Minimum Required Learning and Impact of Information Feedback Delay for Cognitive Users

Yi Su, *Student Member, IEEE*, and Mihaela van der Schaar, *Senior Member, IEEE*

***Abstract*—This paper studies the value of learning for cognitive transceivers in dynamic wireless networks. We quantify the utility improvement that can be obtained by a wideband user that learns the stationary usage pattern of the spectrum occupied by narrowband users and, based on this learned information, adapts its transmission. Specifically, we investigate the basic tradeoff between the learning duration and the achievable performance in stationary environments. We apply optimization and large-deviation theory to analytically derive an upper bound of the minimum required learning duration, given the user's tolerable performance loss and outage probability. Furthermore, since learning techniques require the information feedback of the spectrum usage pattern between the transceivers, we investigate how a cognitive user can further improve its performance by taking into account its feedback delay. The impact of inaccurate delay estimation on the achievable performance is also quantified.**

***Index Terms*—Cognitive users, feedback delay, learning, wireless networks.**

I. Introduction

A PROMISING way of improving the radio spectrum utilization is to build cognitive wireless devices [1] that can benefit from the opportunistic deployment of unused spectral opportunities from various frequency bands [1]–[3]. While conceptually simple, the realization of cognitive wireless devices is highly challenging. Several problems must be solved: sensing over a wide frequency band, identifying and characterizing available spectrum opportunities, exploiting the identified transmission opportunities, etc. In particular, as stated in [1], a cognitive wireless device should be able to "learn from the environment and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit power, carrier frequency, and modulation strategy) in real time."

Learning techniques have already been deployed to improve the performance of a broad class of wired and wireless communication systems. They enable the dynamically interacting communication devices to acquire information, build knowledge, and ultimately improve their performance [4]–[7]. For instance, appropriate learning solutions are studied in distributed environments consisting of players with very limited information about their opponents, such as the Internet [4]. In [5], a reinforcement learning algorithm is proposed to maximize the average throughput in sensor communications without explicitly knowing the model of the environment. By modeling the interaction among noncooperative nodes in wireless ad hoc networks as a repeated game, a reinforcement learning algorithm is proposed to design power control in wireless ad hoc networks [6], where it is shown that the learning dynamics can eventually converge to Nash equilibrium and achieve a satisfactory performance. In [7], a novel learning approach is proposed for wireless users to dynamically and efficiently share spectrum resources by considering the time-varying properties of their traffic and channel conditions.

As opposed to the previous works, which focus on studying the long-term convergence behavior of certain learning algorithms [4]–[6] or determine the operational shorter-term performance without providing any performance guarantees [7], this paper aims at characterizing and analytically quantifying the achievable performance that can be obtained by cognitive users with learning capabilities in wireless networks. We study how much a cognitive device with no prior knowledge should learn about its environment, e.g., the time-varying channel condition or experienced interference, in order to reach its performance (utility) requirement. Particularly, if the environment is stationary, we explicitly quantify the benefits that a user can derive in terms of its improved utility by learning for a longer duration, i.e., based on a larger number of observations about the environment. We apply optimization and large-deviation theory to derive an upper bound of the minimum observation duration, given the performance guarantee desired by the user. Then, noticing that the information required for cognitive devices to perform learning is usually gathered through the information feedback from the receiver to the transmitter and, hence, that this information can be delayed during the feedback process, we study how a cognitive device can improve its performance if it accurately knows the feedback delay. We also quantify the impact of imperfect delay measurements on the achieved performance.

While this paper focuses on studying learning in wireless network settings, the proposed solutions can be generalized to other applications [5], [7] wherein cognitive communication devices deploy strategic learning solutions to accumulate knowledge about its environment and, based on this, improve its performance. The rest of this paper is organized as follows. Section II presents the deployed system model and formulates the problem of learning and adapting to the spectrum usage pattern. In Section III, we analytically derive an upper bound of the minimum required learning duration. Section IV presents

Manuscript received June 11, 2008; revised November 7, 2008 and December 15, 2008. First published January 6, 2009; current version published May 29, 2009. This work was supported by NSF 0830556 and ONR. The review of this paper was coordinated by Dr. Y.-C. Liang.

The authors are with the Electrical Engineering Department, University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: yisu@ee.ucla.edu; mihaela@ee.ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2008.2012107

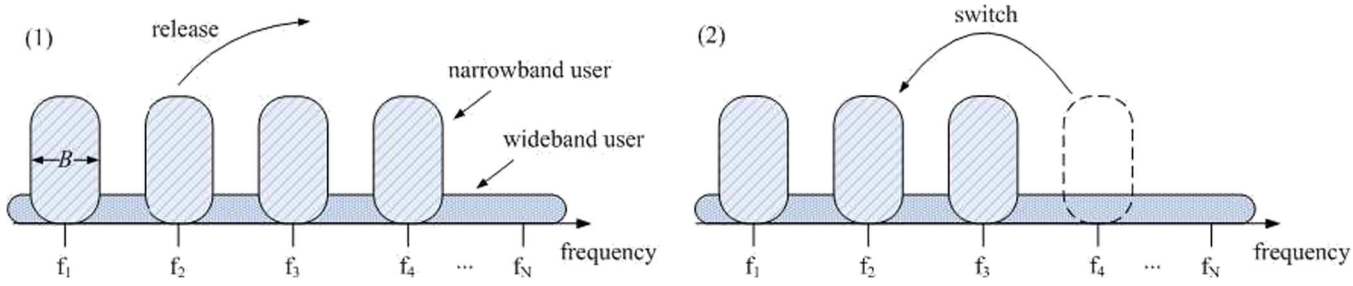

Fig. 1. Investigated cognitive wireless networks.

several illustrative numerical results, and Section V quantifies the impact of spectrum usage information feedback delay. Conclusions are drawn in Section VI.

II. System Model

In this section, we present the mathematical model of the investigated dynamic wireless system and formulate the problem of learning the stationary spectrum occupation pattern by a cognitive transceiver [1].

A. System Description

We assume a cognitive wireless system that is similar with the one studied in [3] (see Fig. 1). The total number of frequency channels in the system is N, and each has a bandwidth of B. The majority of radio devices in this system are narrowband users. These devices can dynamically utilize the idle spectrum bands by enabling carrier frequency switching and "packing" all the active radios tightly in the spectral domain. A simple example is shown in Fig. 1. If one device releases the frequency band f_2, the device occupying frequency f_4 will switch to f_2. The system state is defined as the number of channels n_{nb} that are occupied by the narrowband users. The arrival and departure rates of these devices are assumed to follow a Poisson distribution. As a result, the spectrum usage pattern can be captured as a continuous-time Markov chain [3], [8], [9]. Fig. 2 shows an example of the Markov chain with the infinitesimal generator [10]

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 & & & \\ \mu_1 & -(\lambda_2+\mu_1) & \lambda_2 & & \\ & \mu_2 & -(\lambda_3+\mu_2) & \ddots & \\ & & \ddots & \ddots & \lambda_N \\ & & & \mu_N & -\mu_N \end{bmatrix}. \tag{1}$$

Note that the Markov chain model and its corresponding infinitesimal generator Q can take various forms based on the configuration of the considered wireless network. Denote the steady-state probability vector of the spectrum usage pattern as $\boldsymbol{\pi} = [\pi_0, \pi_1, \ldots, \pi_N]$, in which π_i represents the probability of having active i narrowband devices in the system. No matter what form the infinitesimal generator $\boldsymbol{Q}$ takes, we always have

$$\boldsymbol{\pi Q} = 0. \tag{2}$$

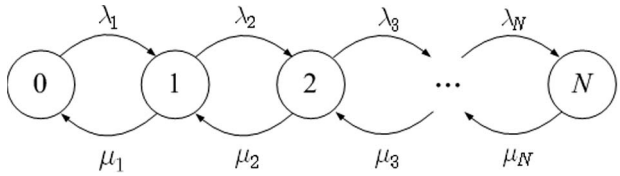

Fig. 2. Example of continuous-time Markov chain model.

