

Almost Optimal Accessing of Nonstochastic Channels in Cognitive Radio Networks

Xiang-Yang Li^{*}, PanLong Yang[†], Yubo Yan[‡], LiZhao You[§], Shaojie Tang^{*}, QiuYuan Huang[¶]

^{*} Department of Computer Science, Illinois Institute of Technology, Chicago, IL

[†] Institute of Communications Engineering, PLA University of Science and Technology

[‡] Institute of Command Automation, PLA University of Science and Technology

[§] State Key Laboratory for Novel Software Technology, NanJing University

[¶] Department of Electrical and Computer Engineering, University of Florida

Abstract—We propose joint channel sensing, probing, and accessing schemes for secondary users in cognitive radio networks. Our method has time and space complexity $O(N \cdot k)$ for a network with N channels and k secondary users, while applying classic methods requires exponential time complexity. We prove that, even when channel states are selected by adversary (thus non-stochastic), it results in a total regret uniformly upper bounded by $\Theta(\sqrt{TN \log N})$, w.h.p, for communication lasts for T timeslots. Our protocol can be implemented in a distributed manner due to the nonstochastic channel assumption. Our experiments show that our schemes achieve almost optimal throughput compared with an optimal static strategy, and perform significantly better than previous methods in many settings.

I. INTRODUCTION

We study dynamic spectrum access in cognitive radio (CR) networks with primary users (PU) and secondary users (SU) where the qualities (e.g., the data rate supported) of different channels often are time dependent, and may differ significantly across all channels. We investigate a joint sensing/probing/accessing mechanism that will result in (almost) optimal expected throughput, compared with the optimal fixed-channel strategy. In our approach, a CR not only senses the busy/idle status of the channel, but also probes the instantaneous quality of the channel using short predefined probing packets. Based on historical observations, a CR will decide whether to transmit data over current channel (called *exploitation*) or to continue sensing/probing some other channels (called *exploration*). Our goal in this work is to design *efficient* (using polynomial space and time) and *effective* (with a small *weak regret*, called *regret* hereafter, i.e., the gap between the overall throughput by the strategy using some fixed best channels when knowing all channel quality for the system lifetime, and the throughput achieved by our scheme) spectrum usage scheme for multiple users.

Designing better schemes for channel usage has been extensively studied [5], [10], [21]. A number of schemes were proposed to exploit the dynamic channel qualities, e.g., [18] assumed a homogeneous distribution of the channel quality. Recently, several results, e.g., [1], [9], [16], were proposed by exploring the parallels between the cognitive medium access and the multi-armed bandit (MAB) problem. These methods can theoretically guarantee almost optimum *regret* under various assumptions. However, these protocols require

either exponential space, or exponential time, or both when multiple users and/or channels present.

We propose a joint sensing, probing, and accessing scheme striking for maximizing the expected throughput. We first consider the system with heterogeneous stochastic channel availability and qualities. For a network of single SU, we derive a throughput optimal strategy based on optimal stopping rule. Optimal centralized method is then proposed for a network with multiple users. We then consider networks with nonstochastic channel qualities. We first propose a centralized scheme, called ϵ -SPA, which runs in polynomial space and polynomial time $O(N \cdot k)$ for a network of N channels and k SUs. Our method is inspired by results in [4] and will adaptively select channels for communication. We theoretically prove that the total regret is $O(\sqrt{TN \ln N})$, when the system lifetime is T . Since our protocol assumes nonstochastic channels, each SU can run our ϵ -SPA scheme individually in a distributed manner, which will also approximate the maximum throughput in each user's view. We also discuss how to achieve better throughput by setting parameters. We conduct extensive experiments using USRP and sensor networks to study the performance of our schemes. Our results show that the total throughput achieved by our SPA scheme is indeed close to the optimum fixed strategy.

We present the system model in Section II, and our optimal sensing/probing/accessing (SPA) strategy in Section III. We report our evaluation results in Section IV, review related work in Section V, and conclude the work in Section VI.

II. SYSTEM MODEL AND PROBLEM

A. Network and System Model

Consider a set $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N\}$ of N channels. When a channel \mathbf{c}_i is used, it is termed as *busy*, otherwise, it is termed as *idle*. Let $p_I(i)$ (resp. $p_B(i)$) be the idle (resp. busy) probability of channel \mathbf{c}_i . There is a set $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k\}$ of k secondary users (SU) equipped with a cognitive radio (CR). Before accessing a channel \mathbf{c}_i , the \mathbf{s}_u needs to sense the channel to determine whether it is idle, and if it is idle then probe \mathbf{c}_i to determine the estimated data rate, $R_{u,i,t}$, that can be achieved in next timeslot. Here we assume that the data rate $R_{u,i,t} \in [0, M]$ could be *arbitrary*, while previous results often assumed it is drawn from a given (maybe unknown) distribution. Here constant M is the maximum data rate for

all channels, $R_{u,i,t} = 0$ denoting that the channel \mathbf{c}_i is busy at time t . We do not assume $R_{u,i,t} = R_{j,i,t}$ for $u \neq j$. Denote $R_{u,i,t}$ as $R_{i,t}$ if \mathbf{s}_u is clear from context.

We assume that each radio on CR needs time t_s for sensing the status of a channel and time t_p for probing its quality. The actual time depends on the technology and device: t_s is about 10ms and t_p is from 10ms to 133ms [2]. Let $t_{sp} = t_s + t_p$. For a pair of transmitter/receiver, we assume that the clock is synchronized, and a common random number generator is shared for determining the sequence of channels to be sensed/probed/accessed. When a channel is idle, transmitter/receiver can only access it for at most t_a time continuously so it can detect the return of a PU. In practice, the value t_a is typically 2s. Observe that, although the channel quality often has a coherence time $t_c > t_a$, to avoid the possible collisions with PUs and other CRs, we assume that a CR will *not* recall (reuse) a previously probed channel. We will extend our mechanisms to the case when recall is allowed.

B. Problem Formulation

The sequential channel sensing/probing/accessing (SPA) problem is to determine when to sensing/probing/accessing which channel, without knowing future channel states, for a pair of cognitive radio nodes so as to improve the throughput. One difference between MAB [7] and the sequential channel SPA problem is that, for MAB problem, at every timeslot t , we will receive a reward and we repeat this for T timeslots; while for SPA problem, at a timeslot t , we will *not* have any gain if we only sense/probe a channel: the actual gain only happens when a CR transmits data using some channel. To address this challenge, we call a block of continuous timeslots spent for sensing/probing a chosen channel as a *round*. The immediate following timeslots spent for transmitting data over a chosen channel are *not* counted in a round. Instead, we will treat the total data transmitted in these transmission timeslots as the gain of the previous sensing/probing round. The strategy for sensing/probing/accessing is then composed of many sequential rounds of sensing/probing.

A sensing/probing/accessing strategy $\chi = \langle \theta(t), \sigma(t) \rangle$ by the pair of CRs will decide its action. At every round t , $\theta(t)$ is the channel index selected, and $\sigma(t) \in \{0, 1\}$ denotes whether the CR decides to transmit data using the probed channel $\mathbf{c}_{\theta(t)}$ after this round immediately: $\sigma(t) = 1$ if they will transmit using channel $\mathbf{c}_{\theta(t)}$ and $\sigma(t) = 0$ otherwise. For a strategy $\chi(t) = \langle \theta(t), \sigma(t) \rangle$ by a SU \mathbf{s}_j , at round t , its gain is $g_\chi(t) = \sigma(t) \cdot R_{j,\theta(t),t} \cdot Z_t \cdot (1 - P_{loss}(t))$, where Z_t denotes the time of the actual transmission and P_{loss} is the probability that the transmission will be destroyed by the return of PU or some other SU during the Z_t time-frame. We assume that $Z_t = t_a$.

