\mathcal{N}} \mathbb{E}_f\{BZ_i\}$. One can now proceed to solve the case with $T=2$ and so on.

3.1.2 Nonparametric Asymptotic Analysis and Asymptotically Optimal Strategies

The optimal solution developed in Section 3.1.1 suffers from prohibitive computational complexity. In particular, the dimensionality of our search dimension grows exponentially with the block length T. Moreover, one can envision many practical scenarios in which it would be difficult for the cognitive user to obtain the prior information $f(\boldsymbol{\theta})$. This motivates our pursuit of low-complexity nonparametric protocols. Toward this end, we study in the following asymptotic performance of several low-complexity approaches. In particular, we analyze nonparametric schemes that do not explicitly use $f(\boldsymbol{\theta})$, and thus the rule Γ considered in this section depends only on $\Psi(j)$ explicitly.

For a certain strategy Γ, the expected number of bits the cognitive user is able to transmit through a block with certain parameters $\boldsymbol{\theta}$ is

$$\mathbb{E}\left\{\sum_{j=1}^{T} BZ_{S(j)}(j)\right\} = \sum_{j=1}^{T} B \sum_{i=1}^{N} \theta_i \Pr\{\Gamma(\Psi(j)) = i\}. \tag{7}$$

Recall that $\Gamma(\Psi(j)) = i$ means that, following strategy Γ, the cognitive user should choose channel i at time slot j, based on the available information $\Psi(j)$. Here, $\Pr\{\Gamma(\Psi(j)) = i\}$ is the probability that the cognitive user will choose channel i at time slot j, following the strategy Γ.

Compared with the ideal case where the exact value of $\boldsymbol{\theta}$ is known, in which the optimal strategy for the cognitive user is to always choose the channel with the largest availability probability, the loss entailed by Γ is given by

$$L(\boldsymbol{\theta};\Gamma) = \sum_{j=1}^{T} B\theta_{i^*} - \sum_{j=1}^{T} B \sum_{i=1}^{N} \theta_i \Pr\{\Gamma(\Psi(j)) = i\}, \tag{8}$$

where $\theta_{i^*} = \max\{\theta_1, \ldots, \theta_N\}$. We say that a strategy Γ is consistent if, for any $\boldsymbol{\theta} \in [0,1]^N$, there exists $\beta < 1$ such that $L(\boldsymbol{\theta};\Gamma)$ scales as[3] $O(T^\beta)$. For example, consider a loyal scheme in which the cognitive user selects channel i at the beginning of a block and sticks to it. If θ_i is the largest one among $\boldsymbol{\theta}$, $L(\boldsymbol{\theta};\Gamma) = 0$. On the other hand, if θ_i is not the largest one, $L(\boldsymbol{\theta};\Gamma) \sim O(T)$. Hence, this loyal scheme is not consistent. The following lemma characterizes the fundamental limits of any consistent scheme.

Lemma 2. *For any $\boldsymbol{\theta}$ and any consistent strategy Γ, we have*

$$\liminf_{T\to\infty} \frac{L(\boldsymbol{\theta};\Gamma)}{\ln T} \geq B \sum_{i\in\mathcal{N}\setminus\{i^*\}} \frac{\theta_{i^*} - \theta_i}{D(\theta_i \| \theta_i^*)}, \tag{9}$$

where $D(\theta_i\|\theta_l)$ is the Kullback-Leibler divergence between two Bernoulli random variables with parameters θ_i and θ_l:

$$D(\theta_i\|\theta_l) = \theta_i \ln(\theta_i/\theta_l) + (1-\theta_i)\ln((1-\theta_i)/(1-\theta_l)).$$

Proof. This follows as an application of a theorem proved in [13]; see Appendix A for details. □

Lemma 2 shows that the loss of any consistent strategy scales at least as $\omega(\ln T)$. An intuitive explanation of this loss is that we need to spend at least $O(\ln T)$ time slots on sampling each of the channels with smaller θ_i, in order to get a reasonably accurate estimate of $\boldsymbol{\theta}$, and hence, use it to determine the channel having the largest θ_i to sense. We say that a strategy Γ is order optimal if $L(\boldsymbol{\theta};\Gamma) \sim O(\ln T)$.

Now, the first question that arises is whether there exist order-optimal strategies. As shown later in this section, we can design suboptimal strategies that have loss

3. In this paper, we use Knuth's asymptotic notations: 1) $g_1(N) = o(g_2(N))$ means $\forall c > 0, \exists N_0, \forall N > N_0, g_1(N) < cg_2(N)$, 2) $g_1(N) = \omega(g_2(N))$ means $\forall c > 0, \exists N_0, \forall N > N_0, g_2(N) < cg_1(N)$, and 3) $g_1(n) = O(g_2(N))$ means $\exists c_2 \geq c_1 > 0, N_0, \forall N > N_0, c_1 g_2(N) \leq g_1(N) \leq c_2 g_2(N)$.

of order $O(\ln T)$. Thus, the answer to this question is affirmative. Before proceeding to the proposed low-complexity order-optimal strategy, we first analyze the loss order of some heuristic strategies that may appear appealing in certain applications.

The first simple rule is the random strategy Γ_r where, at each time slot, the cognitive user randomly chooses a channel from the available N channels. The fraction of time slots the cognitive user spends on each channel is therefore $1/N$, leading to the loss

$$L(\boldsymbol{\theta}; \Gamma_r) = BT \sum_{i=1}^{N} (\theta_{i^*} - \theta_i)/N \sim O(T).$$

The second protocol of interest is the myopic rule Γ_g in which the cognitive user keeps updating $f^j(\boldsymbol{\theta})$, and chooses the channel with the largest value of $\hat{\theta}_i = \int \zeta_i f^j(\zeta) d\zeta$ at each stage. Since there are no convergence guarantees for the myopic rule, that is $\hat{\boldsymbol{\theta}}$ may never converge to $\boldsymbol{\theta}$ due to the lack of sufficiently many samples for each channel [14], the loss of this myopic strategy is $O(T)$.

The third protocol we consider is *staying with the winner and switching from the loser rule* Γ_{SW} where the cognitive user randomly chooses a channel in the first time slot. In the succeeding time slots 1) if the accessed channel is found to be free, it will choose the same channel to sense; 2) otherwise, it will choose one of the remaining channels based on a certain switching rule.

Lemma 3. *Regardless of the switching rule,* $L(\boldsymbol{\theta}; \Gamma_{SW}) \sim O(T)$.

Proof. Please refer to Appendix B. □

There are several strategies that have loss of order $O(\ln T)$. We adopt the following linear complexity strategy which was proposed and analyzed in [15].

Rule 1 (Order-optimal single-index strategy). *The cognitive user maintains two vectors* $\mathbf{X}$ *and* $\mathbf{Y}$, *where* X_i *records the number of time slots for which the cognitive user has sensed channel* i *to be free, and* Y_i *records the number of time slots for which the cognitive user has chosen channel* i *to sense. The strategy works as follows:*

1. *Initialization: at the beginning of each block, sense each channel once.*
2. *After the initialization period, the cognitive user obtains an estimate* $\hat{\boldsymbol{\theta}}$ *at the beginning of time slot* j, *given by* $\hat{\theta}_i(j) = X_i(j)/Y_i(j)$, *and assigns an index*

$$\Lambda_i(j) = \hat{\theta}_i(j) + \sqrt{2 \ln j / Y_i(j)}$$

to the i*th channel. The cognitive user chooses the channel with the largest value of* $\Lambda_i(j)$ *to sense at time slot* j. *After each sensing, the cognitive user updates* $\mathbf{X}$ *and* $\mathbf{Y}$.

The intuition behind this strategy is that as long as Y_i grows as fast as $O(\ln T)$, Λ_i converges to the true value of θ_i in probability, and the cognitive user will choose the channel with the largest θ_i eventually. The loss of $O(\ln T)$ comes from the time spent in sampling the inferior channels in order to learn the value of $\boldsymbol{\theta}$. This price, however, is inevitable as established in the lower bound of Lemma 2.

Finally, we observe that the difference between the myopic rule and the order-optimal single-index rule is the additional term $\sqrt{2 \ln j / Y_i(j)}$ added to the current estimate $\hat{\theta}_i$. Roughly speaking, this additional term guarantees enough sampling time for each channel, since if we sample channel i too sparsely, $Y_i(j)$ will be small, which will increase the probability that Λ_i is the largest index. When $Y_i(j)$ scales as $\ln T$, $\hat{\theta}_i$ will be the dominant term in the index Λ_i, and hence the channel with the largest θ_i will be chosen more frequently.