As shown in Fig. 1, we also consider a wideband device in the system, which can transmit over all N frequency channels. The noise power at frequency band i is N_i, and its channel gain is h_i.[1] Each active narrowband device causes an interference power of I to the wideband receiver. The wideband device is subjected to a total power constraint of $P^{\max}$. Denote the power vector across all frequency bands $\boldsymbol{P} = [P_1, \ldots, P_N]^{\mathrm{T}}$, in which P_i is the power allocated in frequency band i. Note that the probability that the wideband receiver experiences an interference in channel i from the narrowband user equals $\sum_{n\geq i}^{N} \pi_n$, and the probability of having no interference equals $\sum_{n=0}^{n<i} \pi_n$. Hence, the achievable rate is given by

$$R(\boldsymbol{\pi}, \boldsymbol{P}) = \sum_{i=1}^{N} \left(\sum_{n\geq i}^{N} \pi_n B \log\left(1 + \frac{h_i P_i}{N_i + I}\right) + \sum_{n=0}^{n<i} \pi_n B \log\left(1 + \frac{h_i P_i}{N_i}\right) \right). \tag{3}$$

Note that we do not associate priority types (i.e., primary or secondary) with the narrowband and wideband devices. However, it is possible to consider the interaction among devices of different priorities. For example, transmit-power mask for the cognitive wideband radio transmitter can be created [20].

B. Learning Duration and Performance

As previously mentioned, we would like to determine to what extent the wideband device should learn the spectrum usage pattern in order to explore its available spectrum opportunities. Fig. 3 shows this learning process wherein the wideband receiver periodically senses the spectrum and feeds back to its transmitter the number of interfering narrowband devices n_{nb}^t at time t. Specifically, the wideband device models its environment by simply counting the number of active narrowband

[1]This paper assumes a static channel. The same technique can be applied if the channel dynamics is also taken into consideration. For example, a Rayleigh fading channel can also be modeled using the Markov chain model [23].

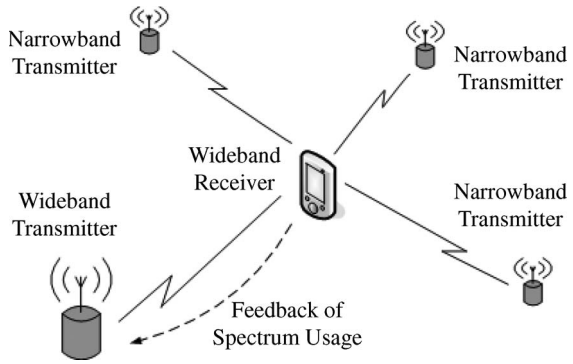

Fig. 3. Spectrum usage feedback of the wideband device.

devices that it encountered in the past and approximating the stationary spectrum usage pattern $\boldsymbol{\pi}$ by the observed frequencies of the system states discussed previously.[2] We define an empirical frequency function

$$\gamma^t(n) = c^t(n) \Big/ \sum_{n=0}^{N} c^t(n) \tag{4}$$

where $c^t(n)$ is a counting function satisfying $c^0(n) = 0, \forall n \in \{0, 1, \ldots, N\}$, and

$$c^t(n) = \begin{cases} c^{t-1}(n) + 1, & \text{if } n_{nb}^t = n \\ c^{t-1}(n), & \text{otherwise.} \end{cases} \tag{5}$$

The wideband user approximates the steady state π using the empirical frequency function $\boldsymbol{\gamma}^t$ and takes the best response action $\boldsymbol{P}(\boldsymbol{\gamma}^t)$ that maximizes

$$R(\boldsymbol{\gamma}^t, \boldsymbol{P}) = \sum_{i=1}^{N} \left(\sum_{n \geq i}^{N} \gamma^t(n) B \log\left(1 + \frac{h_i P_i}{N_i + I}\right) + \sum_{n=0}^{n<i} \gamma^t(n) B \log\left(1 + \frac{h_i P_i}{N_i}\right) \right) \tag{6}$$

i.e., $\boldsymbol{P}(\boldsymbol{\gamma}^t) = \arg\max_{\boldsymbol{P}^{\mathrm{T}}\mathbf{1} \leq P^{\max}} R(\boldsymbol{\gamma}^t, \boldsymbol{P})$, with $\mathbf{1} = [1, \ldots, 1]^{\mathrm{T}}$.[3] We denote the achievable rate when the wideband user takes the best response to the empirical frequency function γ^t as $R_a(\gamma^t) = R(\boldsymbol{\pi}, \boldsymbol{P}(\gamma^t))$. Similarly, we define the maximal achievable rate to be $R_a(\boldsymbol{\pi}) = R(\boldsymbol{\pi}, \boldsymbol{P}(\boldsymbol{\pi}))$.

Throughout this paper, learning *duration* refers to the number of available observed spectrum usage patterns over time for the wideband user to update $\gamma^t(n)$ and approximate the steady-state distribution $\boldsymbol{\pi}$. Intuitively, the performance of learning is expected to improve if more observations are available. In this paper, we aim at determining how many observations are sufficient for a learning user to reach a certain desirable performance guarantee. Specifically, given the tolerable performance loss Δ_R with respect to perfectly knowing $\boldsymbol{\pi}$ and the outage probability δ_R, we want to determine

$$\boxed{\begin{array}{l} \textit{Minimum Required Learning Duration}: \\ \min\ t,\ \ s.t.\ \mathbf{Prob}\left(R_a(\boldsymbol{\pi}) - R_a(\gamma^t) \geq \Delta_R\right) \leq \delta_R. \end{array}} \tag{7}$$

The wideband user's learning and adaptation mechanisms are summarized in Table I.

The next section will investigate this tradeoff between learning duration and its achievable performance.

III. Minimum Required Learning Duration

This section aims at solving the previous stated problem of determining the minimum learning duration for a cognitive user in a stationary environment, given its tolerable performance loss and outage probability. Specifically, we derive an upper bound of the minimum required learning duration and discuss the tradeoff between the learning duration and the achievable performance.

Although similar bounds exist in statistical learning theory, e.g., Hoeffding's inequality [11], it is still difficult to solve the problem in (6) because these bounds do not directly apply to our considered problem. However, we can find an upper bound for the solution of the problem in (6). Having such a bound is important from both a theoretical and a practical perspective, because, due to the real-time adaptation requirement of cognitive networks [1], only limited observations are usually available to cognitive users, and thus, it becomes necessary for them to understand the basic tradeoff that can be made between the obtained performance and the learning duration. For this, we adopt tools from large-deviation theory, which quantifies the exponential decay of probability measures for certain kinds of tail events [12]. According to the large-deviation theory, the empirical frequency function $\gamma^t(n)$ of a random sample of size t drawn from $\boldsymbol{\pi}$ satisfies

$$\mathbf{Prob}\left(D(\boldsymbol{\gamma}^t \| \boldsymbol{\pi}) \geq \delta\right) \leq \binom{N+t}{N} 2^{-\delta t} \qquad \forall \delta > 0 \tag{8}$$

where $D(p\|q)$ is the Kullback–Leibler (KL) distance between two probability mass functions (pmfs) $p(x)$ and $q(x)$ [13]. Then, we need to convert the performance loss $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}^t)$ into the KL distance $D(\boldsymbol{\gamma}^t \| \boldsymbol{\pi})$. Note that these two metrics do not always perfectly align with each other. The basic idea in determining an upper bound is to find a value of δ such that $D(\boldsymbol{\gamma}^t \| \boldsymbol{\pi}) \leq \delta$ always leads to $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}^t) \leq \Delta_R$. By setting t to satisfy $\binom{N+t}{N} 2^{-\delta t} \leq \delta_R$, we have $\mathbf{Prob}(D(\boldsymbol{\gamma}^t \| \boldsymbol{\pi}) \geq \delta) \leq \mathbf{Prob}(R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}^t) \geq \Delta_R)$ and, this value provides an upper bound for the problem in (6). As shown in Fig. 4, we divide this procedure into three steps. First, we construct a convex set $\mathcal{B}$ in the standard probability simplex $\Omega = \{\boldsymbol{\gamma} | \mathbf{1}^{\mathrm{T}} \boldsymbol{\gamma} = 1, \boldsymbol{\gamma} \succeq 0\}$ such that, for all $\boldsymbol{\gamma} \in \mathcal{B}$, it satisfies $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}) \leq \Delta_R$. Second, by solving convex optimization problems that minimize the KL distance between $\boldsymbol{\pi}$ and the pmfs that lie on the boundary of $\mathcal{B}$, we obtain the desired value of δ, which is denoted as $\delta_{D_{\min}}$ in Fig. 4.