Let n be the number of sensing/probing rounds executed during the system's lifetime T , which should satisfy the condition: $n \cdot t_{sp} + \sum_{t=1}^n \sigma(t) \cdot Z_t \leq T$. The first part is time spent for sensing/probing and the second part is the time for data transmissions. Then our objective is to design a strategy that maximizes the expected throughput over the system lifetime T , i.e., $\max_\chi \frac{\sum_{t=1}^T g_\chi(t)}{T}$. A channel i with the highest total throughput, $\sum_{t=1}^T R_{i,t}$, is called the *best channel*

(or *optimal static strategy*). The *regret* ρ after T rounds of an online strategy χ is defined as the difference between the reward sum associated with an optimal static strategy and the sum of the data rates in T rounds by χ . A strategy whose average regret per round $\rho/T \rightarrow 0$ with probability 1 when $T \rightarrow \infty$ is a *zero-regret strategy*. Our objective is design a strategy χ with small regret.

III. ALMOST OPTIMAL SPA SCHEMES

We first propose two methods when centralized decision is possible. The first method assumes channels' statistics follow some unknown distributions and the number of possible rates is limited. The strategy is *learning then transmitting*. A CR will first collect the channel statistics by sensing/probing all channels for some rounds, then transmit data based on the collected information. The second method works for a more general case when the channel states are nonstochastic. The strategy is sort of *learning when transmitting*. The main idea is as follows. We guess an optimal strategy at the beginning. With certain probability, we execute the strategy we guessed; otherwise, we try some new strategies. Based on the feedback, i.e., the throughput, we adjust our guess dynamically. For both methods, we first consider the simple case where the network has only one SU. Then we extend our methods to cases with k users. We also discuss how to optimize the parameters used in our methods.

A. OSP SPA for Stochastic Channels

Assume that each channel \mathbf{c}_i has rates drawn from a set $\mathcal{R} = \{R_1, R_2, \dots, R_r\}$ with $R_j < R_{j+1}$ for $j \in [1, r-1]$, following some distributions with mean μ_i . After learning rate distributions, we can then use the optimal stopping rule (OSP) [8] for the single channel. Intuitively, the *static* optimal strategy should always sense/probe/access a fixed channel that will maximize the expected throughput $p_I(i) \cdot \mu_i$. Unfortunately, this may not be always optimal as we will show later.

When we choose a *fixed* channel \mathbf{c}_i to sense/probe, the *rate of return*, i.e., the achieved data rate for next transmission, is $\frac{R_{i,h} t_a (1 - P_{loss})}{h \cdot t_{sp} + t_a}$ if we stop at the h -th probing after previous transmission. To maximize mean rate of return, it is easy to show that the optimal strategy is as follows (1) keep sensing/probing the channel \mathbf{c}_i until the probed data rate $R_{i,t}$ at t -th probing after previous transmission is at least $\frac{\lambda_i^*}{1 - P_{loss}}$; (2) transmit using data rate $R_{i,t}$. Here λ_i^* is the solution of

$$\sum_{X \geq \frac{\lambda_i^*}{1 - P_{loss}}} \mathbf{Pr}_i(X) (X \cdot (1 - P_{loss}) - \lambda_i^*) = \lambda_i^* t_{sp} / t_a \quad (1)$$

where $\mathbf{Pr}_i(X)$ is the probability that the channel \mathbf{c}_i will have rate X (by integrating the availability $p_I(i)$ of the channel \mathbf{c}_i). Notice that λ_i^* is different from μ_i .

We then design the best strategy if we have heterogeneous channels. We focus on one secondary user \mathbf{s}_u . Let λ_i^* denote the expected data rate achieved if we stick with sensing/probing channel \mathbf{c}_i using the optimal stopping rule (see Eq. (1)). Without loss of generality, we assume $\lambda_i^* \geq \lambda_{i+1}^*$, for $i \in [1, N-1]$. Let Y denote any possible mixed strategy, i.e.,

at any time slot t , we can randomly sense/probe a channel $\theta(t)$. Assume that with probability q_i it will sense/probe channel \mathbf{c}_i , where $\sum_{i=1}^N q_i = 1$. Let variable Y_t denote the data rate that can be observed at time t and λ_Y^* be the expected rate of the return by this strategy Y . Then $\Pr(Y_t = R_j) = \sum_{i=1}^N q_i \cdot \Pr_i(R_j)$. Notice that it is not necessary that $\lambda_Y^* = \sum_{i=1}^N q_i \lambda_i^*$. Let $\lambda_i^*(u)$ be the expected average data rate a SU \mathbf{s}_u will achieve when it keeps sensing/probing and accessing the channel i . We show that this mixed strategy is no better than sticking with the best single channel. Proofs are omitted due to space limitation.

Theorem 1: For any mixed strategy Y by a SU \mathbf{s}_u ,

$$\lambda_1^*(u) \geq \lambda_Y^*(u) \geq \lambda_N^*(u) \quad (2)$$

For multiple secondary users, we define a weighted bipartite graph H over two sets of vertices SUs \mathcal{S} and channels \mathcal{C} , and the weight of an edge $(\mathbf{s}_u, \mathbf{c}_i)$ is $\lambda_i^*(u)$. Let Π be a maximum weighted matching in graph H . Then,

Theorem 2: To maximize overall throughput, the best strategy for a user \mathbf{s}_u is to use the optimal stopping method on channel i if $(\mathbf{s}_u, \mathbf{c}_i)$ is in the maximum weighted matching Π .
B. SPA for Nonstochastic Channels

Previous channel accessing methods often assume i.i.d. distributions of channel data rates (e.g., [1], [9], [16], [18]), or Markovian [19]. Here we design a method whose expected throughput is almost optimal when the channel qualities may not have stochastic distributions.

1) *Our Protocol Overview:* The challenge is that at any time instance, we do not know whether our current strategy is good enough or not. Inspired by [4], we use parameter γ to adjust the fraction of exploration and exploitation. With probability $1 - \gamma$, we will *exploit* and just adapt the strategy used in previous round. Here γ is a relatively small parameter depending on the number of total rounds n (which is mainly decided by the lifetime T and our strategy). And exploitation will promise an almost optimal performance when previously used strategy is almost optimal. With a probability γ we will *explore* new channels, to estimate the rate of each channel with same probability $\frac{1}{N}$. The process of exploring is also important, which eventually improves our strategy to the optimal solution. In all our methods, for technical convenience, all actual data rates are scaled with maximum value 1. The virtual rates used in our method are used to compute the channel sensing/probing probability for next round.

2) *Protocol for Single Secondary User:* Algorithm 1 summarizes our protocol inspired by methods in [4], when there is only one secondary user. We present this method for the completeness of presentation and discussion of an efficient method for multiple users.

To study the performance of our algorithm, we analyze its *regret*. For this purpose, we define the accumulated data rate, denoted as $\mathbb{R}_{i,n}$, of channel \mathbf{c}_i , and the accumulated virtual data rate, denoted as $\mathbb{R}'_{i,n}$, as follows

$$\mathbb{R}_{i,n} = \sum_{t=1}^n R_{i,t} \quad \text{and} \quad \mathbb{R}'_{i,n} = \sum_{t=1}^n R'_{i,t} \quad (5)$$

Algorithm 1 ϵ -SPA Scheme for Single User

Parameters: real number $\beta > 0$, $0 < \eta, \gamma < 1/2$.