3.2 Multichannel Cognitive Users

In certain scenarios, cognitive users may be able to sense more than one channel simultaneously. We assume the presence of a single cognitive user capable of sensing, and subsequently utilizing, $M \leq N$ channels simultaneously. Let $\mathcal{M}(j)$ be the set of channels the cognitive user selects to sense at time slot j, where $|\mathcal{M}(j)| = M$. The average number of bits that the cognitive user is able to send over a block is therefore

$$\mathbb{E}\{W\} = \mathbb{E}\left\{ \sum_{j=1}^{T} \sum_{S(j) \in \mathcal{M}(j)} B Z_{S(j)}(j) \right\}. \tag{10}$$

At the beginning of time slot j, the cognitive user can update the pdf $f^j(\boldsymbol{\theta})$ similarly to (4) and (5). Similarly to Lemma 1, the optimal solution can be characterized by the following optimality condition:

$$\begin{aligned} V^*(f, T) = \max_{\mathcal{M}(1) \subseteq \mathcal{N}, |\mathcal{M}(1)| = M} \mathbb{E}_f \Bigg\{ & \sum_{s(1) \in \mathcal{M}(1)} B Z_{s(1)} \\ & + V^*\left(f_{\{Z_{s(1)} : s(1) \in \mathcal{M}(1)\}}, T - 1 \right) \Bigg\}. \end{aligned} \tag{11}$$

Here, $f_{\{Z_{s(1)} : s(1) \in \mathcal{M}(1)\}}$ is the updated pdf after observing the sensing output of the channels $s(1) \in \mathcal{M}(1)$. We can then follow the same procedure described for the single-channel sensing scenario to obtain the optimal strategy Γ^* according to (11). In the following, however, we focus on low-complexity nonparametric strategies that are asymptotically optimal.

If $\boldsymbol{\theta}$ is known, the cognitive user will choose the M channels with the largest θ's to sense. Without loss of generality, we assume $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_N$. Hence, for any strategy Γ, the loss is

$$L(\boldsymbol{\theta}; \Gamma) = \sum_{j=1}^{T} \sum_{i=1}^{M} B\theta_i - \sum_{j=1}^{T} B \sum_{i=1}^{N} \theta_i P\{i \in \mathcal{M}(j)\}. \tag{12}$$

We have the following order-optimal simple single-index strategy.

Rule 2. *The cognitive user maintains two vectors* $\mathbf{X}$ *and* $\mathbf{Y}$, *where* X_i *is the number of time slots in which the cognitive user has sensed channel* i *to be free, and* Y_i *is the number of time slots in which the cognitive user has chosen channel* i *to sense. The strategy works as follows:*

1. *Initialization: at the beginning of each block, each channel is sensed once. This initialization stage takes* $\lceil N/M \rceil$ *time slots, in which* $\lceil x \rceil$ *denotes the least integer that is no less than* x.

2. *After the initialization period, the cognitive user obtains an estimate* $\hat{\boldsymbol{\theta}}$ *at the beginning of time slot* j *given by* $\hat{\theta}_i(j) = X_i(j)/Y_i(j)$, *and assigns an index*

$$\Lambda_i(j) = \hat{\theta}_i(j) + \sqrt{2 \ln j / Y_i(j)}$$

to the i*th channel. The cognitive user orders these* $\Lambda_i(j)$*s and selects the* M *channels with the largest* $\Lambda_i(j)$*s to sense. After each sensing, the cognitive user updates* $\mathbf{X}$ *and* $\mathbf{Y}$.

Lemma 4. *Rule 2 is asymptotically optimal and* $L(\boldsymbol{\theta}, \Gamma) \sim O(\ln T)$.

Proof. Please refer to Appendix C. □

In the proof presented in Appendix C, the performance bound is derived for general T. Hence, compared with the situation in which $\boldsymbol{\theta}$ is known, the performance loss of the order-optimal rules is bounded by $c_1 \ln T + c_2$ for certain constants c_1 and c_2. Thus, the scheme is not only asymptotically optimal, but also has a performance guarantee for finite value of T.

4 Multiuser Analysis

In this section, we assume the presence of a set $\mathcal{K} = \{1, \dots, K\}$ of cognitive users and consider the distributed medium access decision processes of the multiple users with no prior coordination. The presence of multiple cognitive users adds an element of competition to the problem. In order for a cognitive user to make use of a channel now, it must be free of the primary traffic and traffic from the other competing cognitive users.

4.1 Single Channel

We start with a simpler situation in which each cognitive user can sense one channel at a time. We denote by $\mathcal{K}_i(j) \subseteq \mathcal{K}$ the random set of users who choose to sense channel i at time slot j. We assume that the users follow a generalized version of the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) protocol to access the channel after sensing the main channel to be free, i.e., if channel i is free, each user k in the set $\mathcal{K}_i(j)$ will generate a random number $t_k(j)$ according to a certain probability density function g, and wait the time specified by the generated random number. At the end of the waiting period, user k senses the channel again, and if it is found to be free, the packet from user k will be transmitted. The probability that user k in the set $\mathcal{K}_i(j)$ gains access to the channel is the same as the probability that $t_k(j)$ is the smallest random number generated by the users in the set $\mathcal{K}_i(j)$. Thus, the throughput[4] user k achieves in a block is

$$W_k = \sum_{j=1}^{T} B Z_{S_k(j)}(j) I\left\{ k = \arg \min_{q \in \mathcal{K}_{S_k(j)}(j)} t_q(j) \right\}. \tag{13}$$

4. The impact of the CSMA/CA overhead will be discussed in Section 5.4.

Therefore, user k should devise a sensing rule Γ_k that maximizes

$$\mathbb{E}\{W_k\} = \mathbb{E}\left\{ \sum_{j=1}^{T} B Z_{S_k(j)}(j) I\left\{ k = \arg \min_{q \in \mathcal{K}_{S_k(j)}(j)} t_q(j) \right\} \right\}.$$

Clearly, with multiple cognitive users, it is not optimal anymore for all the users to always choose the channel with the largest θ_i to sense. In particular, if all the users choose the channel with the largest θ_i, the probability that a given user gains control of the channel decreases, while potential opportunities in the other channels in the primary network are wasted.

4.1.1 Known θ Case

To enable a succinct presentation, we first consider the case in which the values of $\boldsymbol{\theta}$ are known to all the cognitive users. The users distributively choose channels to sense and compete for access if the channels are free.

Without loss of generality, we consider a mixed strategy where user k will choose channel i with probability $p_{k,i}$. Furthermore, we let $\mathbf{p}_k = [p_{k,1}, \dots, p_{k,N}]$ and consider the symmetric solution in which $\mathbf{p} = \mathbf{p}_1 = \cdots = \mathbf{p}_K$. The symmetry assumption implies that all the users in the network distributively follow the same rule to access the spectral opportunities present in the primary network, in order to maximize the same average throughput that each user can obtain. The following result derives the optimal solution in this situation.

Lemma 5. *For a cognitive network with* $K > 1$ *cognitive users and* N *channels with probabilities* $\boldsymbol{\theta}$ *of being free, the optimal* $\mathbf{p}^*$ *is given by*

$$p_i^* = \begin{cases} \left\{ 1 - \left(\frac{\lambda^*}{K\theta_i} \right)^{1/(K-1)} \right\}^+, & \text{for} \quad \theta_i > 0, \\ 0, & \text{for} \quad \theta_i = 0, \end{cases} \tag{14}$$

where λ^* *is a constant such that* $\sum_{i=1}^N p_i^* = 1$. *Here,* $\{x\}^+ = \max\{0, x\}$.

Proof. Please refer to Appendix D. □

Note that, if $K = 1$, then $p_{i^*}^* = 1$ satisfies the Karush-Kuhn-Tucker (KKT) conditions for optimality [16]. Here, $i^* = \arg\max_{i \in \mathcal{N}} \theta_i$, $p_l^* = 0$, and $l \in \mathcal{N} \backslash \{i^*\}$.

So, the total throughput of the K cognitive users is

$$\begin{aligned} KW &= BKT \sum_{i=1}^{N} \frac{\theta_i}{K} \left\{ 1 - \left(1 - p_i^*\right)^K \right\} \\ &= BT \sum_{i=1}^{N} \theta_i \left\{ 1 - \left(1 - p_i^*\right)^K \right\}. \end{aligned} \tag{15}$$

On the other hand, the average total spectral opportunity in the primary network is $BT \sum_{i=1}^N \theta_i$. This upper bound can be achieved by a centralized channel allocation strategy when $K > N$ (simply by assigning one cognitive user to each channel). Therefore, the loss of the distributed protocol as compared with the centralized scheduling is

$$L = BT \sum_{i=1}^{N} \theta_i \left(1 - p_i^*\right)^K. \tag{16}$$

There is an intuitive explanation of this loss. If there is a spectral opportunity in channel i but there are no users choosing channel i to sense, a loss occurs. The probability that there is no user choosing channel i to sense is $(1-p_i^*)^K$, and hence the probability of loss occurring at channel i is $\theta_i(1-p_i^*)^K$. To obtain further insights into the performance of the cognitive network, we study the following special cases:

$\diamond$ $N \geq 1, K = 1$. As stated above, $p_{i^*}^* = 1$, and $p_l^* = 0$, $l \in \mathcal{N}\backslash\{i^*\}$. Hence, the user should choose the channel with the largest availability probability to sense, and

$$L = BT \sum_{i \in \mathcal{N}\backslash\{i^*\}} \theta_i.$$

$\diamond$ $N = 2, K = 2$. Substituting $N = 2$ and $K = 2$ into (14), we obtain

$$p_1^* = \theta_1/(\theta_1+\theta_2) \quad \text{and} \quad p_2^* = \theta_2/(\theta_1+\theta_2). \tag{17}$$

Furthermore,

$$\begin{aligned} W &= \frac{BT\theta_1}{2}\left[1 - \frac{\theta_2^2}{(\theta_1+\theta_2)^2}\right] \\ &\quad + \frac{BT\theta_2}{2}\left[1 - \frac{\theta_1^2}{(\theta_1+\theta_2)^2}\right], \quad \text{and} \\ L &= \frac{BT\theta_1\theta_2}{2(\theta_1+\theta_2)}. \end{aligned} \tag{18}$$

$\diamond$ N is fixed, and $K \to \infty$. We have the following asymptotic characterization:

Lemma 6. *Let $2 \leq Q \leq N$ be the number of channels for which $\theta_i > 0$. We have $p_i^* \to 1/Q$, and $L \to 0$ exponentially as K increases, i.e., $L \sim O(e^{-c_1 K})$, where $c_1 = \ln \frac{Q}{Q-1}$.*

Proof. Please refer to Appendix E. □

From this lemma, we can see that if the number of cognitive users is large, the optimal strategy is independent of the exact value of $\boldsymbol{\theta}$. The reason for the exponential decrease in the loss is that, as the number of cognitive users increases, the probability that there is no user sensing any particular channel decreases exponentially. If $Q = 1$, then there is no loss of performance, since all the users will always sense the channel with nonzero availability probability.