[2] Typical detection methods include energy detection, coherent detection, etc. [21].

[3] Note that, here, we normalize the feedback period, and we implicitly assume that this period is sufficiently large such that the spectrum usage pattern will be independent of the previous sampled usage pattern. Section V will discuss the optimal strategies for various feedback delays and sampling intervals.

TABLE I
WIDEBAND USER'S LEARNING AND ADAPTATION MECHANISMS

Initialization : $t=0, c^0(n)=0, \forall n \in \{0,1,\cdots,N\}$

Repeat

I. Measure the number of currently active narrowband users;

II. Update the counting function and empirical frequency function according to (5) and (6);

III. Update the power allocation using the best response $\boldsymbol{P}(\boldsymbol{\gamma}^t) = \arg \max_{\boldsymbol{P}^T \mathbf{1} \le P^{\max}} R(\boldsymbol{\gamma}^t, \boldsymbol{P})$;

IV. $t = t+1$.

until the minimum required learning duration is attained.

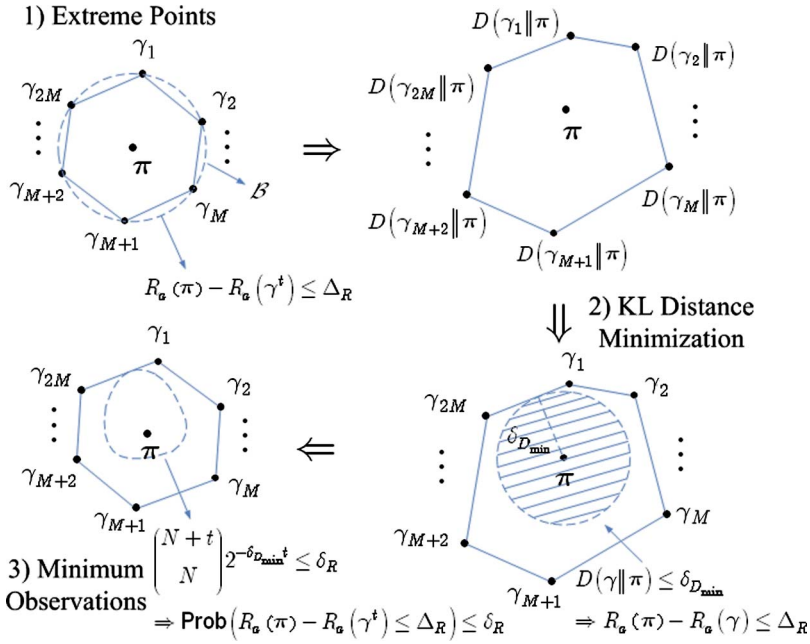

Fig. 4. Performance loss and KL distance.

Third, we apply large-deviation theory and derive an upper bound of the minimum required observations. In the following sections, we will explain each step in detail.

A. Extreme Points With Performance Loss Constraints

First, in the probability simplex Ω, we construct a convex set $\mathcal{B}$ that contains the actual pmf. Let $A = \{\{k,j\} : k,j \in \{0,1,\ldots,N\} \text{ and } k<j\}$ such that A contains a total number of $M = \binom{N+1}{2}$ combinations of any two different integers in $\{0,1,2,\ldots,N\}$. Let $(S)_m$ denote the mth element of set S. Based on the tolerable performance loss Δ_R, we choose a total number of $2M$ pmfs and view them as "extreme points" of the set $\mathcal{B}$ in which we are going to derive an upper bound of the minimum required learning duration. For $m = 1,2,\ldots,2M$, the $2M$ pmfs that we are interested in satisfy the following:

P1) $\gamma_m \in \Omega$;
P2) $\gamma_m(n) = \pi_n, \quad \text{if } n \notin (\mathrm{A})_m$.

Note that P2) ensures that these pmfs have only two elements that are different from the stationary distribution $\boldsymbol{\pi}$. The pmfs that satisfy P1) and P2) can be rewritten as $\gamma_m(n, \delta_m)$ defined by

$$\gamma_m(n,\delta_m) = \begin{cases} \pi_n - \delta_m, & \text{if } n = ((A)_m)_1 \\ \pi_n + \delta_m, & \text{if } n = ((A)_m)_2 \\ \pi_n, & \text{if } n \notin (A)_m. \end{cases}, \quad m = 1,2,\ldots,2M \tag{9}$$

Denoting $\boldsymbol{\gamma}_m(\delta_m) = [\gamma_m(0,\delta_m), \ldots, \gamma_m(N,\delta_m)]$, now, we can determine the extreme points by setting the parameter δ_m based on the tolerable performance loss Δ_R. For $m = 1,2,\ldots,M$

$$\delta_m = \begin{cases} \pi_l \text{ with } l = ((A)_m)_1, & \text{if } S_\delta = \varnothing \\ \min \delta \in S_\delta, & \text{otherwise} \end{cases} \tag{10}$$

in which $S_\delta = \{\delta : R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}_m(\delta)) \ge \Delta_R \text{ and } \delta \ge 0\}$, and

$$\delta_{m+M} = \begin{cases} -\pi_l, \text{ with } l = ((A)_m)_2, & \text{if } S_{-\delta} = \varnothing \\ \min \delta \in S_{-\delta}, & \text{otherwise} \end{cases} \tag{11}$$

where $S_{-\delta} = \{\delta : R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}_m(-\delta)) \ge \Delta_R \text{ and } \delta \ge 0\}$. Due to the nonnegative property in P1), when $n \in (\mathrm{A})_m$, if

$S_\delta = \varnothing$ or $S_{-\delta} = \varnothing$, we set $\gamma_m(n, \delta_m)$ to be zero to ensure that the performance loss is as close to Δ_R as possible. On the other hand, if $S_\delta \neq \varnothing$ or $S_{-\delta} \neq \varnothing$, the "extreme points" are the pmfs that cause an exact performance loss of Δ_R.

Using the convex hull of the aforementioned $2M$ extreme points, we construct a convex set $\mathcal{B}$ within which to derive an upper bound of the minimum required learning duration in (6), i.e.,

$$\mathcal{B} = \left\{ \boldsymbol{\gamma} : \boldsymbol{\gamma} = \sum_{m=1}^{2M} \alpha_m \boldsymbol{\gamma}_m(\delta_m), \quad \alpha_m \geq 0, \quad \text{and} \ \sum_{m=1}^{2M} \alpha_m = 1 \right\}. \tag{12}$$

Proposition 1 (Satisfying Performance Loss Constraints): Any $\gamma \in \mathcal{B}$ satisfies $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}) \leq \Delta_R$.

The proof is given in Appendix A. Proposition 1 ensures that any convex combinations of the extreme points still satisfy the tolerable performance loss requirement, which enables us to apply optimization theory to convert the metric of performance loss Δ_R into KL distance $\delta_{D_{\min}}$ in the following step.