Initialization: Set $w_{i,0} = 1$ for all $1 \leq i \leq N$, and $W = N$. Divide all data rates by the maximum possible data rate M .

- 1: At t th round, randomly select a channel $\theta(t) = c_i$ according to the following distribution $p_{i,t} \forall i \in [1, N]$:

$$p_{i,t} = (1 - \gamma) \frac{w_{i,t-1}}{W_{t-1}} + \frac{\gamma}{N} \quad (3)$$

- 2: Sense and probe the channel $\theta(t)$, get the scaled data rate (i.e., in the range $[0, 1]$), denoted as $R_{\theta(t),t}$, of channel $\theta(t)$ at time t . Calculate virtual rates $R'_{i,t}, \forall i$,

$$R'_{i,t} = \begin{cases} \frac{R_{i,t} + \beta}{\beta} & \text{if } \theta(t) = c_i \\ \frac{p_{i,t}}{p_{i,t}} & \text{otherwise.} \end{cases} \quad (4)$$

- 3: Update $w_{i,t} = w_{i,t-1} e^{\eta R'_{i,t}}$, and $W_t = \sum_{i=1}^N w_{i,t}$.
 - 4: Access the channel $\theta(t)$ with probability ϵ , i.e., set $\sigma(t) = 1$ with probability ϵ .
-

Lemma 3: For any $\delta \in (0, 1)$, $\beta \in [0, 1]$ and $i \in [1, N]$, we have $\Pr\left(\mathbb{R}_{i,n} > \mathbb{R}'_{i,n} + \frac{1}{\beta} \ln \frac{N}{\delta}\right) \geq \frac{\delta}{N}$

Define $\hat{\mathbb{R}}_n(\epsilon)$ as the expected total rates that can be achieved by an ϵ -SPA scheme over n rounds. Theorem 4 bounds the regret, $\max_{1 \leq i \leq N} \mathbb{R}_{i,n} - \hat{\mathbb{R}}_n(1)$ of Algorithm 1 when $\epsilon = 1$.

Theorem 4: For any $\delta \in (0, 1)$, when $\beta = \sqrt{\frac{\ln \frac{N}{\delta}}{Nn}}$, $\gamma = 2\eta N$, $\eta = \sqrt{\frac{\ln N}{4nN}}$ and $n \geq \max\{\frac{\ln \frac{N}{\delta}}{N}, 4N \ln N\}$, we have $\Pr\left(\max_{1 \leq i \leq N} \mathbb{R}_{i,n} - \hat{\mathbb{R}}_n(1) \leq 6\sqrt{nN \ln N}\right) \geq 1 - \delta$.

According to our ϵ -SPA scheme, CR will transmit ϵn times in expectation during n rounds. It is easy to show $E[\hat{\mathbb{R}}_n(\epsilon)] = \epsilon E[\hat{\mathbb{R}}_n(1)]$, which implies

$$E[\hat{\mathbb{R}}_n(\epsilon)] \geq \epsilon \max_{1 \leq i \leq N} E[\mathbb{R}_{i,n}] - 6\epsilon \sqrt{nN \ln N} \quad (6)$$

Let R_{max} be the largest expected data rate among all channels. We have $\max_{1 \leq i \leq N} E[\mathbb{R}_{i,n}] = n \cdot R_{max}$. Assume $t_a = \alpha t_{sp}$ where constant $\alpha \gg 1$. Then we have

Theorem 5: The expected throughput of ϵ -SPA is at least

$$\frac{\hat{\mathbb{R}}_n(\epsilon) t_a}{T} \geq \frac{R_{max} - 6\sqrt{\frac{N \ln N}{n}}}{\frac{1}{\alpha} + 1} = \frac{R_{max} - 6\sqrt{\frac{(1+\alpha\epsilon)t_{sp} N \ln N}{T}}}{1 + \frac{1}{\alpha\epsilon}}$$

with probability $1 - \delta$, where $T = nt_{sp} + \epsilon nt_a$.

When T is sufficiently large, the expected throughput achieved is at least $\frac{R_{max}}{1 + \frac{1}{\alpha\epsilon}}$, which is maximized when $\epsilon = 1$. Clearly, the expected throughput that can be achieved is no more than $\frac{R_{max} t_a}{t_{sp} + t_a} = \frac{R_{max}}{1 + \frac{1}{\alpha}}$, because each transmission takes at least $t_{sp} + t_a$ time while the expected data rate is no more than R_{max} . Thus, we have

Theorem 6: When T is sufficiently large, our ϵ -SPA scheme is almost optimal.

3) *Multiple Secondary Users:* Here we consider the case when there are k secondary users with a centralized decision. For our SPA scheme, its main idea is still to *explore/exploit* channels. The difference is that our new method needs to select

k channels in each round. Therefore, there are totally $\binom{N}{k}$ different strategies when $k \leq N$ (here we view each channel as an arm in the MAB problem, and each SU as a player). When $k \geq N$, each channel should be sensed/probed/accessed. Then we will view each SU as an arm, and each channel as a player, thus, there are $\binom{k}{N}$ different strategies when $k \geq N$. The rest of discussions focuses on the case $k \leq N$. All our results can be converted easily when $k \geq N$.

Let χ_j denote the j th strategy which includes a set of k channels, where $1 \leq j \leq \binom{N}{k}$. Here we still use $\theta(t)$ to denote the set of channels we choose at time t . A simple method to address multiple users is to modify Algorithm 1 with new settings: each strategy is a combination of k channels, and each strategy χ is associated with a weight $w_{\chi,t}$ which is recursively updated as

$$w_{\chi,t} = w_{\chi,t-1} \prod_{i \in \mathcal{X}} e^{\eta R'_{i,t}}; \quad \text{and} \quad W_t = \sum_{\chi} w_{\chi,t}$$

Then at time t , a strategy χ is chosen with probability

$$p_{\chi,t} = (1 - \gamma) \frac{w_{\chi,t-1}}{W_{t-1}} + \gamma \frac{\sum_{S \in \mathcal{C}} I((\chi \cap S) \neq \emptyset)}{|\mathcal{C}|}. \quad (7)$$

Here the set $\mathcal{C} = \{\{\mathbf{c}_1, \dots, \mathbf{c}_k\}, \{\mathbf{c}_{k+1}, \dots, \mathbf{c}_{2k}\}, \dots\}$ is called a covering set of \mathcal{C} . Notice \mathcal{C} includes strategies such that each channel appears in at least one of the strategies. So the size of a covering set is $\lceil N/k \rceil$. This simple implementation will have the same regret bound as our new method (to be discussed), but it will have time and space complexity $O(N^k)$, exponential in number of users, which could be very expensive if we have a large number of users.

To reduce the complexity, we propose a novel approach that utilizes the internal structure of our method. We will not choose a strategy from the strategy set (*i.e.*, any combination of k channels) directly. In our new method (Algorithm 2), we make decision on each channel one by one. Assume each channel \mathbf{c}_i is still associated with a weight $w_{i,t}$ at time t . And the weight of a strategy χ_j is defined as the product of weights of all channels in that strategy. Let $S(p, q, k)$ denote the set of all strategies that will choose exactly k channels from subset $\{\mathbf{c}_p, \mathbf{c}_{p+1}, \dots, \mathbf{c}_q\}$. Define the weight of $S(p, q, k)$ at time t as $W_t(p, q, k) = \sum_{\chi_j \in S(p, q, k)} \prod_{\mathbf{c}_i \in \chi_j} w_{i,t}$. For example, the total weight of all strategies choosing k channels from N channels is $W_t(1, N, k)$. Here the probability that we choose channel \mathbf{c}_1 in our strategy at time t is $\frac{w_{1,t} W_t(2, N, k-1)}{W_t(1, N, k)}$, which is the total weight of strategies choosing channel \mathbf{c}_1 over the total weight of all strategies. Similarly, the probability that channel \mathbf{c}_1 is not chosen at time t is $\frac{W_t(2, N, k)}{W_t(1, N, k)}$.