The optimality of the distributed protocol proposed above hinges on the assumption that all the users will follow the symmetric rule. However, it is straightforward to see that if a single cognitive user deviates from the rule specified in Lemma 5, it will be able to transmit more bits. If this selfish perspective propagates through the network, it may lead to a significant reduction in the overall throughput. This observation motivates our next step in which the channel selection problem is modeled as a noncooperative game, where the cognitive users are the players, the Γ_ks are the strategies, and the average throughput of each user is the payoff. The following result derives a sufficient condition for the Nash equilibrium in the asymptotic scenario $K \to \infty$.

Lemma 7. *$(\Gamma_1, \ldots, \Gamma_K)$ is a Nash equilibrium if K is sufficiently large and, at each time slot, there are $\tau_i K$ users sensing channel i, where τ_i satisfies*

$$\tau_i = \theta_i \Big/ \sum_{i=1}^{N} \theta_i. \tag{19}$$

At this equilibrium, each user has probability $\sum_{i=1}^N \theta_i/K$ of transmitting at each time slot.

Proof. We prove this result by backward induction. At the last time slot T, if the τ_is satisfy (19), the probability of user k gaining a channel is

$$p_k = \theta_i/(\tau_i K) = \sum_{i=1}^{N} \theta_i/K. \tag{20}$$

Now, if user k deviates from this strategy, and chooses channel i', the number of users sensing channel i' is $\tau_{i'}K + 1$, and the probability of user k gaining the channel is

$$p_k' = \theta_{i'}/(\tau_{i'}K+1) < \theta_{i'}/(\tau_{i'}K) = p_k. \tag{21}$$

Hence, the strategy that has $\tau_i K$ users sensing channel i at time slot T is a Nash equilibrium. Now, we know the optimal strategy for the last time slot, so we can ignore this time slot. Then time slot $T-1$ becomes the last slot, in which this strategy is optimal. Similarly, we show by induction that this strategy is optimal for all other time slots. □

We note that in the lemma we implicitly assume that $\tau_i K$ is an integer. In practice, this is not always true. However, if K is large, then rounding $\tau_i K$ to the nearest integer will have minimal effect. The Nash equilibrium is also optimal from a system perspective, in the sense that this strategy maximizes the entire throughput of the network by fully utilizing the available spectral opportunities when K is large (i.e., on the average, each user will be able to transmit $(BT\sum\theta_i)/K$ bits per block, and the total throughput of the network is $BT\sum\theta_i$).

With this equilibrium result, the cognitive users can use the following stochastic sensing strategy to approximately work on the equilibrium point for a large but finite K. Let $s_k(j)$ be the channel chosen by user k at time slot j. At each time slot, each user independently selects channel i with probability $\tau_i = \theta_i/\sum_{i\in\mathcal{N}}\theta_i$, i.e., $\Pr\{s_k(j) = i\} = \tau_i$. Then, at each time slot, the number of users sensing channel i will be $\sum_{k=1}^K I\{s_k(j) = i\}$, where the $I\{s_k(j) = i\}$s are independent and identically distributed (i.i.d) Bernoulli random variables. Hence, the total number of users sensing channel i is a binomial random variable, and the fraction of users sensing channel i converges to τ_i in probability as K increases, i.e.,

$$\tau' = \left(\sum_{k=1}^{K} I\{s_k(j) = i\}\right) \Big/ K \to \tau_i \tag{22}$$

in probability. Hence, as K increases, the operating point will converge to the Nash equilibrium in probability.

For any K, the probability that there is no user choosing channel i to sense is $(1-\tau_i)^K$. Hence, the performance loss compared with the centralized scheme is

$$L = BT\sum \theta_i(1-\tau_i)^K = BT\sum_{i=1}^{N}\theta_i\left(\frac{\sum_{l=1}^{N}\theta_l-\theta_i}{\sum_{l=1}^{N}\theta_l}\right)^K. \quad (23)$$

It is easy to check that

$$\lim_{K\to\infty}\frac{L}{\exp^{-c_2K}} = BT\theta_{l^*}, \quad (24)$$

where $\theta_{l^*} = \min\{\theta_i : \theta_i > 0\}$, and

$$c_2 = \ln\frac{\sum\theta_i}{\sum_{l=1}^{N}\theta_l-\theta_{l^*}}.$$

It is now clear that the loss of the game-theoretic scheme goes to zero exponentially, though the decay rate is smaller than that of the scheme specified in Lemma 5. On the other hand, compared with the scheme in Lemma 5, the game-theoretic scheme has the advantage that the individual cognitive users do not need to know the total number of cognitive users K in the network and, more importantly, they have no incentive to deviate unilaterally.

4.1.2 Unknown θ Case

If $\boldsymbol{\theta}$ is unknown, the cognitive users need to estimate $\boldsymbol{\theta}$ (in addition to resolving their competition). Combining the results from Sections 3.1.2 and 4.1.1, we design the following low-complexity strategy which is asymptotically optimal.

Rule 3.

1. *Initialization: Each user k maintains the following two vectors: $\mathbf{X}_k$, which records the number of time slots in which user k has sensed each channel to be free; and $\mathbf{Y}_k$, which records the number of time slots in which user k has sensed each channel. At the beginning of each block, user k senses each channel once and transmits through this channel if the channel is free and it wins the competition. Also, set $X_{k,i}=1$, regardless of the sensing result of this stage.*
2. *At the beginning of time slot j, user k estimates $\hat{\theta}_i$ as*

$$\hat{\theta}_i(j) = X_{k,i}(j)/Y_{k,i}(j),$$

and chooses each channel $i \in \mathcal{N}$ with probability

$$\hat{\theta}_i(j)\Big/\sum_{i=1}^{N}\hat{\theta}_i(j). \quad (25)$$

After each sensing, $\mathbf{X}_k$ and $\mathbf{Y}_k$ are updated.

Lemma 8. *If K is large, the scheme in Rule 3 converges to the Nash equilibrium specified in Lemma 7 in probability, as T increases.*

Proof. Please refer to Appendix F. □

The intuition behind this scheme is that, each user will sample each channel at least $O(T)$ times, and hence as T increases, the estimate $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}$ in probability, implying that the unknown $\boldsymbol{\theta}$ case will eventually reduce to the case in which $\boldsymbol{\theta}$ is known to all the users. Hence, if K is sufficiently large, the operating point converges to the Nash equilibrium in probability.

If one can assume that the users will follow the prespecified rule, then we can design the following strategy that converges to the optimal operating point in probability for any K, as T increases.

Rule 4.

1. *Initialization: Same as Rule 3.*
2. *At the beginning of time slot $j \le \ln T$, user k estimates $\hat{\theta}_i$ as*

$$\hat{\theta}_i(j) = X_{k,i}(j)/Y_{k,i}(j),$$

and chooses each channel $i \in \mathcal{N}$ with probability $\hat{\theta}_i(j)/\sum_{i=1}^{N}\hat{\theta}_i(j)$. For $j \ge \ln T$, the ith channel is sensed with probability

$$\hat{p}_i^* = \left\{1-(\lambda^*/\hat{\theta}_i)^{1/(K-1)}\right\}^+. \quad (26)$$

After each sensing, $\mathbf{X}_k$ and $\mathbf{Y}_k$ are updated.

Lemma 9. *The proposed scheme converges in probability to the optimal operating point specified in Lemma 5, as T increases.*

Proof. Following the same steps as in the proof of Lemma 8, one can show that $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}$ in probability as T increases. Hence the operating point specified by (26) converges in probability to the optimal point specified in Lemma 5 as T increases. □

4.2 Multiple Channels

Now, consider the scenario in which each cognitive user is able to sense and utilize M channels simultaneously. Let $\mathcal{M}_k(j)$ be the set of channels cognitive user k selects to sense at time slot j, where $|\mathcal{M}_k(j)| = M$. As in Section 4.1, we denote by $\mathcal{K}_i(j) \subseteq \mathcal{K}$ the random set of users who choose to sense channel i at time slot j. The users follow the same protocol described in Section 4.1 to access channel i, after sensing channel i to be free of primary users' traffic.

From N channels, we have $L = \binom{N}{M}$ different subsets, each containing M channels. We arbitrarily label these subsets from 1 to L, and let $\mathcal{A}_l$ be the lth subset. Let $p_{k,l}$ be the probability that user k chooses the set $\mathcal{A}_l$ to sense. Furthermore, we let $\mathbf{p}_k = [p_{k,1},\ldots,p_{k,L}]$. For simplicity, we consider only symmetric solutions here, and thus we assume $\mathbf{p} = \mathbf{p}_1 = \cdots = \mathbf{p}_K$. With $\mathbf{p}$, the probability that channel i will be selected by any user at each time slot is $\sum_{l:i\in\mathcal{A}_l} p_l$.