B. KL Distance Minimization in Convex Set

In the first step, a convex set $\mathcal{B}$ is constructed based on the tolerable performance loss Δ_R. Next, we apply large-deviation theory to translate the performance loss Δ_R into another metric, the KL distance δ_D. The basic idea is to solve an optimization problem to find the minimum KL distance $\delta_{D_{\min}}$ such that, for any $\boldsymbol{\gamma}$ that satisfies $D(\boldsymbol{\gamma}\|\boldsymbol{\pi}) \leq \delta_{D_{\min}}$, we have $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}) \leq \Delta_R$. Particularly, the optimization problem can be formulated as

$$\begin{aligned} &\min_{\boldsymbol{\gamma}} D(\boldsymbol{\gamma}\|\boldsymbol{\pi}) \\ &\text{s.t. } \boldsymbol{\gamma} \in \mathcal{S}(\mathcal{B}) \end{aligned} \tag{13}$$

where $\mathcal{S}(\mathcal{B})$ represents the surface of the convex set $\mathcal{B}$, i.e., $\mathcal{S}(\mathcal{B}) = \mathcal{B} \setminus \text{int}(\mathcal{B})$. Here, we denote the interior of the set $\mathcal{B}$ as int($\mathcal{B}$) [14].

Note that the KL distance $D(\boldsymbol{\gamma}\|\boldsymbol{\pi})$ is convex in the pair $(\boldsymbol{\gamma}, \boldsymbol{\pi})$, and $\gamma \in \mathcal{S}(\mathcal{B})$ is a linear constraint [13]. Therefore, the problem in (13) essentially belongs to convex programming, and the optimal solution can efficiently be obtained by solving the optimization problem for each polyhedron on the boundary $\mathcal{S}(\mathcal{B})$ [15]. Because the convex combinations of the extreme points in $\mathcal{B}$ cover the adjacent region of the actual stationary distribution $\boldsymbol{\pi}$, the minimum of (13) that ensures that $D(\boldsymbol{\gamma}\|\boldsymbol{\pi}) \leq \delta_{D_{\min}}$ is a sufficient condition to ensure that $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}) \leq \Delta_R$.

C. Minimum Learning Duration Calculation

In the second step, we show that $D(\boldsymbol{\gamma}\|\boldsymbol{\pi}) \leq \delta_{D_{\min}}$ always leads to $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}) \leq \Delta_R$. Hence, an upper bound of the solution to the problem in (6) can be obtained by solving

$$\begin{aligned} &\min t \\ &\text{s.t.} \quad \mathbf{Prob}\left(D(\boldsymbol{\gamma}^t\|\boldsymbol{\pi}) \geq \delta_{D_{\min}}\right) \leq \delta_R. \end{aligned} \tag{14}$$

Applying (8) from large-deviation theory, we have the following proposition.

Proposition 2 (An Upper Bound of Minimum Required Learning Duration): Suppose the wideband device updates its empirical frequency function $\boldsymbol{\gamma}^t$ and takes the best-response action with respect to $\boldsymbol{\gamma}^t$. An upper bound T of the solution of problem (6) is

$$T = Min\_t(\delta_{D_{\min}}, N, \delta_R) \tag{15}$$

in which

$$Min\_t(x, y, z) = \min\left\{ t : t \in \mathcal{Z}^+ \quad \text{and} \quad \binom{y+t}{y} \cdot 2^{-tx} \leq z \right\}.$$

Proof: Combining (8) and (14), we know that any t that satisfies

$$\binom{N+t}{N} 2^{-\delta_{D_{\min}} t} \leq \delta_R \tag{16}$$

is an upper bound of the solution of problem (6). Let $F(t) = \binom{N+t}{N} 2^{-\delta_{D_{\min}} t}$. We have $F(t+1)/F(t) = (1 + (N/t+1)) 2^{-\delta_{D_{\min}}}$ and $\lim_{t\to\infty}(F(t+1)/F(t)) = 2^{-\delta_{D_{\min}}} < 1$. Therefore, we can conclude that $\lim_{t\to\infty} F(t) = 0$. As a result, by choosing $T = Min\_t(\delta_{D_{\min}}, N, \delta_R)$ as the minimum integer in the feasible region of inequality (16), we obtain an upper bound of the optimum solution of (6). ■

Subsequently, we provide some intuition to interpret the previously derived upper bound. Define $f : \mathcal{R} \to \mathcal{R}$ to be the function that maps the tolerable performance loss Δ_R into the minimum KL distance $\delta_{D_{\min}}$. Obviously, f is a nonincreasing function because a larger Δ_R enlarges the set $\mathcal{B}$ and increases the corresponding $\delta_{D_{\min}}$. The upper bound of the minimum learning duration can be rewritten as

$$T = Min\_t\left(f(\Delta_R), N, \delta_R\right). \tag{17}$$

We can make several key observations by examining this upper bound.

Remark 1: Decreasing the acceptable performance loss Δ_R will lead to a larger minimum observation duration T, which is a direct consequence of the nonincreasing property of function f.

Remark 2: Decreasing the outage probability δ_R will increase T. This remark is also quite intuitive.

Remark 3: If the number of channels N is increased, the upper bound of the required observations T also increases in order to ensure that the outage probability is smaller than the threshold of δ_R. This argument holds because a larger number of channels N will cause $\binom{N+t}{N}$ to increase and $\delta_{D_{\min}}$ to be smaller than or equal to its original value (given that the steady-state probability distribution $\boldsymbol{\pi}$ is unchanged). Intuitively, a larger N adds additional uncertainty in the learning process and increases the upper bound T.

IV. Illustrative Examples

This section simulates an example to illustrate all the previously proposed procedures. We consider a cognitive system

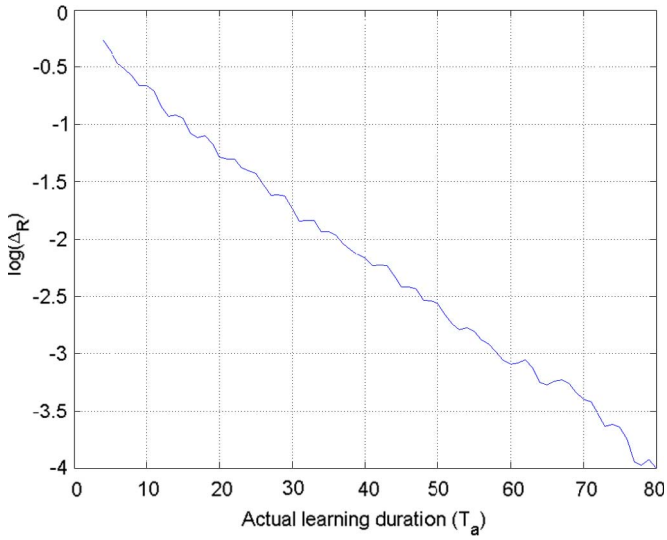

Fig. 5. Learning duration and performance loss.

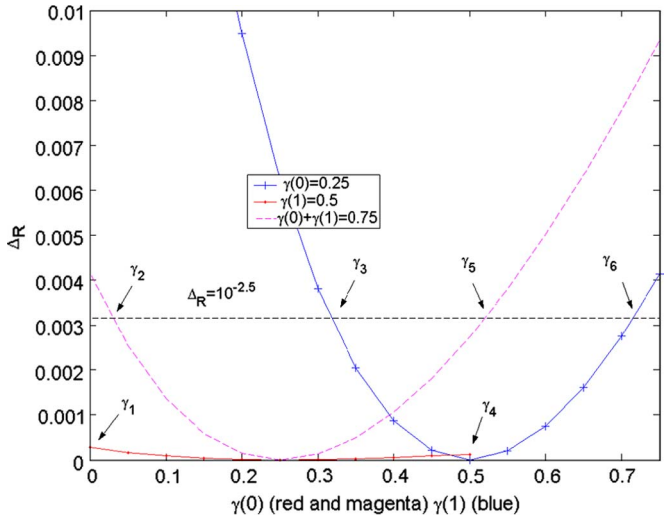

Fig. 6. Constructing the extreme points.