For channel \mathbf{c}_2 , if channel \mathbf{c}_1 is chosen, the probability that channel \mathbf{c}_2 is in the strategy is $\frac{w_{1,t} w_{2,t} W_t(3, N, k-2)}{w_{1,t} W_t(2, N, k-1)} = \frac{w_{2,t} W_t(3, N, k-2)}{W_t(2, N, k-1)}$; if channel \mathbf{c}_1 is not chosen, the probability that channel \mathbf{c}_2 is in the strategy is $\frac{w_{2,t} W_t(3, N, k-1)}{W_t(2, N, k)}$. Thus, for channel \mathbf{c}_i , if k' channels has been chosen among $\mathbf{c}_1, \dots, \mathbf{c}_{i-1}$, the probability that \mathbf{c}_i is in the strategy is $\frac{w_{i,t} W_t(i+1, N, k-k'-1)}{W_t(i, N, k-k')}$.

For current timeslot, repeat the previous steps for all channels. We can show that, for this new method,

Algorithm 2 Coordinated ϵ -SPA Scheme for k Users

Parameters: real number $\beta > 0$, $0 < \eta, \gamma < 1$.

Initialization: Set $w_{i,0} = 1 \forall i \in [1, N]$; $W_0(N, N, 1) = 1$; $W_0(N, N, k') = 0 \forall k' \geq 2$; $W_0(i, N, 1) = N - i + 1 \forall i \geq 1$.

1: At t th round, initialize $\theta(t) = \emptyset$, $k' = k$, and update $W_t(i, N, k')$, $\forall i \in [1, N]$ and $\forall k' \in [1, k]$ as follows.

$$W_t(i, N, k') = w_{i,t} W_t(i+1, N, k'-1) + W_t(i+1, N, k'-1)$$

2: **for** $i = 1$ to N **do**

3: With prob. $\frac{w_{i,t} W_t(i+1, N, k'-1)}{W_t(i+1, N, k')}$, choose channel \mathbf{c}_i ,

$$\theta(t) = \theta(t) \cup \mathbf{c}_i \quad \text{and} \quad k' = k' - 1$$

4: **end for**

5: Compute $p_{i,t} = \gamma \frac{\sum_{S \in \mathcal{C}} I((\chi \cap S) \neq \emptyset)}{|\mathcal{C}|} + (1 - \gamma) \sum_{k'=0}^k \frac{W_{t-1}(1, i-1, k') w_{i,t-1} W_{t-1}(i+1, N, k-k'-1)}{W_{t-1}(1, N, k)}$.

6: Sense and probe all channels in $\theta(t)$, get the scaled data rates. Calculate virtual rates $R'_{i,t}$, $\forall i \in [1, N]$:

$$R'_{i,t} = \begin{cases} \frac{R_{i,t} + \beta}{p_{i,t}} & \text{if } \mathbf{c}_i \in \theta(t) \\ \frac{\beta}{p_{i,t}} & \text{otherwise.} \end{cases}$$

7: Update the weights, $\forall i$, $w_{i,t} = w_{i,t-1} e^{\eta R'_{i,t}}$.

8: Access all channels in $\theta(t)$ with probability ϵ , *i.e.*, set $\sigma(t) = 1$ with probability ϵ .

Lemma 7: Exactly k channels will be chosen. The time and space complexity at each timeslot is $O(k \cdot N)$ and the probability that a strategy χ will be chosen is $p_{\chi,t}$.

For regret analysis, similarly, we define

$$\begin{cases} \mathbb{R}_{\chi_j, n} = \sum_{t=1}^n R_{\chi_j, t} = \sum_{t=1}^n \sum_{\mathbf{c}_i \in \chi_j} R_{i,t} \\ \mathbb{R}'_{\chi_j, n} = \sum_{t=1}^n R'_{\chi_j, t} = \sum_{t=1}^n \sum_{\mathbf{c}_i \in \chi_j} R'_{i,t} \end{cases} \quad (8)$$

Lemma 3 still holds and we have following theorem.

Theorem 8: For $\delta \in (0, 1)$, $\beta = \sqrt{\frac{\ln \frac{kN}{\delta}}{Nn}}$, $\gamma = 2\eta N$, $\eta = \sqrt{\frac{\ln N}{4nkN}}$ and $n \geq \max\{\frac{k \ln \frac{N}{\delta}}{N}, 4 \frac{N}{k} \ln N\}$, we have

$$\Pr \left(\max_{\chi_j} \mathbb{R}_{\chi_j, n} - \hat{\mathbb{R}}_n(1) \leq 6k \sqrt{nN \ln N} \right) \geq 1 - \delta \quad (9)$$

Let R_{max}^k be the largest total expected data rates of k channels among all channels. Similar to Theorem 6, we have

Theorem 9: When T is sufficiently large, our ϵ -SPA scheme for k users is almost optimal.

C. Further Discussions on SPA

1) *Exploit Channel Coherence:* Here we consider the affect of channel coherence time. Within the coherence time t_c we assume the data rate does not change with high probability. Recalling a previously sensed/probed channel is possible when t_c is long enough. Assume that $t_c \geq m(t_{sp} + \epsilon t_a) + t_a$ for some integer $m \geq 1$. Then we can recall m previously sensed/probed channels whose data rates are assumed to remain same.

Our method with recall is similar to our ϵ -SPA scheme. The difference is as follows. At each round t , we randomly select a channel $\theta(t)$, sense and probe it. At same time, we

“virtually” sense and probe the last $m - 1$ channels we just sensed and probed, *i.e.*, channels $\theta(t - 1), \dots, \theta(t - m + 1)$. Here “virtually” means we don’t really sense and probe them. We update the weights just as if we sensed and probed them. Thus, in each round, we sense and probe m channels, and choose the best one with probability ϵ . Then,

Theorem 10: When $\beta = \sqrt{\frac{\ln \frac{N}{\delta}}{Nnm}}$, $\gamma = 2\eta N$, $\eta = \sqrt{\frac{\ln N}{4nmN}}$

and $nm \geq \max\{\frac{\ln \frac{N}{\delta}}{N}, 4N \ln N\}$, the regret of our algorithm 1-SPA with m channel recall satisfies

$$\Pr \left(\max_{1 \leq i \leq N} \mathbb{R}_{i,n} - \hat{\mathbb{R}}_n(1) \leq 6\sqrt{\frac{nN \ln N}{m}} \right) \geq 1 - \delta.$$

Proof: The proof is similar to that of Theorem 4. We can consider each round as a m virtual sub-rounds, where the $(jm + k)$ th sub-round, for $1 \leq j \leq n$, $0 \leq k \leq m - 1$, is mapped to the $(j + m - k)$ th round in the problem without recall. Then the access probability of each sub-round is $1/m$ for our 1-SPA scheme and we have $n \cdot m$ sub-rounds totally. By replacing ϵ with $\frac{1}{m}$, and n with nm in Eq. (6), we have

$$\Pr \left(E[\hat{\mathbb{R}}_{nm}(1)] \geq \frac{1}{m} \max_{1 \leq i \leq N} E[\mathbb{R}_{i,nm}] - 6\sqrt{\frac{nN \ln N}{m}} \right) \geq 1 - \delta.$$

The theorem then follows from $E[\hat{\mathbb{R}}_n(1)] = E[\hat{\mathbb{R}}_{nm}(1)]$, and $\max_{1 \leq i \leq N} E[\mathbb{R}_{i,n}] = \frac{1}{m} \max_{1 \leq i \leq N} E[\mathbb{R}_{i,nm}]$. ■

Similarly with a proper parameter setting we have

$$\Pr \left(E[\hat{\mathbb{R}}_n(\epsilon)] \geq \epsilon \max_{1 \leq i \leq N} E[\mathbb{R}_{i,n}] - 6\epsilon \sqrt{\frac{nN \ln N}{m}} \right) \geq 1 - \delta$$

for our ϵ -SPA scheme. Comparing with Eq. (6), allowing recall improves the convergence speed of our method.