First, consider the situation in which $\boldsymbol{\theta}$ is known. Following the same argument as that of Section 4.1, the transmission opportunities per time slot not utilized is

$$y = B\sum_{i=1}^{N}\theta_i\left(1-\sum_{l:i\in\mathcal{A}_l}p_l\right)^K. \quad (27)$$

Hence, we can obtain the optimal value of $\mathbf{p}$ by solving the following optimization problem:

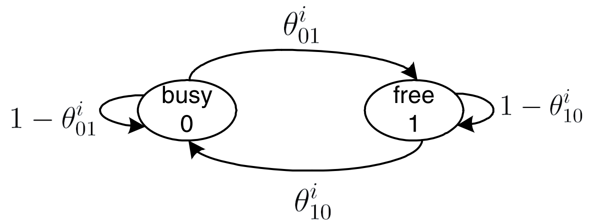


Fig. 2. State transition of channel i.

$$\min y = B\sum_{i=1}^{N}\theta_i\left(1-\sum_{l:i\in\mathcal{A}_l}p_l\right)^K \tag{28}$$

$$\text{s.t.} \sum_{l=1}^{L}p_l = 1 \quad \text{and} \quad \mathbf{p}\geq\mathbf{0}. \tag{29}$$

The KKT conditions for optimality are

$$\begin{aligned} &\mathbf{p}^*\geq\mathbf{0}, \quad \sum_{l=1}^{L}p_l^* = 1,\\ &p_l^*\left(\lambda^*-\sum_{i=1}^{N}\theta_i\left(1-\sum_{l:i\in\mathcal{A}_l}p_l^*\right)^{K-1}\right)=0, \quad \text{and}\\ &\lambda^*\geq\sum_{i=1}^{N}\theta_i\left(1-\sum_{l:i\in\mathcal{A}_l}p_l^*\right)^{K-1}, \end{aligned} \tag{30}$$

where λ^* is a Lagrange multiplier.

For the scenario in which $\boldsymbol{\theta}$ is unknown, we have the following simple rule that converges to the optimal operating point as T increases.

Rule 5.

1. *Initialization: Same as Rule 3.*
2. *At the beginning of time slot $j\leq\ln T$, user k estimates $\hat{\theta}_i$ as*

$$\hat{\theta}_i(j) = X_{k,i}(j)/Y_{k,i}(j).$$

Now, user k first chooses one channel $i\in\mathcal{N}$ with probability $\hat{\theta}_i(j)/\sum_{i=1}^{N}\hat{\theta}_i(j)$. Then, user k chooses the remaining $M-1$ channels randomly. For $j\geq\ln T$, the lth subset is chosen with probability computed using (30) with θ_i being replaced with the estimated value $\hat{\theta}_i$. After each sensing, $\mathbf{X}_k$ and $\mathbf{Y}_k$ are updated.

5 Further Discussions and Numerical Results

5.1 Markovian Model

Certain results presented above can be extended to a more general Markovian model. As shown in Fig. 2, we assume that the availability of channel i follows a Markov chain with parameters θ_{01}^i and θ_{10}^i, in which θ_{01}^i denotes the transition probability of channel i from the "busy" state (denoted by 0) to the "free" state (denoted by 1) and θ_{10}^i denotes the transition probability of channel i from the "free" state to the "busy" state. We assume that the value of θ_{01}^i and θ_{10}^i are unknown a priori to the cognitive users. We denote $\boldsymbol{\theta}_{01}=[\theta_{01}^1,\ldots,\theta_{01}^K]$ and $\boldsymbol{\theta}_{10}=[\theta_{10}^1,\ldots,\theta_{10}^K]$. In the following, we set

$$\theta_i^M = \frac{\theta_{01}^i}{\theta_{01}^i+\theta_{10}^i}, \tag{31}$$

and denote $\boldsymbol{\theta}^M=[\theta_1^M,\ldots,\theta_K^M]$.

If there is only one cognitive user in the network, i.e., the problem considered in Section 3, the problem under this Markovian model can be formulated as a restless bandit problem [17]. It has been shown that the restless bandit problem is PSPACE-hard [18]. Hence, finding the optimal solution for $K=1$ is a very difficult problem. We note that some interesting results have been obtained for the case in which the payoff is the discounted sum of payoffs obtained at each time slot in [19]. On the other hand, in a network setting, the number of cognitive users K is usually large, and so all the results in Section 4 can be extended to the Markovian model. More specifically, one only needs to replace θ_i with θ_i^M, and all the proofs in Section 4 follow. Details of the proofs of these results can be found in [20].

5.2 Impact of Finite T

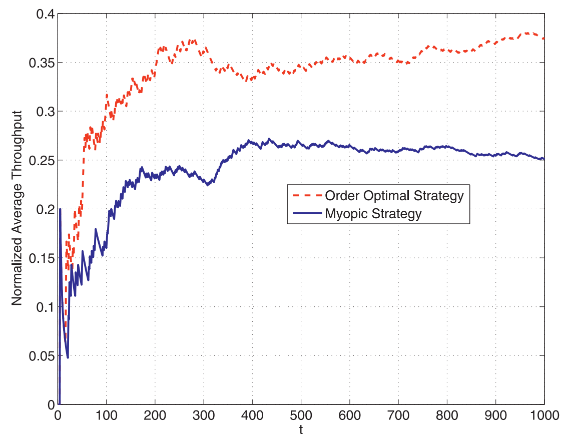


Fig. 3. Comparison of the order-optimal strategy in Rule 1 and the myopic strategy with finite T.

Here, we show some numerical results to consider the impact of finite T on the low-complexity rules proposed in the single-user and multiple-user cases. As mentioned in Section 3.2, the proposed low-complexity scheme is not only order optimal, but also offers a strict performance guarantee for any finite T. On the other hand, for the myopic strategy there is no convergence guarantee, i.e., there is a nonzero probability that the myopic strategy will stay with channels with smaller availability probabilities. Fig. 3 exhibits such a scenario for the single-user scenario. In this figure, we show the time average of the throughput obtained before t by using Rule 1. In generating this simulation result, we assume that the user can access one channel per time and that there are $N=10$ channels. In the figure, B is set to be 1. We randomly generate the parameter $\boldsymbol{\theta}$, and keep the value fixed for T slots. In this figure, $\boldsymbol{\theta}=[0.0211,\ 0.1461,\ 0.0981, 0.1780, 0.2539, 0.4742, 0.4480, 0.1691, 0.0431, 0.1206]$. Using the same parameters, we run the simulation 10,000 times. We find that 62 percent of the time, the myopic strategy will stay with a channel with a smaller availability probability. Fig. 4 shows the performance of

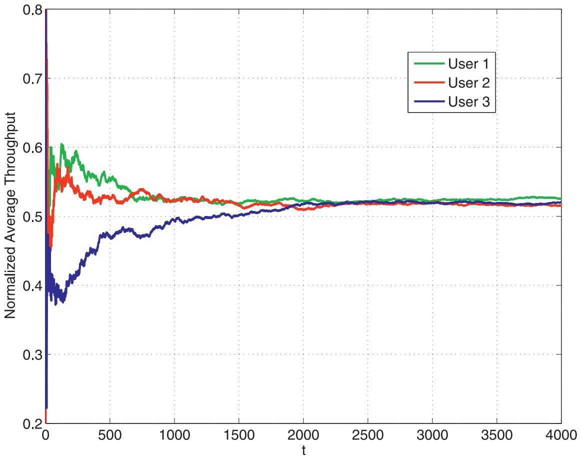


Fig. 4. The performance of each user with finite T in Rule 3.

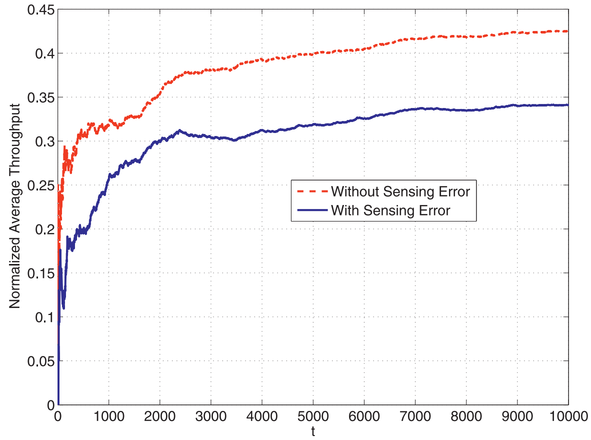


Fig. 5. Performance comparison of Rule 1 for the cases with sensing error and without sensing error.

Rule 3 in the multiple cognitive user case for finite T. In this simulation, we assume that there are $K = 3$ users and $N = 10$ channels with randomly generated parameters $\boldsymbol{\theta} = [0.9000, 0.3000, 0.4894, 0.2193, 0.4840, 0.6711, 0.3685, 0.4065, 0.2390, 0.8689]$. From this figure, we can see that the time average throughput of each user converges to the average value, and that these throughputs are close to each other for finite values of T.