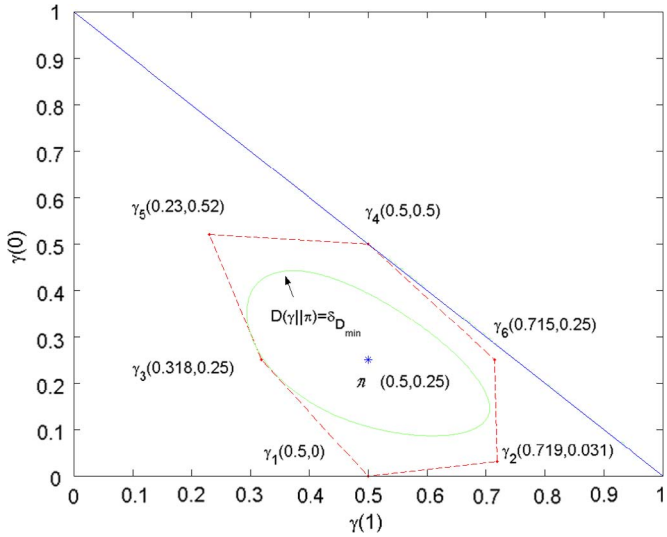

Fig. 7. KL distance minimization in $\mathcal{S}(\mathcal{B})$.

with $N = 2$, $\lambda_1 = \mu_2 = 2$ users/time slot, and $\lambda_2 = \mu_1 = 1$ user/time slot, and the power constraint of the wideband device is $P^{\max} = 40$ dBm. Its channel gain and the power of noise and interference are given by $h_1 = -117$ dB, $h_2 = -120$ dB, and $N_1 = N_2 = I = -80$ dBm. It is easy to solve that the stationary distribution is $\boldsymbol{\pi} = [0.25 \quad 0.5 \quad 0.25]$. Fig. 5 shows the simulated learning curve that indicates the learning duration versus the resulting performance loss. We can see that the performance loss Δ_R is decreased by learning for a longer duration.

We set the parameters in problem (6) to be $\Delta_R = 10^{-2.5}$ and $\delta_R = 10^{-2}$. Figs. 6 and 7 show the procedure of obtaining the upper bound in Section III-B. Noting that $N = 2$, we choose six extreme points $\gamma_1, \ldots, \gamma_6$ in total, which are determined based on the rate–pmf curves in Fig. 6. These plotted curves indicate the achievable rates for three pmfs, including $\gamma(0) = 0.25$, $\gamma(1) = 0.5$, and $\gamma(2) = 0.25$, i.e., $\gamma(0) + \gamma(1) = 0.75$. The convex hull of these extreme points $\gamma_1, \ldots, \gamma_6$ is the extreme point set $\mathcal{B}$. The dashed hexagon in Fig. 7 is the surface $\mathcal{S}(\mathcal{B})$ on which we minimize the KL distance. Solving the convex optimization problem (13) leads to $\delta_{D_{\min}} = 0.1265$. Using (15), we obtain that $T = Min\_t(0.1265, 2, 10^{-2}) = 161$. As shown in Fig. 7, if the learning duration is larger than T, the KL distance between the actual stationary distribution $\boldsymbol{\pi}$ and the observed empirical frequency function $\boldsymbol{\gamma}^t$ will lie within the solid circle with an outage probability that is less than δ_R.

We also examine the tightness of the upper bound in different settings. The tolerable performance loss Δ_R is varied to be 10^{-2}, $10^{-2.5}$, and 10^{-3}, while the outage probability δ_R is set to be a constant of 10^{-2}. In each scenario, we use Monte Carlo methods to calculate the actual required learning duration T_a. The results are summarized in Table II. From the table, we can see that the bound is not very tight, which can be explained by the observation that the space between the solid circle and the dashed hexagon is large. At the same time, we also find that the ratio of T/T_a increases when the performance loss Δ_R is decreased, because the mismatch between the contours is increased as Δ_R decreases. Moreover, we can also see that, by carefully choosing the extreme points, $\mathcal{S}(\mathcal{B})$ can be enlarged to approach the contour of $R_a(\boldsymbol{\gamma})$ and therefore improve the tightness of the upper bound. However, since we are mostly interested in deriving the minimum required learning duration for intermediate values of Δ_R, the actual value T_a and the upper bound T are still of the same magnitude. In addition, it is important to note that, even though the bound is not tight, it still guarantees that, sensing the environment and learning for this time interval, the cognitive device can achieve the desired performance.

TABLE II
LEARNING DURATIONS FOR DIFFERENT PERFORMANCE LOSS REQUIREMENTS

Performance loss Δ_R	10^{-2}	$10^{-2.5}$	10^{-3}
KL distance $\delta_{D_{\min}}$	0.1887	0.1265	0.0406
Actual value T_a	38	50	62
Upper bound T	101	161	593
T/T_a	2.7	3.2	9.6

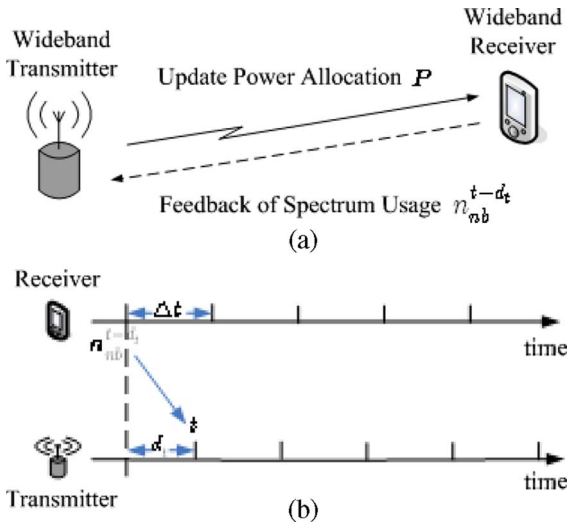

Fig. 8. Feedback delay of the spectrum usage.

V. Impact of Feedback Delay

In this section, we discuss the impact of the feedback delay of spectrum usage information, which causes the received information to be out of date and degrades the performance. The feedback delay exists due to several reasons, e.g., wireless propagation, signal processing expense, and protocol overhead.

We denote the feedback delay of the spectrum usage pattern n_{nb} from the receiver to the transmitter as d_t. As shown in Fig. 8, the spectrum usage pattern that the transmitter receives at time t is actually the usage pattern that the receiver experienced at time $t - d_t$.

As stated in Section III, the infinitesimal generator Q of the Markov chain can take various forms based on the system specification. Define the transition probability matrix $\boldsymbol{S}(t)$, in which $S_{i,j}(t)$ is the probability that a Markov process is in state j at time t, given that it is in state i at time 0. Based on the stochastic process theory [10], we know that $\boldsymbol{S}(t)$ is the solution of the Kolmogorov equation, which takes the form of

$$\boldsymbol{S}(t) = \sum_{i=1}^{N+1} \boldsymbol{v}_i e^{t\xi_i} \boldsymbol{\omega}_i \tag{18}$$

in which $\xi_1, \xi_2, \ldots, \xi_{N+1}$ are the $N+1$ distinct eigenvalues of matrix $\boldsymbol{Q}$, and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{N+1}$ and $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_{N+1}$ are the corresponding right and left eigenvectors of matrix $\boldsymbol{Q}$. In particular, the matrix $\boldsymbol{Q}$ for the considered Markov process has an eigenvalue $\xi_1 = 0$ with the corresponding right and left eigenvectors $\boldsymbol{v}_1 = [1, 1, \ldots, 1]^{\mathrm{T}}$ and $\boldsymbol{\omega}_1 = \boldsymbol{\pi}$. All the other eigenvalues $\xi_2, \ldots, \xi_{N+1}$ of $\boldsymbol{Q}$ have strictly negative real parts.