2) *Impact of sensing time:* Notice that we don’t consider the false alarm probability of sensing in previous analysis. Since we consider energy detector for channel sensing, the false alarm probability [14] is approximated by $P_{fa}(t_s) = Q((\frac{\epsilon_0}{\sigma_u^2} - 1)\sqrt{t_s f_s})$, where $\frac{\epsilon_0}{\sigma_u^2}$ is the decision threshold for sensing, f_s is the channel bandwidth, and $Q(\cdot)$ is the Q -function for the tail probability of the standard normal distribution. Consider the false alarm probability, we have

Lemma 11: With probability $1 - \delta$, the expected throughput of ϵ -SPA scheme (Algorithm 1) is at least

$$\left(R_{max} - 6\sqrt{\frac{(1 + \alpha\epsilon_0)t_{sp}N \ln N}{(1 - P_{fa})T}} \right) / \left(1 + \frac{1}{\alpha\epsilon_0} \right)$$

The proof is similar to that of Theorem 10. For each round, we sense and probe channel successfully $1 - P_{fa}$ in expectation. Replacing m with $1 - P_{fa}$, we get the above result. Here P_{fa} is a function of t_s , and $\alpha = \frac{t_a}{t_s + t_p}$. Treating t_s as a variable, we can compute the optimal t_s which maximizes the expected throughput by numerical analysis.

3) *Impact of probing time and others:* The step of probing is not necessary in our problem. The reason why we need to probe the channel is that we want to make sure the data rate is good enough. This is important when the qualities of the channels are not good. On the other hand, when the qualities of the channels are good enough, a sense/access scheme may achieve better throughput since there is no probing overhead. Our ϵ -SPA scheme also can be extended to a simplified ϵ -SA scheme without probing steps. In ϵ -SA scheme, we can only get the observation on the data rate after each successful transmission

and ACK. In other word, in n rounds, ϵn data rates will be observed in expectation by the ϵ -SA scheme. Replacing m with ϵ in Theorem 10, we can show that the expected throughput of ϵ -SA scheme is $\left(R_{max} - 6\sqrt{\frac{(1 + \alpha'\epsilon)t_s N \ln N}{\epsilon T}} \right) / \left(1 + \frac{1}{\alpha'\epsilon} \right)$, where $\alpha' = \frac{t_a}{t_s}$. Let t_p^* be the probing time which satisfies $(R_{max} - 6\sqrt{\frac{(1 + \alpha'\epsilon)t_s N \ln N}{\epsilon T}}) / (1 + \frac{1}{\alpha'\epsilon}) = (R_{max} - 6\sqrt{\frac{(1 + \alpha\epsilon)(t_s + t_p^*)N \ln N}{T}}) / (1 + \frac{1}{\alpha\epsilon})$. When $t_p \leq t_p^*$, we will use ϵ -SPA; when $t_p \geq t_p^*$, we will use ϵ -SA.

The transmission may be destroyed by the return of PUs. Thus with the knowledge of p_I we can optimize t_a to maximize the expected throughput. By numerical analysis, we can also find ϵ that maximizes the expected throughput.

D. Decentralized Protocol With Multiple Users

It is challenging to design optimal decentralized protocol without using a common control channel (CCC). Since the energy-detection cannot differentiate spectrum usage of PUs and SUs, the view of each SU is also affected by other SUs. The channels quality and availability are thus dynamic and nonstochastic. Notice that the protocol ϵ -SPA developed previously applies to a more general observation model as long as SUs have the common set of k best channels and each of these channels has the same mean across players.

Our method is to let each user in the network run the ϵ -SPA method based on their own view. Eventually, each user will almost maximize its own throughput when others do not change their strategies. We expect to prove that the regret of each user is upper-bounded in this nonstochastic setting. This could be proved by applying an approach similar to the TDFS scheme [16]. We leave the detailed analysis as a future work.

IV. PERFORMANCE STUDY BY EXPERIMENTS

A. Testbed Implementation and Results

We build a cognitive radio system which is composed of 20 sensor nodes and 6 USRP devices and conduct extensive experiment studies on the performance of our system. The maximum transmission speed for each sensor node is $250kbps$. Using this software programmable platform by USRP, we can adaptively select the channel and transmission power. We use these USRPs as cognitive nodes and use the other sensor nodes as primary users. More importantly, we can also monitor and diagnose the wireless network using USRPs. There are 16 channels (from channel 11 to 26) available for the primary user network, and in each channel, the transmission bandwidth of the primary users is 2MHz, while the bandwidth of USRPs is 500kHz. We use the software suit built upon GNU radio, and add our SPA model into the model. There are two additional modules for running the SPA algorithm efficiently.

1) *Overall Design, Workflow, and Parameters:* We assume that all the secondary users will agree upon a common control channel (CCC). In our testing, we use channel 26 as the CCC. If the sender wants to initiate a communication, it will send a packet INIT_PKT to the receiver using CCC and then start the SPA process. The INIT_PKT also contains a seed that will be used to generate the sequence of random channel numbers

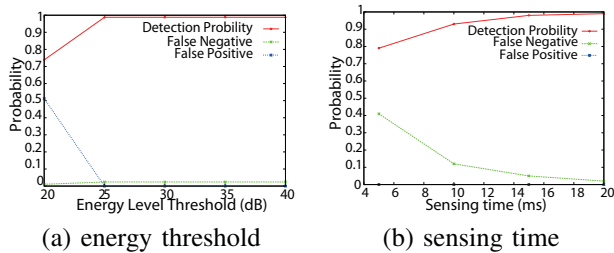


Fig. 1. Energy and time threshold for channel sensing.

used by SPA. The receiver will wait for the communication request from the sender. It will start the SPA process after it received the packet INIT_PKT.

Channel Sensing: We use the energy threshold method to detect whether a given channel is busy or idle. We revise this method (*e.g.*, performing signal smooth and filtering) such that it works well under the unique characteristics of sensor networks. The channel is considered to be busy if the energy level of the signal is above a certain threshold. Fig. 1(a) shows the different success probability using different energy threshold for detecting the busy/idle status of a channel in our testing.

For this experiment, using $25dB$ may be the best choice. The success probability also depends on the duration we examine a channel. From Fig. 1 (b), we find that when we exam a channel for about $15ms$, the success probability is already good enough. We choose $200ms$ in experiments for better accuracy.