5.3 Impact of Sensing Errors

Sensing errors are inevitable in any spectrum sensing situation. In practice, sensing algorithms have the following two categories of sensing errors [21]: missed detection and false alarm. Let η be the missed detection probability, which is the probability that a channel is sensed to be free while it is actually busy. If a missed detection happens, there will be a transmission collision. Also let ϵ be the false alarm probability, which is the probability that a channel is sensed to be busy while it is actually available. Suppose that the true availability probability of a channel is θ_i; then the sensing result $Z_i(j)$ (recall that $Z_i(j)$ is the random variable that equals 1 when channel i is sensed to be free in time slot j, and equals 0 otherwise) is a Bernoulli random variable with the parameter $(1-\theta_i)\eta + (1-\epsilon)\theta_i$. Thus, as long as we sample each channel with enough number of samples, $X_i(j)/Y_i(j)$ converges to $(1-\theta_i)\eta + (1-\epsilon)\theta_i$. Hence, the computationally simple rules still work with additional modification:

$$\hat{\theta}_i = \frac{X_i(j)/Y_i(j) - \eta}{1 - \epsilon - \eta}. \tag{32}$$

For any implementation, we can choose η according to the collision probability that the primary user can tolerate. Then, from the detection algorithms, we know the parameter ϵ. Fig. 5 shows the simulation result for Rule 1 for the parameters $\eta = 0.01$, $\epsilon = 0.3$, and

$$\boldsymbol{\theta} = [0.0193, 0.2113, 0.0368, 0.4656, 0.2159, 0.2251, 0.1312, 0.2975, 0.1609, 0.4347].$$

In this figure, the upper curve shows the results with no sensing errors, and the lower curve shows the results with sensing errors generated according to η and ϵ. We can see that with sensing errors, the performance still converges to the optimal value corresponding to $(1-\theta_{i^*})\eta + (1-\epsilon)\theta_{i^*}$.

5.4 Impact of CSMA/CA Overhead

In a practical system, when CSMA/CA protocol is used to resolve the contention among users that try to access the same channel, the procedure in each time slot is described as follows: If the channel is sensed free, a user randomly chooses an integer number (called the backoff timer) from its contention window, denoted by [0, CW] (here CW is called the contention window size). Then, the user counts down its backoff timer by one after each small duration (called a minislot). Once its backoff timer reaches zero, the user senses the channel again, and transmits in the time slot if the channel is sensed free. So in each time slot, the user with the smallest backoff timer will transmit. Two kinds of overhead exist in this procedure: backoff duration and possible collision (i.e., when two or more users choose the same smallest backoff timer).

As shown in Lemma 7, at the Nash equilibrium there are $\tau_i K$ users sensing channel i. Based on this, we can determine the optimal contention window size in channel i, which maximizes the effective throughput in the channel (taking the two kinds of overhead into consideration). Thus, we can obtain the ratio of the effective throughput in all channels to the ideal throughput (i.e., when the backoff duration is zero and there is no collision), for different total numbers of users (K). In the calculations, we set $N = 10$. The availability probabilities of the 10 channels are randomly chosen within (0, 1) to be $\boldsymbol{\theta} = [0.9501, 0.2311, 0.6068, 0.4860, 0.8913, 0.7621, 0.4565, 0.0185, 0.8214, 0.4447]$. Similarly to [22], the slot duration is 100 ms. The minislot duration is 20 μs, in accordance with the IEEE 802.11 Standard. It is observed that, when the total number of users (K) ranges through from 50, 100, 150, 200, 300, 400, and 500, the ratio of the effective throughput to the ideal throughput is still high, taking values of 0.98, 0.97, 0.96, 0.95, 0.93, 0.90, and 0.88, respectively. Hence, in this range of parameters, the impact of overhead is not significant.

6 Conclusions

This work has taken a first step in the design and analysis of cognitive medium access operating in uncertain environments, based on the classical bandit problem. In the single-user scenario, our formulation is inspired by the equivalence with the multiarmed bandit problem. This equivalence is used to highlight the trade-off between exploration and exploitation in cognitive channel selection. A linear complexity cognitive medium access algorithm, which is asymptotically optimal as $T \to \infty$, has been proposed. The multiuser setting has also been formulated, as a competitive bandit problem enabling the design of efficient and game theoretically fair medium access protocols. These ideas have also been extended to the multichannel scenario in which the cognitive user is capable of sensing and utilizing several channels simultaneously.

Appendix A

Proof of Lemma 2

Proof. The proof is an application of a theorem proved in [13]. More specifically, for a general bandit problem, let X be the random payoff obtained by choosing bandit i (not necessarily Bernoulli), and we let $h_{\theta_i}(x)$ be the pdf of X for a given θ_i.

Let μ_{θ_i} denote the average payoff of bandit i, i.e., $\mu_{\theta_i} = \int x h_{\theta_i}(x) dx$, and note that the Kullback-Leibler divergence between bandits i and l is given by

$$D(\theta_i \| \theta_l) = \int [\ln h_{\theta_i}(x) - \ln h_{\theta_l}(x)] h_{\theta_i}(x) dx. \tag{33}$$

Let $i^* = \arg\max_{i \in \mathcal{N}} \mu_i$, i.e., the index of the channel with the largest average payoff. It has been proved in [13, Theorem 1] that if

1. $0 < D(\theta_i \| \theta_l) < \infty$, for any $\theta_i \neq \theta_l$, and
2. $\forall \epsilon > 0$, and $\forall \mu_{\theta_l} > \mu_{\theta_i}$, there exists $\delta > 0$, so that we have $|D(\theta_i \| \theta_l) - D(\theta_i \| \theta')| < \epsilon$ whenever $\mu_{\theta_i} \leq \mu_{\theta'} \leq \mu_{\theta_i} + \delta$,

then for any consistent strategy Γ, we have

$$\liminf_{T \to \infty} \frac{L(\boldsymbol{\theta}; \Gamma)}{\ln T} \geq \sum_{i \in \mathcal{N} \setminus \{i^*\}} \frac{\mu_{i^*} - \mu_i}{D(\theta_i \| \theta_i^*)}. \tag{34}$$

In our cognitive radio channel selection problem, given $\boldsymbol{\theta}$, X is a random variable with pdf

$$h_{\theta_i}(x) = \theta_i \delta(x - B) + (1 - \theta_i)\delta(x);$$

hence $\mu_i = B\theta_i$, and

$$D(\theta_i \| \theta_l) = \theta_i \ln(\theta_i / \theta_l) + (1 - \theta_i) \ln((1 - \theta_i)/(1 - \theta_l)).$$

It is straightforward to verify that the technical conditions are satisfied. Thus, on substituting these parameters into (34), the proof is complete. □

Appendix B

Proof of Lemma 3

Proof. Let $i^* = \arg\max_{i \in \mathcal{N}} \theta_i$ and $i^{**} = \arg\max_{i \in \mathcal{N} \setminus \{i^*\}} \theta_i$; i.e., i^* is the best channel, and i^{**} is the second best channel. To avoid trivial conditions, without loss of generality we assume that $\theta_{i^*} \neq \theta_{i^{**}}$ and $\theta_{i^*} \neq 1$. We can upper bound the performance of the staying with the winner and switching from the loser rule by assuming that the cognitive user has the following extra knowledge.

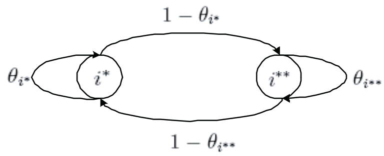


Fig. 6. A Markov process representation of the optimistic strategy Γ^*_{SW}.

1. In the first time slot, the cognitive user is able to choose i^* correctly.
2. Once i^* is sensed to be busy, the cognitive user somehow knows which channel is the second best, and switches to i^{**}.
3. Once i^{**} is sensed to be busy, the cognitive user is always able to switch back to i^*.

We denote this optimistic rule by Γ^*_{SW}. With any realistic switching rule Γ_{SW}, we have

$$L(\boldsymbol{\theta}; \Gamma_{SW}) \geq L(\boldsymbol{\theta}; \Gamma^*_{SW}).$$

Now with the optimistic rule Γ^*_{SW}, the system can be modeled as the following Markov process as shown in Fig. 6, in which we have two states: 1) sensing channel i^* and 2) sensing channel i^{**}. The transition probability matrix of this Markov chain is

$$P = \begin{bmatrix} \theta_{i^*}, & 1 - \theta_{i^*} \\ 1 - \theta_{i^{**}}, & \theta_{i^{**}} \end{bmatrix}. \tag{35}$$

The probability $P_{i^{**}}$ that the cognitive user will sense channel i^{**} can be obtained by solving the following equation:

$$P_{i^{**}} = (1 - \theta_{i^*})(1 - P_{i^{**}}) + \theta_{i^{**}} P_{i^{**}}, \tag{36}$$

from which we obtain

$$P_{i^{**}} = \frac{1 - \theta_{i^*}}{1 - \theta_{i^*} + 1 - \theta_{i^{**}}}. \tag{37}$$

Hence, in the nontrivial cases, we have

$$L(\boldsymbol{\theta}; \Gamma^*_{SW}) = B P_{i^{**}} (\theta_{i^*} - \theta_{i^{**}}) T, \tag{38}$$

implying that, for any switching rule, $L(\boldsymbol{\theta}; \Gamma_{SW}) \sim O(T)$. □