Given the latest feedback $n_{nb}^{t-d_t}$, the optimization of power allocation at the transmitter is converted into

$$\max_{\boldsymbol{P}^{\mathrm{T}} 1 \le P^{\max}} R\left(\boldsymbol{\pi}^t, \boldsymbol{P} \middle| n_{nb}^{t-d_t}\right) \tag{19}$$

in which $\boldsymbol{\pi}^t = [\pi_0^t, \pi_1^t, \ldots, \pi_N^t]$ is the probability vector of the spectrum usage pattern n_{nb}^t with $\pi_n^t | n_{nb}^{t-d_t} = S_{n_{nb}^{t-d_t},n}(d_t) = \Pr(n_{nb}^t = n | n_{nb}^{t-d_t})$. From (18), we have

$$\lim_{t \to +\infty} \boldsymbol{S}(t) = \boldsymbol{v}_1 \boldsymbol{\omega}_1. \tag{20}$$

Therefore, when $d_t \to +\infty$, $\boldsymbol{\pi}^t \to \boldsymbol{\omega}_1 = \boldsymbol{\pi}$, which is independent of $n_{nb}^{t-d_t}$. As a result, $R(\boldsymbol{\pi}^t, \boldsymbol{P} | n_{nb}^{t-d_t})$ in (19) is reduced to $R(\boldsymbol{\pi}, \boldsymbol{P})$ in (3). We can conclude that learning the stationary distribution $\boldsymbol{\pi}$ of frequency usage pattern and optimizing the power allocation with respect to this distribution are optimal only when the feedback delay is large.

On the other hand, we note that the achievable rate in (3) can be further improved, if both the transmitter and the receiver have perfect and instantaneous channel state information [22], i.e., the delay of information feedback is zero. In fact, in the limited feedback delay scenarios, the best strategy is not to learn the stationary distribution, and the transmitter needs to explore the timeliness of the feedback information $n_{nb}^{t-d_t}$, because $\boldsymbol{\pi}^t$ in (19) is a function of the limited feedback delay d_t. In particular, $R(\boldsymbol{\pi}^t, \boldsymbol{P} | n_{nb}^{t-d_t})$ in the optimal transmission strategy of (19) will become

$$\begin{aligned} & R\left(\boldsymbol{\pi}^t, \boldsymbol{P} | n_{nb}^{t-d_t}\right) \\ &= R\left(S_{n_{nb}^{t-d_t},:}(d_t), \boldsymbol{P}\right) \\ &= \sum_{i=1}^{N} \left(\sum_{n \ge i}^{N} S_{n_{nb}^{t-d_t},n}(d_t) \cdot B \log\left(1 + \frac{h_i P_i}{N_i + I}\right) \right. \\ &\qquad \left. + \sum_{n=0}^{n<i} S_{n_{nb}^{t-d_t},n}(d_t) B \log\left(1 + \frac{h_i P_i}{N_i}\right) \right) \end{aligned} \tag{21}$$

where $S_{i,:}(d_t)$ represents the ith row of $\boldsymbol{S}(d_t)$. The problem is converted into how to accurately estimate $\boldsymbol{S}(t)$ at $t = d_t$. Due to the periodic nature of the feedback information n_{nb}^t, the wideband device is able to sample the transition probability matrix $\boldsymbol{S}(t)$ at $t = \Delta t, 2\Delta t, \ldots$ by updating empirical frequency functions and use numerical algorithms [16], such as curve fitting, to estimate $\boldsymbol{S}(t)$ for noninteger multiples of Δt. As long as the environment is stationary and the sampling data are large enough, the wideband device can accurately estimate $\boldsymbol{S}(d_t)$.

Now, we investigate the impact of imperfect estimation of the feedback delay d_t. Practical methods of measuring the feedback can be found in [17] and [18]. Suppose that the estimate that the wideband device has about the feedback delay d_t is d_t'. The performance degradation $\Delta R(d_t')$ of imperfect estimation d_t' is given by

$$\begin{aligned} \Delta R\left(d_t'\right) = \sum_{i=0}^{N} \pi_i \left[R\left(S_{i,:}(d_t), \boldsymbol{P}\left(S_{i,:}(d_t)\right)\right) \right. \\ \left. - R\left(S_{i,:}(d_t), \boldsymbol{P}\left(S_{i,:}\left(d_t'\right)\right)\right) \right]. \end{aligned} \tag{22}$$

We derive an upper bound of this performance degradation based on Markov chain theory and formally state the result as Theorem 1.

Theorem 1: The performance degradation $\Delta R(d_t')$ defined in (22) depends on two terms $|d_t' - d_t|$ and $\min(d_t', d_t)$.

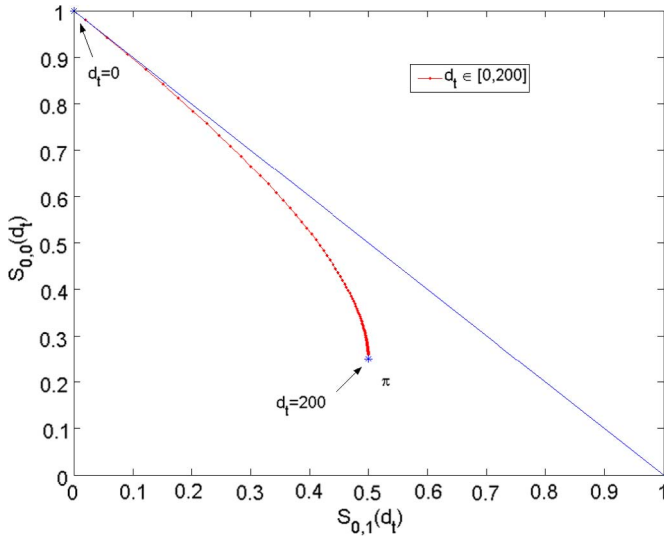

Fig. 9. Transition probability of $S_{0,0}(d_t)$ and $S_{0,1}(d_t)$.

Specifically, $\Delta R(d_t')$ is bounded as

$$0 \leq \Delta R\left(d_t'\right) \leq \alpha\left(\left|d_t' - d_t\right|\right) e^{-\beta \min\left(d_t', d_t\right)} \tag{23}$$

in which $\alpha(\bullet)$ is a nonnegative function satisfying $\alpha(0) = 0$ and $\lim_{t\to+\infty} \alpha(t)$ exists, and $\beta > 0$.

Proof: See Appendix B.

Two key observations can be made from the aforementioned theorem. First, it is straightforward to see that the performance loss is a function of $|d_t' - d_t|$ and the performance loss is zero if $d_t' = d_t$. More importantly, the theorem indicates that the performance loss decreases at least exponentially with $\min(d_t', d_t)$. This result indicates the significance of the timeliness of the information feedback. In addition, the existence of $\lim_{t\to+\infty} \alpha(t)$ implies that the infinite estimation error of the feedback delay causes bounded performance loss. With the increase of $\min(d_t', d_t)$, the effect of inaccurate estimation of the delay d_t over the performance diminishes at least exponentially.

We verify the performance improvement by considering the feedback delay. We use an example with the parameters $N = 2, \lambda_1 = \mu_2 = 0.02$ user/time slot, and $\lambda_2 = \mu_1 = 0.01$ user/time slot. It is easy to show that, for example, the three eigenvalues of $\boldsymbol{Q}$ are $\xi_1 = 0, \xi_2 = -0.02$, and $\xi_3 = -0.04$, and the transition probability matrix $\boldsymbol{S}(t)$ is given by (24), shown at the bottom of the page.