Channel Probing: If the sender determines that the channel is idle using channel sensing method described previously, it then starts probing the channel quality and uses a timer, **PT**, for probing time. Timer **PT** starts with the maximum value *Probing_time*. The sender will send Probe_num of probing packets when **PT** is reduced to $Probing_time - switch_delay$. Here we use a delay, *switch_delay*, to avoid the negative impact caused by the processing delay of USRP. The receiver could send back the ACKs for the probing packets. Such feedback messages can be collected for the SNR and PRR (packet reception ratio) information. Before the expiration of the timer **PT**, when **PT** is reduced to a value $update_delay + switch_delay$, the sender updates the channel observations based on the probing results. Here *update_delay* is the delay for the SPA algorithm to update its internal parameters. This could be negligible for USRP node, but not for sensor nodes when our algorithm runs in sensor nodes. They decide whether to access some channel or continue sensing/probing other channels.

For the receiver, it also starts a timer **PT** with starting value *Probing_time*, when it knows that the sender/receiver starts the probing phase. It will send an ACK packet to the sender when **PT** is reduced to time $prob_RT + update_delay + switch_delay$. Here *prob_RT* is the round-trip delay for the probing packet. In our experiments, we set *prob_RT* to be a reasonable value that can also cover the errors of clock synchronization between the sender and the receiver,

TABLE I
PARAMETER VALUES FOR IMPLEMENTATIONS

Parameters	Values	Parameters	Values
Sensing_time	200ms	Probing_time	800ms
Accessing_time	2000ms	Running_time	100s to 600s
switch_delay	50ms	proc_delay	3ms
update_delay	50ms	Packet_interval	10ms
probe_num	50	probe_RT	200ms

the clock shifting of the sender and receiver. A proper choice of this parameter can let the sender receive the ACK packet in time. It also updates its own SPA internal parameters. Then it will decide to either access some channel or continue sensing/probing other channels based on SPA method using the common random number generator as the sender.

If the sender or the receiver determines the channel is busy, they will start another round of sensing/probing after delay of *probing_time*.

Parameters: Table I summarizes some parameters used for our experiments. These parameters are selected according to the performance of the working USRP system, and to improve the stability of the system.

Channel Accessing: During the channel accessing period, the sender will send data packets to the receiver using the maximum data rate under the current channel condition. Transmission of two consecutive packets are separated by a delay *proc_delay*. This delay is introduced to avoid the packet-drops by the receiver when the receiver has a processing speed lower than that of the sender.

Coding and modulation effects: We made extensive tests on multiple coding and modulation combinations over different SNRs. The interferences are from the external and internal. The external interferences are generated by the sensor nodes, where packet transmissions occurred periodically around the USRP nodes. The internal “interferences” are generated by the power control ability on the USRP mother board. With different levels of the output gain at the transmitter side, the SNR value changes accordingly. We can use different combination tests for the optimal transmission rate. As shown in Fig. 2 (a), we adjust the transmission power to different levels and achieve the mapping table of SNR and transmission power. We then conduct extensive experimental study to investigate the SNR-PER relationship. Fig. 2(c) illustrates the relationship between SNR and the PER. In our experiments, we have 8 different combinations of modulation and coding rate: (BPSK, 1/2), (BPSK, 3/4), (QPSK, 1/2), (QPSK, 3/4), (16-QAM, 1/2), (16-QAM, 3/4), (64-QAM, 1/2), (64-QAM, 3/4). The corresponding data rates (units kbps) are 150, 225, 300, 450, 600, 900, 1200, and 1350 respectively. We implement this based on rawOFDM (see <http://people.csail.mit.edu/szym/rawofdm/README.html>).

We applied two different approaches to build heterogeneous channel qualities: (1) adjust the transmission power of the transmitter (We use RFX2400 daughter board in the experiment, which does not allow us to change output gain, therefore we adjust the amplitude of transmission signal to change power) (2) using a separate wireless sensor network that will

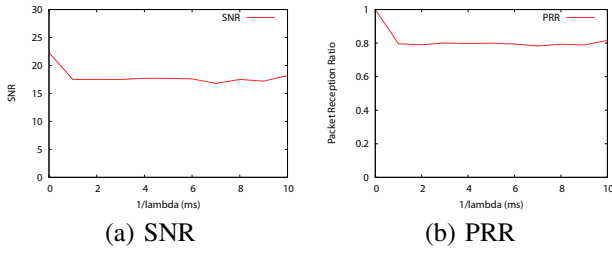


Fig. 3. Impact on SNR and PRR by arrival rates.

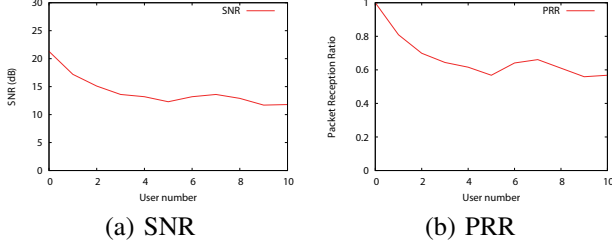


Fig. 4. Impact on SNR and PRR by user numbers.

transmit the collected data where the data arrival follows a Poisson distribution with arrival rate λ_0 . Fig. 2 shows the impact of different amplifications on SNR and PER. Fig. 3 shows the impact of different data arrival rate on the SNR and PRR. Fig. 4 shows the impact of different number of users on the SNR and PRR.

2) *Experimental Results:* In our experiment, we use 6 USRPs (numbered 1 to 6) to form 3 pairs, denoted as (1,2), (3,4), (5,6), of sender/receiver as secondary users. We first measure the energy level on different channels when the SPA is applied (this is also used for idle/busy detection). Fig. 5 plots the energy level at different time instances for a link.

Impact of Synchronization: Due to different processing abilities and asynchronous clocks by the sender and the receiver, they may have different views on the current system status, which could greatly impact the throughput of the system and the delay of accessing the channels. In our testbed implementation, we introduced several delay mechanisms by the sender and receiver, such as *switch_delay*, *proc_delay*, and *update_delay* to address this notoriously challenging issue. Unfortunately, these delay mechanisms will reduce the ideal achievable maximum throughput and increase the ideal minimum accessing-delay (time duration between two consecutive successful transmissions) to some extent. Accessing-delay is called delay hereafter. To study the impact of time synchronization on the throughput and delay, we conduct two separate experiments (1) the first experiment will connect each pair of the communicating USRPs to a computer, which we assume that time synchronization error is negligible; (2) the second experiment will let each pair of the communicating USRPs to connect to two separate computers, which we assume that the synchronization error is not negligible. Fig. 6 shows our experimental results of the impact of synchronization error on the throughput and delay. The throughput drops significantly and the delay increases significantly after 300s when synchronization errors exist. On the other hand,

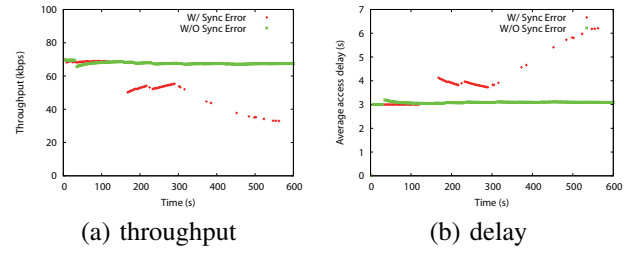


Fig. 6. Impact of clock synchronization error on throughput and delay.

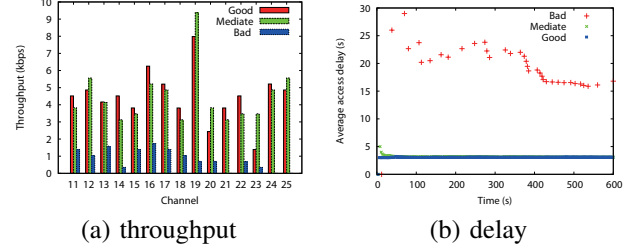


Fig. 7. Impact of amplification-controlled channel qualities on throughput and delay.

the throughput and delay remain stable when synchronization error is negligible.