Appendix C

Proof of Lemma 4

Proof. We bound $Y_i(T)$ for $i \geq M + 1$, i.e., the channels that are not among the channels having the M largest values of θ. Note that $Y_i(T)$ is the total number of time slots in which the cognitive user has sensed channel i in a block with T time slots. We have

$$\begin{aligned} Y_i(T) &= 1 + \sum_{j = \lceil N/M \rceil + 1}^{T} I\{i \in \mathcal{M}(j)\} \\ &\leq m + \sum_{j = \lceil N/M \rceil + m}^{T} I\{i \in \mathcal{M}(j) \big| Y_i(j) \geq m\}, \end{aligned}$$

for any $m \geq 1$, where $I\{x|y\}$ is the conditional indicator function, which equals 1 if, conditioning on y, x is satisfied, and otherwise equals 0. Since $Y_i(j) \geq m$, it follows that $i \in \mathcal{M}(j)$ only if $\Lambda_i(j)$ is among the M largest indices. Hence, a necessary condition for $i \in \mathcal{M}(j)$ is $\Lambda_i(j) \geq \min\{\Lambda_l(j) : 1 \leq l \leq M\}$. Otherwise, if $\Lambda_i(j) < \min\{\Lambda_l(j) : 1 \leq l \leq M\}$, then the indices of these M channels are already larger than that of channel i, and channel i will not be selected. Thus,

$$\begin{aligned} & I\{i \in \mathcal{M}(j)|Y_i(j) \geq m\} \\ & \quad \leq I\{\Lambda_i(j) \geq \min\{\Lambda_l(j) : 1 \leq l \leq M\}|Y_i(j) \geq m\} \\ & \quad \leq \sum_{l=1}^{M} I\{\Lambda_i(j) \geq \Lambda_l(j)|Y_i(j) \geq m\}. \end{aligned} \tag{39}$$

Hence,

$$\begin{aligned} Y_i(T) & \leq m + \sum_{j=\lceil N/M\rceil+m}^{T} \sum_{l=1}^{M} I\{\Lambda_i(j) \geq \Lambda_l(j)|Y_i(j) \geq m\} \\ & \leq \sum_{l=1}^{M} \left\{ m + \sum_{j=\lceil N/M\rceil+m}^{T} I\{\Lambda_i(j) \geq \Lambda_l(j)|Y_i(j) \geq m\} \right\}. \end{aligned}$$

In order for $\Lambda_i(j) \geq \Lambda_l(j)$, one of the following three conditions must be satisfied:

$$\Lambda_l(j) \leq \theta_l, \; \Lambda_i(j) \geq \theta_i + 2\sqrt{\frac{2\ln j}{Y_i(j)}}, \quad \text{or} \quad \theta_l \leq \theta_i + 2\sqrt{\frac{2\ln j}{Y_i(j)}}.$$

One can easily check that, if none of these three conditions is satisfied, we will have $\Lambda_i(j) < \Lambda_l(j)$. In the following, we bound the probability of each event.

$$\begin{aligned} & \Pr\{\Lambda_l(j) \leq \theta_l|Y_i(j) \geq m\} \\ & \quad = \Pr\{\hat{\theta}_l + \sqrt{2\ln j/Y_l(j)} \leq \theta_l|Y_i(j) \geq m\} \\ & \quad \leq \Pr\{|\hat{\theta}_l - \theta_l| \geq \sqrt{2\ln j/Y_l(j)}|Y_i(j) \geq m\} \\ & \quad = \sum_{q=1}^{j} \Pr\{Y_l(j) = q|Y_i(j) \geq m\} \\ & \qquad \Pr\left\{ |\hat{\theta}_l - \theta_l| \geq \sqrt{\frac{2\ln j}{Y_l(j)}} \,\middle|\, Y_i(j) \geq m, Y_l(j) = q \right\} \\ & \quad \leq \sum_{q=1}^{j} \Pr\{|\hat{\theta}_l - \theta_l| \geq \sqrt{2\ln j/Y_l(j)}|Y_l(j) = q\} \\ & \quad \leq 2j \exp^{-4\ln j} = 2j^{-3}, \end{aligned} \tag{40}$$

where (40) follows from the following Chernoff-Hoeffding bound, which says that for n i.i.d Bernoulli random variables $X_j, j = 1, \ldots, n$ with mean $\bar{\theta}$, we have

$$\Pr\left\{ \left| \frac{\sum X_j}{n} - \bar{\theta} \right| \geq \epsilon \right\} \leq 2\exp^{-2n\epsilon^2}, \quad \text{for all} \quad \epsilon > 0. \tag{41}$$

To see this, we note that in our case, $X_l(j)$ is the sum of $Y_l(j)$ i.i.d Bernoulli random variables with parameter θ_l. On setting

$$n = Y_l(j), \quad \text{and} \quad \epsilon = \sqrt{2\ln j/Y_l(j)},$$

and also using the fact that

$$\hat{\theta}_l = \sum Z_l(j)/Y_l(j),$$

we have (40).

Similarly, we have

$$\begin{aligned} & \Pr\{\Lambda_i(j) \geq \theta_i + 2\sqrt{2\ln j/Y_l(j)}|Y_i(j) \geq m\} \\ & \quad = \Pr\{\hat{\theta}_i \geq \theta_i + \sqrt{2\ln j/Y_l(j)}|Y_i(j) \geq m\} \\ & \quad = \sum_{q=m}^{j} \Pr\{Y_i(j) = q|Y_i(j) \geq m\} \\ & \qquad \Pr\{\hat{\theta}_i \geq \theta_i + \sqrt{2\ln j/Y_l(j)}|Y_i(j) \geq m, Y_i(j) = q\} \\ & \quad \leq \sum_{q=1}^{j} \Pr\{\hat{\theta}_i \geq \theta_i + \sqrt{2\ln j/Y_l(j)}|Y_i(j) = q\} \\ & \quad \leq 2j \exp^{-4\ln j} = 2j^{-3}. \end{aligned} \tag{42}$$

At the same time, if we set $m = \lceil \frac{8\ln T}{(\theta_i - \theta_M)^2} \rceil$, then, for any $1 \leq l \leq M$, if $Y_i(j) \geq m$, we have

$$\begin{aligned} & \theta_i + 2\sqrt{2\ln j/Y_i(j)} \leq \theta_i + 2\sqrt{2\ln j/m} \\ & \quad \leq \theta_i + (\theta_M - \theta_i)\sqrt{\ln j/\ln T} < \theta_M \leq \theta_l. \end{aligned} \tag{43}$$

Hence, with this m,

$$\Pr\{\theta_l \leq \theta_i + 2\sqrt{2\ln j/Y_l(j)}|Y_i(j) \geq m\} = 0,$$

for each $1 \leq l \leq M$.

Thus,

$$\begin{aligned} & \Pr\{\Lambda_i(j) \geq \Lambda_l(j)|Y_i(j) \geq m\} \\ & \quad \leq \Pr\{\Lambda_l(j) \leq \theta_l|Y_i(j) \geq m\} \\ & \qquad + \Pr\{\Lambda_i(j) \geq \theta_i + 2\sqrt{2\ln j/Y_l(j)}|Y_i(j) \geq m\} \\ & \qquad + \Pr\{\theta_l \leq \theta_i + 2\sqrt{2\ln j/Y_l(j)}|Y_i(j) \geq m\} \\ & \quad \leq 4j^{-3}. \end{aligned} \tag{44}$$

Moreover,

$$\begin{aligned} & \mathbb{E}\{Y_i(T)\} \\ & \quad \leq \mathbb{E}\left\{ \sum_{l=1}^{M} \left\{ m + \sum_{j=\lceil N/M\rceil+m}^{T} I\{\Lambda_i(j) \geq \Lambda_l(j)|Y_i(j) \geq m\} \right\} \right\} \\ & \quad = \sum_{l=1}^{M} \left\{ \left\lceil \frac{8\ln T}{(\theta_i - \theta_M)^2} \right\rceil + \sum_{j=\lceil N/M\rceil+m}^{T} \right. \\ & \qquad \left. \mathbb{E}\left\{ I\left\{ \Lambda_i(j) \geq \Lambda_l(j) \middle| Y_i(j) \geq \left\lceil \frac{8\ln T}{(\theta_i - \theta_M)^2} \right\rceil \right\} \right\} \right\} \\ & \quad \leq M\left\{ \left\lceil \frac{8\ln T}{(\theta_i - \theta_M)^2} \right\rceil + \sum_{j=\lceil N/M\rceil+m}^{T} 4j^{-3} \right\} \\ & \quad \sim O(\ln T), \end{aligned} \tag{45}$$

since

$$\sum_{j=\lceil N/M\rceil+m}^{T} 4j^{-3} \leq \sum_{j=1}^{\infty} 4j^{-3}, \tag{46}$$

and $\sum_{j=1}^{\infty} j^{-3}$ exists.

Hence from (45), we have that, for any channel that is not among the best M channels, the average number of time slots for which this channel is selected is bounded by $O(\ln T)$. Thus, the loss is of order $O(\ln T)$.