The transition probability of $S_{0,0}$ and $S_{0,1}$ is plotted as a function of the feedback delay d_t in Fig. 9. As we expect, if the feedback delay $d_t \to 0$, the spectrum usage pattern n_{nb}^t has a large possibility to be equal to $n_{nb}^{t-d_t}$, i.e., the transmitter exactly knows how many narrowband users are currently active. On the other hand, if $d_t \to +\infty$, the spectrum usage pattern will converge to the stationary distribution $\boldsymbol{\pi}$. Therefore, if d_t is not sufficiently large, the wideband transmitter should optimize its power allocation with respect to the transition probability matrix $\boldsymbol{S}(d_t)$ rather than the stationary distribution $\boldsymbol{\pi}$.

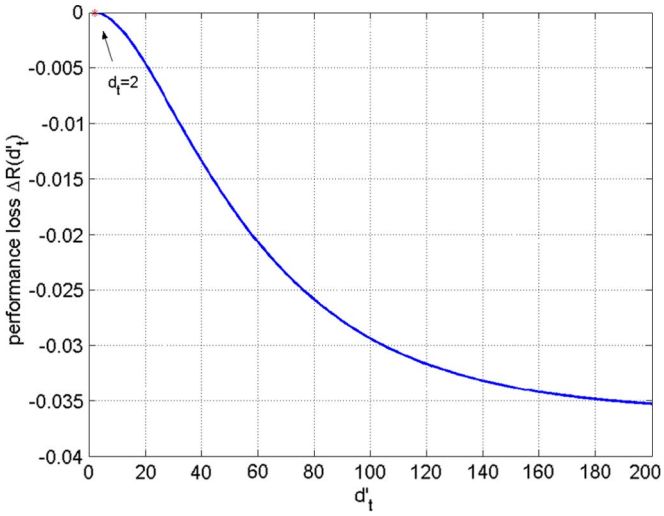

Fig. 10. Performance loss of inaccurate estimate over d_t.

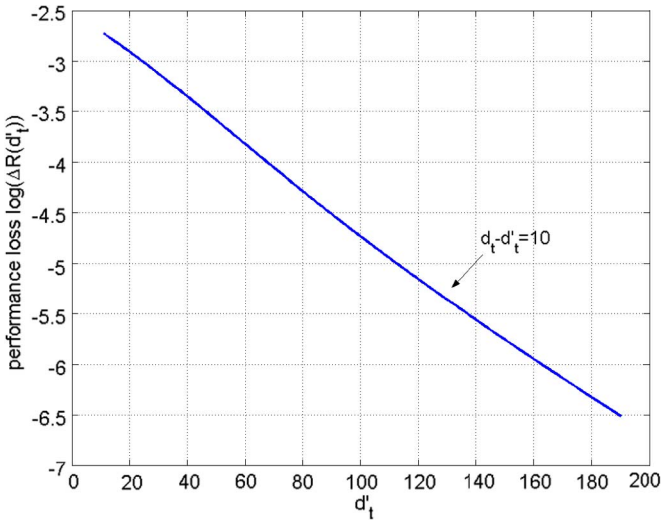

Fig. 11. Performance loss of inaccurate estimate for fixed $d_t - d_t'$.

Next, we numerically show the improvement of measuring the feedback delay d_t. The feedback delay d_t is assumed to be two, and the performance loss $\Delta R(d_t')$ is shown in Fig. 10. We can see that it agrees with the argument that $\alpha(0) = 0$ and that $\lim_{t\to+\infty} \alpha(t)$ exists in Theorem 1. Compared with taking the best response to the stationary distribution, perfectly knowing the value of feedback delay can increase the achievable rate by 3.5%. We also vary d_t' while fixing $d_t - d_t'$ to be ten and plot the corresponding $\Delta R(d_t')$ in Fig. 11. We can see that the performance loss $\Delta R(d_t')$ decreases exponentially with d_t', which complies, as expected, with Theorem 1.

$$\boldsymbol{S}(t) = 0.25 \times \begin{bmatrix} 1 + 2e^{-0.02t} + e^{-0.04t} & 2 - 2e^{-0.04t} & 1 - 2e^{-0.02t} + e^{-0.04t} \\ 1 - e^{-0.04t} & 2 + 2e^{-0.04t} & 1 - e^{-0.04t} \\ 1 - 2e^{-0.02t} + e^{-0.04t} & 2 - 2e^{-0.04t} & 1 + 2e^{-0.02t} + e^{-0.04t} \end{bmatrix} \tag{24}$$

VI. Conclusion

This paper studies the minimum required observations that a wideband user should have in order to learn about the stationary probability distribution of its experienced environment, given the required performance guarantee. The derived results provide several insights for understanding the basic tradeoff that can be made in communication systems between the learning duration and the achievable performance. We also consider the impact of information feedback delay and quantify the performance loss for imperfect estimation of the delay. Such insights are important for designing and evaluating future communication protocols with learning capabilities such that engineers can build practical systems which are able to achieve the desired complexity versus performance tradeoff.

Appendix A

Proof of Proposition 1: We provide the proof for the case of $N=2$. Similar proofs can be established for $N>2$.

For any $\boldsymbol{P}=[P_1 \quad P_2]^{\mathrm{T}}$ satisfying $\boldsymbol{P}^{\mathrm{T}}\mathbf{1}=P^{\max}$, because $R(\boldsymbol{\pi},\boldsymbol{P})$ is concave in $\boldsymbol{P}$, there exists a region $[\underline{P}_1,\overline{P}_1]$ such that $R_a(\boldsymbol{\pi})-R(\boldsymbol{\pi},\boldsymbol{P})\le\Delta_R$ if and only if $P_1\in[\underline{P}_1,\overline{P}_1]$.

It is easy to verify that $((\partial R(\boldsymbol{\gamma},\boldsymbol{P}))/\partial P_i)=(h_i/(N_i+h_iP_i))-\sum_{n\ge i}^{N}\gamma_n(h_iI/((N_i+h_iP_i)(N_i+h_iP_i+I)))$. Based on optimization theory, we know that the optimal solution $\boldsymbol{P}^{\gamma}=[P_1^{\gamma} \quad P_2^{\gamma}]^{\mathrm{T}}$ maximizing $R(\boldsymbol{\gamma},\boldsymbol{P})$ satisfies

$$\left.\frac{\partial R(\boldsymbol{\gamma},\boldsymbol{P})}{\partial P_i}\right|_{P_i=P_i^{\gamma}} = \begin{cases} \frac{h_i}{N_i+h_iP_i^{\gamma}}-\sum\limits_{n\ge i}^{N}\gamma_n\frac{h_iI}{\left(N_i+h_iP_i^{\gamma}\right)\left(N_i+h_iP_i^{\gamma}+I\right)}=\lambda, & \text{if } P_i^{\gamma}>0 \\ \frac{h_i}{N_i}-\sum\limits_{n\ge i}^{N}\gamma_n\frac{h_iI}{N_i(N_i+I)}<\lambda, & \text{if } P_i^{\gamma}=0 \end{cases} \tag{25}$$

in which λ is a constant.

Note that, for any $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ that satisfy $R_a(\boldsymbol{\pi})-R_a(\boldsymbol{\gamma}_i)\le\Delta_R$, $i=1,2$, we have $P_1^{\gamma_i}\in[\underline{P}_1,\overline{P}_1]$. Because $\partial R(\boldsymbol{\gamma},\boldsymbol{P})/\partial P_i$ monotonically decreases in P_i, we have $P_1^{\theta\gamma_1+(1-\theta)\gamma_2}\in[\min(P_1^{\gamma_1},P_1^{\gamma_2}),\max(P_1^{\gamma_1},P_1^{\gamma_2})]$ for any $\theta\in[0,1]$. It follows that $P_1^{\theta\gamma_1+(1-\theta)\gamma_2}\in[\underline{P}_1,\overline{P}_1]$.