Impact of different channel qualities: We then study the impact of different channel qualities on the performance. Recall that two different approaches are used in controlling the channel qualities: adjusting the signal amplitude, or the number of primary users already using some channels.

For the first approach, we apply three different groups of amplifications, denoted as *bad*, *mediate*, and *good*. The amplifications of 15 channels are as follows: (1) Bad: all channels use amplification 500, (2) Mediate: [500, 500, 500, 500, 500, 1000, 1000, 1000, 1000, 1000, 2000, 2000, 2000, 2000, 2000]; (3) Good: all channels use amplification 2000. Fig. 7 shows our experimental results of the impact of different channel qualities on the throughput and delay. There is a significant throughput and delay dropoff for the case “Bad”, compared with cases “Good” and “Mediate”.

For the second approach, we will adjust the number of primary users (sensor nodes in our experiment) in each of the 15 channels 11, 12, ..., 25. Note that channel 26 is reserved as CCC. Fig. 8 shows the experimental results of the number of existing primary users on the throughput and delay. Here the number of primary users for each channel $\in [11, 25]$ is set as follows: [3,3,3,3,3,1,1,1,1,0,0,0,0]. We found that SUs can still achieve good throughput and delay when there is at most one PU, while the throughput drops significantly in channels with 3 PUs.

Adaptivity of SPA system: We then study whether the SPA system can adjust its strategy when the channel qualities change dramatically after a certain time duration, such that the historical channel observation will have negative impact on the performance. We conduct the following experiment. In the first 300s, the amplification of each channel is as follows: [500,500,500, 500, 500, 1000, 1000, 1000, 1000, 1000, 2000, 2000, 2000, 2000, 2000]. After the system is

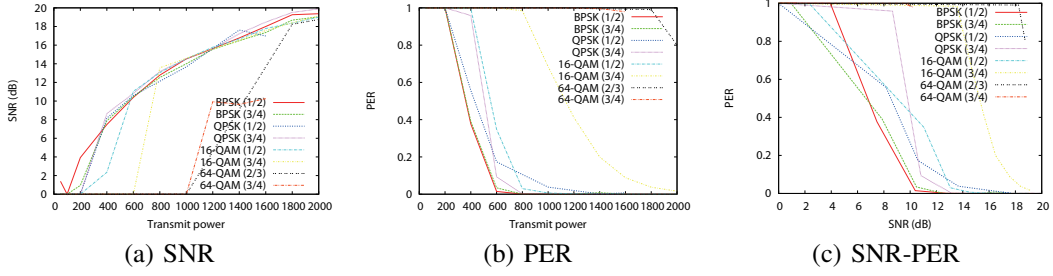


Fig. 2. Impact on SNR and PER by different amplifications, and the relationship between SNR and PER.

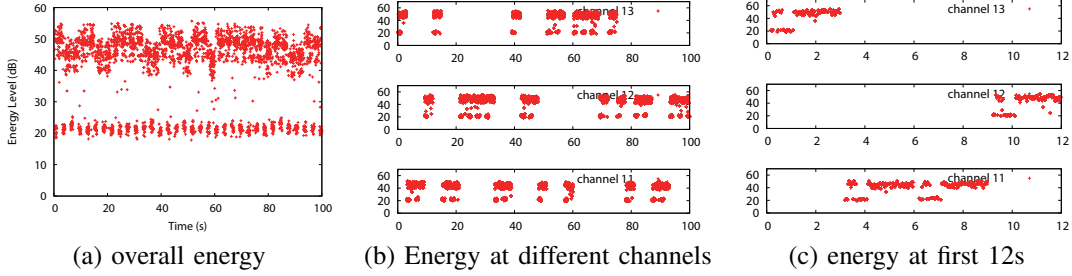


Fig. 5. Energy level when SPA is used to access channels idled by PUs.

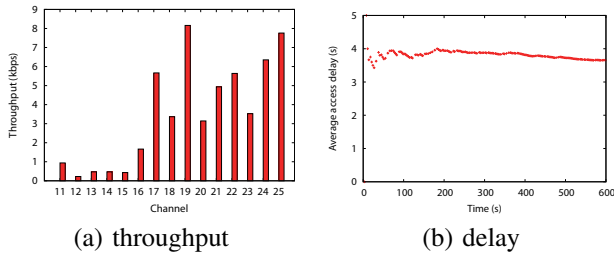


Fig. 8. Impact of the number of primary users on throughput and delay.

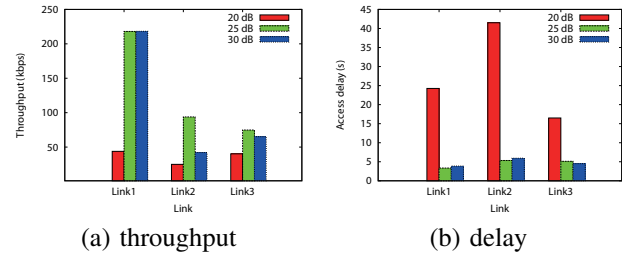


Fig. 10. Impact of different sensing thresholds.

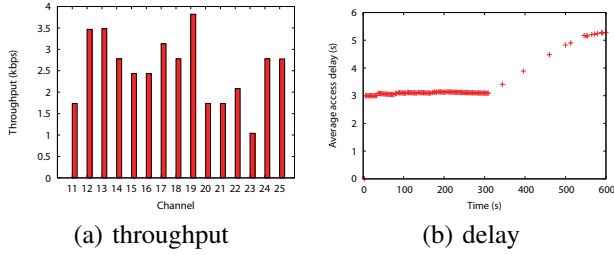


Fig. 9. Adaptivity of SPA scheme.

run for 300s, we change the amplification of each channel as follows: [2000, 2000, 2000, 2000, 2000, 1000, 1000, 1000, 1000, 1000, 500, 500, 500, 500, 500]. In other words, good channels will become bad and bad channels will become good. Fig. 9 demonstrates that our SPA scheme can quickly adapt to the new environment.

Impact of channel status sensing threshold: We then study the impact of different channel status sensing thresholds on the final achievable throughput and delay. Recall that all previous experimental results assume that the energy threshold for channel status sensing is set as 25dB. Fig. 10 shows that our choice of energy-detection thresholds is indeed the best for all links.

V. RELATED WORK

There is a rich body of results for allocating spectrum channels. Li *et al.* [12], [13] designed efficient and truthful mechanisms for various dynamic spectrum assignment problems. Zhou *et al.* [26] proposed a truthful and efficient dynamic spectrum auction system to serve many small players. Several results [20], [27] designed truthful double spectrum auctions with provable performance. All these results are based on offline models. For the online models, Xu *et al.* [22]–[25] studied online spectrum allocation and truthful mechanisms when secondary users could bid arbitrarily.