On the other hand, it has been proved in [23] that for any consistent strategy,

$$\liminf_{T\to\infty} \frac{L(\boldsymbol{\theta};\Gamma)}{\ln T} \geq c_3, \tag{47}$$

with some constant c_3. This completes the proof. □

APPENDIX D
PROOF OF LEMMA 5

Proof. With a strategy $\mathbf{p}$, the probability that user k chooses channel i and, at the same time, there are l other users choosing channel i to sense is

$$p_i \binom{K-1}{l} p_i^l (1-p_i)^{K-1-l}. \tag{48}$$

Under this scenario, the average number of bits transmitted in one slot of each user is $B\theta_i/(l+1)$. Hence, the average throughput W_k of user k is

$$W_k = T\sum_{i=1}^{N} \frac{B\theta_i}{l+1} \sum_{l=0}^{K-1} p_i \binom{K-1}{l} p_i^l (1-p_i)^{K-1-l}. \tag{49}$$

Based on our symmetry assumption, we drop the subscript k and write the average throughput of each user as W, leading to

$$\begin{aligned} W &= BT\sum_{i=1}^{N} p_i\theta_i \sum_{l=0}^{K-1} \binom{K-1}{l} \frac{p_i^l(1-p_i)^{K-1-l}}{l+1} \\ &= BT\sum_{i=1}^{N} p_i\theta_i \sum_{l=0}^{K-1} \frac{(K-1)!}{l!(K-1-l)!} \frac{p_i^l(1-p_i)^{K-1-l}}{l+1} \\ &= BT\sum_{i=1}^{N} \frac{\theta_i}{K} \sum_{l=0}^{K-1} \binom{K}{l+1} p_i^{l+1}(1-p_i)^{K-1-l} \\ &= BT\sum_{i=1}^{N} \frac{\theta_i}{K} \left\{ \sum_{l'=0}^{K} \binom{K}{l'} p_i^{l'}(1-p_i)^{K-l'} - (1-p_i)^K \right\} \\ &= BT\sum_{i=1}^{N} \frac{\theta_i}{K} \{1-(1-p_i)^K\}. \end{aligned} \tag{50}$$

Now, we should solve the following optimization problem:

$$\begin{aligned} \max \quad & W = BT\sum_{i=1}^{N} \frac{\theta_i}{K}\{1-(1-p_i)^K\}, \\ \text{s.t.} \quad & \sum_{i=1}^{N} p_i = 1, \quad \text{and} \quad \mathbf{p} \geq \mathbf{0}. \end{aligned} \tag{51}$$

This optimization problem is equivalent to the following:

$$\begin{aligned} \min \quad & y = \sum_{i=1}^{N} \theta_i (1-p_i)^K, \\ \text{s.t.} \quad & \sum_{i=1}^{N} p_i = 1, \quad \text{and} \quad \mathbf{p} \geq \mathbf{0}. \end{aligned} \tag{52}$$

Since the mixed partials of y are all zero, and since

$$\frac{\partial^2 y}{\partial^2 p_i} = \theta_i K(K-1)(1-p_i)^{K-2} \geq 0,$$

for $0 \leq p_i \leq 1$, y is a convex function of $\mathbf{p}$ in the region of interest, i.e., $\mathbf{p} \in [0,1]^N$. Also, the constraints are the intersection of a convex set and a linear constraint. Therefore, our problem reduces to a convex optimization problem whose KKT conditions for optimality are

$$\begin{aligned} &\mathbf{p}^* \geq \mathbf{0}, \\ &\sum_{i=1}^{N} p_i^* = 1, \\ &p_i^*(\lambda^* - K\theta_i(1-p_i^*)^{K-1}) = 0, \quad \text{and} \\ &\lambda^* \geq K\theta_i(1-p_i^*)^{K-1}, \end{aligned} \tag{53}$$

where λ^* is a Lagrange multiplier.

It is easy to check that, if $K > 1$, then

$$p_i^* = \begin{cases} \left\{1 - \left(\frac{\lambda^*}{K\theta_i}\right)^{1/(K-1)}\right\}^+, & \text{for} \quad \theta_i > 0, \\ 0, & \text{for} \quad \theta_i = 0, \end{cases} \tag{54}$$

satisfies the KKT conditions, in which λ^* is the constant that satisfies $\sum p_i^* = 1$. □

APPENDIX E
PROOF OF LEMMA 6

Proof. Without loss of generality, we assume that $\theta_i \neq 0$, for $1 \leq i \leq Q$. At the moment, we assume that (we will show that this is true, if K is large enough) if $\theta_i \neq 0$, then

$$p_i^* = \left\{1 - \left(\frac{\lambda^*}{K\theta_i}\right)^{1/(K-1)}\right\}^+ = 1 - \left(\frac{\lambda^*}{K\theta_i}\right)^{1/(K-1)}. \tag{55}$$

Together with $\sum_{i=1}^N p_i^* = \sum_{i=1}^Q p_i^* = 1$, we have

$$(\lambda^*)^{1/(K-1)} = \frac{K^{1/(K-1)}(Q-1)}{\sum_{i=1}^{Q} \theta_i^{-1/(K-1)}} \tag{56}$$

and

$$p_i^* = 1 - \frac{(Q-1)\theta_i^{-1/(K-1)}}{\sum_{i=1}^{Q} \theta_i^{-1/(K-1)}}, \quad \text{for} \quad 1 \leq i \leq Q. \tag{57}$$

To satisfy the condition $\mathbf{p} \geq \mathbf{0}$, we need to show

$$\frac{(Q-1)\theta_i^{-1/(K-1)}}{\sum_{i=1}^{Q} \theta_i^{-1/(K-1)}} \leq 1, \tag{58}$$

for all i with $\theta_i > 0$.

With $i^* = \arg\max_{i\in\mathcal{N}}\theta_i$ and $l^* = \arg\min_{1\leq l\leq Q}\theta_l$, we have for all i

$$\frac{(Q-1)\theta_i^{-1/(K-1)}}{\sum_{i=1}^{Q} \theta_i^{-1/(K-1)}} \leq \frac{(Q-1)\theta_{l^*}^{-1/(K-1)}}{Q\theta_{i^*}^{-1/(K-1)}}. \tag{59}$$

For any $\vartheta \leq Q/(Q-1)$, if K is large enough, we have

$$(\theta_{i^*}/\theta_{l^*})^{\frac{1}{K-1}} \leq \vartheta \tag{60}$$

since

$$\lim_{K\to\infty} (\theta_{i^*}/\theta_{l^*})^{1/(K-1)} = 1.$$

Hence, for all $1 \leq i \leq Q$, we have

$$\frac{(Q-1)\theta_i^{-1/(K-1)}}{\sum_{i=1}^{Q} \theta_i^{-1/(K-1)}} \leq \frac{Q-1}{Q}\vartheta \leq 1. \tag{61}$$

Now, straightforward limit calculation shows that $p^* \to 1/Q$, as K increases, and

$$\lim_{K\to\infty} \frac{L}{\exp^{-c_1 K}} = \lim_{K\to\infty} \frac{BT\sum_{i=1}^{Q} \theta_i (1-p_i^*)^K}{\exp^{-c_1 K}} = BT\sum_{i=1}^{Q} \theta_i$$

with $c_1 = \ln Q/(Q-1)$. □

Appendix F
Proof of Lemma 8

Proof. $X_{k,i}$ is the sum of $Y_{k,i}$ i.i.d Bernoulli random variables with parameter θ_i. We use the following form of the Chernoff bound. Let X be the sum of n independent Bernoulli random variables with parameter $\bar{\theta}$, then

$$\Pr\{X \leq (1-\delta)n\bar{\theta}\} < \exp(-n\bar{\theta}\delta^2/2) \tag{62}$$

for any $\delta < 1$.

At time slot j, if we replace X with $X_{k,i}(j)$, n with $Y_{k,i}(j)$, $\bar{\theta}$ with θ_i and let $\delta = 1/2$, then we have

$$\Pr\{X_{k,i}(j) \leq \theta_i Y_{k,i}(j)/2\} < \exp(-Y_{k,i}(j)\theta_i/8). \tag{63}$$

Hence,

$$\begin{aligned}\Pr\{X_{k,i}(j)/Y_{k,i}(j) \geq \theta_i/2\} &\geq 1-\exp(-Y_{k,i}(j)\theta_i/8) \\ &\geq 1-\exp(-\theta_i/8),\end{aligned} \tag{64}$$

since after the initialization period, $Y_{k,i}(j) \geq 1$.

Note that $Y_{k,i}(T)$ is the total number of time slots in which user k has sensed channel i in each block with T time slots. We have

$$\begin{aligned}\mathbb{E}\{Y_{k,i}(T)\} &= \mathbb{E}\left\{\sum_{j=1}^{T} I\{S_k(j) = i\}\right\} \\ &= \sum_{j=1}^{T} \mathbb{E}\left\{\frac{X_{k,i}(j)/Y_{k,i}(j)}{\sum_{i\in\mathcal{N}} X_{k,i}(j)/Y_{k,i}(j)}\right\} \\ &\overset{(a)}{\geq} \sum_{j=1}^{T} \mathbb{E}\left\{\frac{X_{k,i}(j)/Y_{k,i}(j)}{N}\right\} \\ &\overset{(b)}{\geq} \sum_{j=1}^{T} \theta_i(1-\exp(-\theta_i/8))/(2N) \\ &= T\theta_i(1-\exp(-\theta_i/8))/(2N) = c_i T,\end{aligned}$$

where (a) follows from the fact that $X_{k,i}(j)/Y_{k,i}(j) \leq 1$, and (b) follows from (64).

The probability that $Y_{k,i}(T) \leq (1-\delta)\mathbb{E}\{Y_{k,i}(T)\}$ can also be bounded using the Chernoff bound since $Y_{k,i}(T)$ is also the sum of independent Bernoulli random variables. In particular, we have

$$\Pr\{Y_{k,i}(T) \leq (1-\delta)\mathbb{E}\{Y_{k,i}(T)\}\} \leq \exp\left(-\delta^2\mathbb{E}\{Y_{k,i}(T)\}/2\right).$$

On letting $\delta = 1/2$, we have

$$\Pr\{Y_{k,i}(T) \leq 1/2\mathbb{E}\{Y_{k,i}(T)\}\} \leq \exp^{-c_i T/8}. \tag{65}$$

Using the union bound, and the weak law of large numbers, $X_{k,i}(j)/Y_{k,i}(j)$ converges to θ_i in probability as T increases. The scheme becomes the same as the known $\boldsymbol{\theta}$ case, in which we know that the operating point is approximately at the Nash equilibrium, if K is sufficiently large. □

Acknowledgments

This work was supported by the Qatar National Research Fund under Grant NPRP-08-522-2-211 and an Alberta Ingenuity New Faculty Award from the Alberta Innovates-Technology Futures, Canada.

References

[1] J. Mitola, "Cognitive Radio: Making Software Radios More Personal," *IEEE Personal Comm.,* vol. 6, no. 4, pp. 13-18, Aug. 1999.

[2] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE J. Selected Areas in Comm.,* vol. 23, no. 2, pp. 201-220, Feb. 2005.

[3] Q. Zhao, S. Geirhofer, L. Tong, and B.M. Sadler, "Opportunistic Spectrum Access via Periodic Channel Sensing," *IEEE Trans. Signal Processing,* vol. 56, no. 2, pp. 785-796, Feb. 2008.

[4] S. Geirhofer, L. Tong, and B.M. Sadler, "Dynamic Spectrum Access in the Time Domain: Modeling and Exploiting White Space," *IEEE Comm. Magazine,* vol. 45, no. 5, pp. 66-72, May 2007.

[5] Z. Sahinoglu and S. Tekinay, "On Multimedia Networks: Self-Similar Traffic and Network Performance," *IEEE Comm. Magazine,* vol. 37, no. 1, pp. 48-52, Jan. 1999.

[6] D.A. Berry and B. Fristedt, *Bandit Problems: Sequential Allocation of Experiments.* Chapman and Hall, 1985.

[7] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction.* MIT Press, 1998.

[8] A. Motamedi and A. Bahai, "Dynamic Channel Selection for Spectrum Sharing in Unlicensed Bands," *European Trans. Telecomm. and Related Technologies,* submitted, 2007.

[9] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework," *IEEE J. Selected Areas in Comm.,* vol. 25, no. 3, pp. 589-600, Apr. 2007.

[10] Q. Zhao and B. Krishnamachari, "Structure and Optimality of Myopic Sensing for Opportunistic Spectrum Access," *Proc. IEEE Int'l Conf. Comm.,* pp. 6476-6481, June 2007.

[11] H. Jiang, L. Lai, R. Fan, and H.V. Poor, "Optimal Selection of Channel Sensing Order in Cognitive Radios," *IEEE Trans. Wireless Comm.,* vol. 8, no. 1, pp. 297-307, Jan. 2009.

[12] R. Fan and H. Jiang, "Channel Sensing-Order Setting in Cognitive Radio Networks: A Two-User Case," *IEEE Trans. Vehicular Technology,* vol. 58, no. 9, pp. 4997-5008, Nov. 2009.

[13] T.L. Lai and H. Robbins, "Asymptotically Efficient Adaptive Allocation Rules," *Advances in Applied Math.,* vol. 6, no. 1, pp. 4-22, 1985.

[14] P.R. Kumar, "A Survey of Some Results in Stochastic Adaptive Control," *SIAM J. Control and Optimization,* vol. 23, pp. 329-380, May 1985.

[15] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-Time Analysis of the Multiarmed Bandit Problem," *Machine Learning,* vol. 47, pp. 235-256, 2002.

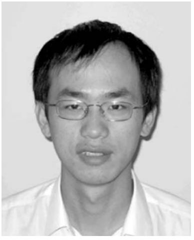
[16] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge Univ. Press, 2004.

[17] P. Whittle, "Restless Bandits: Activity Allocation in a Changing World," *J. Applied Probability,* vol. 25A, pp. 287-298, 1988.

[18] C.H. Papadimitriou and J.N. Tsitsiklis, "The Complexity of Optimal Queueing Network Control," *Math. of Operations Research,* vol. 24, no. 2, pp. 293-305, 1999.

[19] S.H.A. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of Myopic Sensing in Multichannel Opportunistic Access," *IEEE Trans. Information Theory,* vol. 55, pp. 4040-4050, Sept. 2009.

[20] L. Lai, H. Jiang, and H.V. Poor, "Medium Access in Cognitive Radio Networks: A Competitive Multi-Armed Bandit Framework," *Proc. Asilomar Conf. Signals, Systems and Computers,* Oct. 2008.
[21] H.V. Poor, *An Introduction to Signal Detection and Estimation.* Springer-Verlag, 1994.
[22] H. Kim and K.G. Shin, "Efficient Discovery of Spectrum Opportunities with MAC-Layer Sensing in Cognitive Radio Networks," *IEEE Trans. Mobile Computing,* vol. 7, no. 5, pp. 533-545, May 2008.
[23] V. Anantharam, P. Varaiya, and J. Walrand, "Asymptotically Efficient Allocation Rules for the Multiarmed Bandit Problem with Multiple Plays—Part I: I.I.D Rewards," *IEEE Trans. Automatic Control,* vol. 32, no. 1, pp. 968-976, Nov. 1987.



Lifeng Lai received the BE and ME degrees from Zhejiang University, Hangzhou, China, in 2001 and 2004, respectively, and the PhD degree from the The Ohio State University at Columbus in 2007. He was a postdoctoral research associate at Princeton University from 2007 to 2009. He is now an assistant professor at the University of Arkansas, Little Rock. He was a distinguished university fellow of the Ohio State University from 2004 to 2007. He coauthored a paper that received a Best Paper Award from the IEEE Global Communications Conference 2008. He is a member of the IEEE.

Hesham El Gamal received the BS and MS degrees in electrical engineering from Cairo University, Egypt, in 1993 and 1996, respectively, and the PhD degree in electrical and computer engineering from the University of Maryland at College Park in 1999. Since January 2001, he has been with the Electrical and Computer Engineering Department at the Ohio State University where he is now a professor. He held visiting appointments at the University of California, Los Angeles, Institut Eurecom, and Nile University. He is the founding director of the Wireless Intelligent Networks Center (WINC) at Nile University, Cairo, Egypt. He is a recipient of the HNS Annual Achievement Award (2000), the OSU College of Engineering Lumley Research Award (2003 and 2008), the OSU Electrical Engineering Department FARMER Young Faculty Development Fund (2003-2008), the OSU Stanley E. Harrison Award (2008), and the US National Science Foundation CAREER Award (2004). He holds 10 patents and has nine more pending applications. He is a fellow of the IEEE and served as an associate editor for the *IEEE Transactions on Communications* (2001-2005), as an associate editor for the *IEEE Transactions on Mobile Computing* (2003-2007), as a guest editor for the *IEEE Transactions on Information Theory* special issue on cooperative communications (2007), and a member of the SP4COM technical committee (2002-2005). He currently serves as the cochair of the Globecom 2008 communication theory symposium.

Hai Jiang received the BSc and MSc degrees in electronics engineering from Peking University, Beijing, China, in 1995 and 1998, respectively, and the PhD degree (with an Outstanding Achievement in Graduate Studies award) in electrical engineering from the University of Waterloo, Ontario, Canada, in 2006. Since July 2007, he has been an assistant professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. His research interests include radio resource management, cognitive radio networking, and cross-layer design for wireless multimedia communications. He is an associate editor for the *IEEE Transactions on Vehicular Technology*. He served as a cochair for the General Symposium at the International Wireless Communications and Mobile Computing Conference in 2007, the Communications and Networking Symposium at the Canadian Conference on Electrical and Computer Engineering in 2009, and the Wireless and Mobile Networking Symposium at the IEEE International Conference on Communications in 2010. He received an Alberta Ingenuity New Faculty Award in 2008 and a Best Paper Award from the IEEE Global Communications Conference (Globecom) in 2008. He is a member of the IEEE.

H. Vincent Poor received the PhD degree in electrical engineering and computer science from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton, where he is the Michael Henry Strater University professor of electrical engineering and dean of the School of Engineering and Applied Science. His research interests are in the areas of stochastic analysis, statistical signal processing and information theory, and their applications in wireless networks and related fields. Among his publications in these areas are the recent books *Quickest Detection* (Cambridge University Press, 2009) and *Information Theoretic Security* (Now Publishers, 2009). He is a member of the National Academy of Engineering, a fellow of the American Academy of Arts and Sciences, and an International fellow of the Royal Academy of Engineering (United Kingdom). He is also a fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as president of the IEEE Information Theory Society, and from 2004-2007, he served as the editor-in-chief of the *IEEE Transactions on Information Theory*. He was the recipient of the 2005 IEEE Education Medal. Recent recognition of his work includes the 2009 Edwin Howard Armstrong Achievement Award of the IEEE Communications Society, the 2010 IET Ambrose Fleming Medal for Achievement in Communications, and the 2011 IEEE Eric E. Sumner Award. He is a fellow of the IEEE.