Many results have been developed for dynamic spectrum access in cognitive radio networks. Huang *et al.* [10] presented a threshold-based sensing-transmission structure that is optimal under a technical constraint to maximize the SU's utility. Xu and Liu [21] proposed an optimal transmitting, sensing, and sleeping structure. However, all these works ignored different data rates across all channels and over a time period. Recently several results were presented by using the connections between channel access and the multi-armed bandits problem. Shu and Krunz [18] proposed a throughput-efficient sensing/probing/access scheme with sensing errors. The difference with our work is that they assume stochastic

homogeneous channels while we consider nonstochastic channels. Gai *et al.* [9] used a combinatorial MAB formulation to address the multi-user channel allocation. Their method is centralized and assumes i.i.d. stochastic channels. Anandkumar *et al.* [1] designed distributed policy for learning and allocation and it achieves logarithmic growth of regret. It also assumes i.i.d. stochastic channels and cooperation among users in distributed implementation. Liu *et al.* [15], [16] presented a novel distributed learning in MAB problem with multiple players when the reward is i.i.d. from some distribution. The time and space requirement is exponential in the number of users. Proutiere *et al.* [17] presented a decentralized channel access protocol using learning approach. Recently, Tekin and Liu [19] modeled each channel as a restless Markov chain. They presented an algorithm using a sample-mean based index policy, and showed that under mild conditions this algorithm achieves logarithmic regret uniformly over time.

Our problem is related to classical MAB problem [7]. When the awards are i.i.d. according to an unknown law with unknown expectation, several protocols [11] can achieve the optimum logarithmic regret asymptotically. When the awards of actions are chosen by adversary, logarithmic regret cannot be achieved. For example, when the results of *all* N possible actions at each round are known, the regret per round can be bounded by $O(\sqrt{\ln N/n})$ [3]. When only the result of the action that the decision maker performs is known, the best possible regret per round is bounded by $\Theta(\sqrt{N \ln(Nn/\delta)/n})$ with probability at least $1-\delta$ [4]. Cesa-Bianchi *et al.* [6] shows that the regret per round is $\Theta(\sqrt{\ln N/m})$ where m queries are allowed during n rounds.

VI. CONCLUSIONS

We proposed efficient channel accessing methods for optimizing the throughput achieved by secondary users. Our methods have negligible regret even the channel data rates and channel availabilities are given by an adversary. Our methods can be extended to deal with several settings not specifically discussed here. For example, when the idle probabilities of all channels are low, sometimes it is better for the secondary user not to always sense/probe/access channels. To reflect such choices, we can introduce a third action, *idle*, by a secondary user. Some interesting future work is to study the Nash equilibriums of the system when multiple users exist and design distributed protocol with optimal regret.

ACKNOWLEDGMENT

The research of authors is partially supported by NSF CNS-0832120, NSF CNS-1035894, National Basic Research Program of China (973 Program) under grant No. 2010CB328101, 2010CB334707, 2011CB302705, National Natural Science Foundation of China under Grant No. 61170216, Tsinghua National Laboratory for Information Science and Technology (TNList), program for Zhejiang Provincial Key Innovative Research Team, and program for Zhejiang Provincial Overseas High-Level Talents (One-hundred Talents Program).

REFERENCES

- [1] ANANDKUMAR, A., MICHAEL, N., AND TANG, A. Opportunistic Spectrum Access with Multiple Users: Learning under Competition. In *IEEE INFOCOM* (2010).
- [2] ARBAUGH, W. Improving the latency of the probe phase during 802.11 handoff. manuscript, 2009
- [3] AUER, P., CESA-BIANCHI, N., FREUND, Y., AND SCHAPIRE, R. E. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *IEEE FOCS* (1995), p. 322.
- [4] AUER, P., CESA-BIANCHI, N., FREUND, Y., AND SCHAPIRE, R. E. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* 32, 1 (2003), 48–77.
- [5] BUDDHIKOT, M.M. Understanding dynamic spectrum access: Models, taxonomy and challenges. In *IEEE DySPAN*, p. 649-663, 2007
- [6] CESA-BIANCHI, N., LUGOSI, G., AND STOLTZ, G. Minimizing regret with label efficient prediction. *Information Theory, IEEE Transactions on* 51, 6 (june 2005), 2152 – 2162.
- [7] EVEN-DAR, E., MANNOR, S., AND MANSOUR, Y. PAC Bounds for Multi-armed Bandit and Markov Decision Processes. In *COLT* (2002), p. 255.
- [8] FERGUSON, T. S. Optimal stopping and applications. available on line at <http://www.math.ucla.edu/~tom/Stopping/Contents.html>.
- [9] GAI, Y., KRISHNAMACHARI, B., AND JAIN, R. Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation. In *IEEE DySPAN* (2010).
- [10] HUANG, S., LIU, X., AND DING, Z. Optimal sensing-transmission structure for dynamic spectrum access. In *IEEE INFOCOM* (2009).
- [11] LAI, T., AND ROBBINS, H. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics* 6, 1 (1985), 4–22.
- [12] LI, X.-Y., XU, P., TANG, S., AND CHU, X. Spectrum bidding in wireless networks and related. In *COCOON* (2008), pp. 558–567.
- [13] XU, P., LI, X.-Y., TANG, S.-J., AND ZHAO, J. Efficient and Strategyproof Spectrum Allocations in Multi-Channel Wireless Networks, *IEEE Transactions on Computers*, 2009.
- [14] LIANG, Y.-C., ZENG, Y., PEH, E., AND HOANG, A. T. Sensing-throughput tradeoff for cognitive radio networks. *IEEE ICC*, 2007, pp. 5330 –5335.
- [15] LIU, K., AND ZHAO, Q. Decentralized multi-armed bandit with multiple distributed players. In *IEEE Information Theory and Applications Workshop (ITA)* (2010), pp. 1–10.
- [16] LIU, K., AND ZHAO, Q. Distributed Learning in Multi-Armed Bandit with Multiple Players. *Signal Processing, IEEE Transactions on*, 99 (2010), 1.
- [17] PROUTIERE, A., AND KASBEKAR, G. Decentralized opportunistic medium access in multi-channel wireless systems: A learning approach. In *Allerton* (2010).
- [18] SHU, T., AND KRUNZ, M. Throughput-efficient sequential channel sensing and probing in cognitive radio networks under sensing errors. In *ACM MobiCom* (2009), pp. 37–48.
- [19] TEKIN, C., AND LIU, M. Online learning in opportunistic spectrum access: a restless bandit approach, in *IEEE INFOCOM*, 2011.
- [20] WANG, S.-G., XU, P., XU, X., TANG, S.-J., LI, X.-Y., AND LIU, X. TODA: Truthful online double auction for spectrum allocation in wireless networks. *IEEE DySpan* (2010).
- [21] XU, D., AND LIU, X. Opportunistic spectrum access in cognitive radio networks: when to turn off the spectrum sensors. In *4th International Conference on Wireless Internet* (2008).
- [22] XU, P., AND LI, X.-Y. Online market driven spectrum scheduling and auction. In *ACM CoRoNet* (2009), pp. 49–54.
- [23] XU, P., AND LI, X.-Y. SOFA: Strategyproof online frequency allocation for multihop wireless networks. In *ISAAC* (2009), pp. 311–320.
- [24] XU, P., AND LI, X.-Y. TOFU: Semi-Truthful Online Frequency Allocation Mechanism for Wireless Networks, *IEEE/ACM Transactions on Networking*, 2011.
- [25] XU, P., WANG, S.-G., AND LI, X.-Y. SALSA: Strategyproof Online Spectrum Admissions for Wireless Networks, *IEEE Transactions on Computers*, 2010
- [26] ZHOU, X., GANDHI, S., SURI, S., AND ZHENG, H. eBay in the Sky: strategy-proof wireless spectrum auctions. In *ACM MobiCom* (2008), pp. 2–13.
- [27] ZHOU, X., AND ZHENG, H. TRUST: A general framework for truthful double spectrum auctions. In *IEEE INFOCOM* (2009).