Since any $\boldsymbol{\gamma}\in\mathcal{B}$ can be expressed as a convex combination of the extreme points $\boldsymbol{\gamma}_m$ and these extreme points satisfy that $R_a(\boldsymbol{\pi})-R_a(\boldsymbol{\gamma}_i)\le\Delta_R$, we can conclude that $R_a(\boldsymbol{\pi})-R_a(\boldsymbol{\gamma})\le\Delta_R$ for any $\boldsymbol{\gamma}\in\mathcal{B}$. ■

Appendix B

To show Theorem 1, we first derive a lemma that describes the relative distance between the rows of $\boldsymbol{S}(d_t)$ and $\boldsymbol{S}(d_t')$ as a function of d_t and d_t'.

Lemma 1: There exist a nonnegative function $\alpha'(\bullet)$ and a constant $\beta'>0$, such that the difference between the ith row of $\boldsymbol{S}(d_t)$ and $\boldsymbol{S}(d_t')$ is bounded as

$$\sum_{n=0}^{N}\left|S_{i,n}(d_t)-S_{i,n}\left(d_t'\right)\right|\le\alpha_i'\left(\left|d_t'-d_t\right|\right)e^{-\beta'\min(d_t',d_t)} \tag{26}$$

in which $\alpha_i'(\bullet)$ is a nonnegative function satisfying $\alpha_i'(0)=0$ and $\lim_{t\to+\infty}\alpha_i'(t)$ exists.

Proof of Lemma 1: Following the arguments and the remarks in [19], we have

$$\begin{aligned}&\sum_{n=0}^{N}\left|S_{i,n}(d_t)-S_{i,n}\left(d_t'\right)\right| \\ &\quad\le\frac{1}{2}\left\|\left[S_{i,:}\left(\left|d_t-d_t'\right|\right)-S_{i,:}(0)\right]S\left(\min\left(d_t',d_t\right)\right)\right\|_{L^2(1/\pi)} \\ &\quad\le\frac{1}{2}\left\|S_{i,:}\left(\left|d_t-d_t'\right|\right)-S_{i,:}(0)\right\|_{L^2(1/\pi)}e^{-\beta'\min(d_t',d_t)}\end{aligned}$$

in which the definition of $\|\bullet\|_{L^2(1/\pi)}$ and the positive constant β' can be found in [19].

Denote $\alpha_i'(|d_t'-d_t|)=(1/2)\|S_{i,:}(|d_t-d_t'|)-S_{i,:}(0)\|_{L^2(1/\pi)}$. We have $\alpha_i'(0)=(1/2)\|S_{i,:}(0)-S_{i,:}(0)\|_{L^2(1/\pi)}=0$ and $\lim_{t\to+\infty}\alpha_i'(t)=(1/2)\|S_{i,:}(+\infty)-S_{i,:}(0)\|_{L^2(1/\pi)}=(1/2)\|\boldsymbol{\pi}-S_{i,:}(0)\|_{L^2(1/\pi)}$. ■

Proof of Theorem 1: It is easy to see that $\Delta R(d_t')\ge0$ because $\boldsymbol{P}(S_{i,:}(d_t))=\arg\max_{\boldsymbol{P}^{\mathrm{T}}\mathbf{1}\le P^{\max}}R(S_{i,:}(d_t),\boldsymbol{P})$.

To show the second inequality, we have

$$\begin{aligned}\Delta R\left(d_t'\right)&=\sum_{i=0}^{N}\pi_i\left[R(S_{i,:}(d_t),\boldsymbol{P}(S_{i,:}(d_t)))\right. \\ &\qquad\left.-R(S_{i,:}(d_t),\boldsymbol{P}(S_{i,:}\left(d_t'\right)))\right] \\ &=\sum_{i=0}^{N}\pi_i\left[\sum_{n=0}^{N}S_{i,n}(d_t)R_n(\boldsymbol{P}(S_{i,:}(d_t)))\right. \\ &\qquad\left.-\sum_{n=0}^{N}S_{i,n}(d_t)R_n(\boldsymbol{P}(S_{i,:}\left(d_t'\right)))\right] \\ &=\sum_{i=0}^{N}\pi_i\left[\sum_{n=0}^{N}S_{i,n}(d_t)R_n(\boldsymbol{P}(S_{i,:}(d_t)))\right. \\ &\qquad\left.-\sum_{n=0}^{N}S_{i,n}\left(d_t'\right)R_n(\boldsymbol{P}(S_{i,:}\left(d_t'\right)))\right] \\ &\quad+\sum_{i=0}^{N}\pi_i\sum_{n=0}^{N}R_n(\boldsymbol{P}(S_{i,:}(d_t')))(S_{i,n}\left(d_t'\right)-S_{i,n}(d_t))\end{aligned}$$

in which $R_i(\boldsymbol{P})$ represents the achievable rate of $\boldsymbol{P}$ when the number of active narrowband users is i. Applying the Cauchy–Schwarz inequality and Lemma 1, we derive again $\Delta R\left(d_t'\right)$, shown at the top of the next page.

Denote $\alpha(|d_t'-d_t|)=\sum_{i=0}^{N}\pi_i\sqrt{2\cdot\sum_{n=0}^{N}\max_{t>0}R_n(\boldsymbol{P}(S_{i,:}(t)))\cdot\alpha_i'(|d_t'-d_t|)}$ and $\beta=\beta'/2$. It is easy to verify that $\alpha(0)=0$ and $\lim_{t\to+\infty}\alpha(t)$ exists. ■

$$
\begin{aligned}
\Delta R\left(d_t'\right) \leq & \sum_{i=0}^{N} \pi_i \sum_{n=0}^{N} \max\left\{R_n\left(\boldsymbol{P}\left(S_{i,:}(d_t)\right)\right), R_n\left(\boldsymbol{P}\left(S_{i,:}\left(d_t'\right)\right)\right)\right\} \\
& \cdot \left|S_{i,n}(d_t) - S_{i,n}\left(d_t'\right)\right| + \sum_{i=0}^{N} \pi_i \sum_{n=0}^{N} R_n\left(\boldsymbol{P}\left(S_{i,:}\left(d_t'\right)\right)\right) \\
& \cdot \left|S_{i,n}\left(d_t'\right) - S_{i,n}(d_t)\right| \\
\leq & \sum_{i=0}^{N} \pi_i \sqrt{\sum_{n=0}^{N} \left|S_{i,n}\left(d_t'\right) - S_{i,n}(d_t)\right|} \\
& \cdot \sqrt{\sum_{n=0}^{N} \left(\max\left\{R_n\left(\boldsymbol{P}\left(S_{i,:}(d_t)\right)\right), R_n\left(\boldsymbol{P}\left(S_{i,:}\left(d_t'\right)\right)\right)\right\} + R_n\left(\boldsymbol{P}\left(S_{i,:}\left(d_t'\right)\right)\right)\right)} \\
\leq & e^{-\frac{\beta'}{2}\min\left(d_t', d_t\right)} \cdot \sum_{i=0}^{N} \pi_i \cdot \sqrt{\alpha_i'\left(\left|d_t' - d_t\right|\right)} \\
& \cdot \sqrt{\sum_{n=0}^{N} \left(\max\left\{R_n\left(\boldsymbol{P}\left(S_{i,:}(d_t)\right)\right), R_n\left(\boldsymbol{P}\left(S_{i,:}\left(d_t'\right)\right)\right)\right\} + R_n\left(\boldsymbol{P}\left(S_{i,:}\left(d_t'\right)\right)\right)\right)} \\
\leq & e^{-\frac{\beta'}{2}\min\left(d_t', d_t\right)} \cdot \sum_{i=0}^{N} \pi_i \sqrt{2 \cdot \sum_{n=0}^{N} \max_{t>0} R_n\left(\boldsymbol{P}\left(S_{i,:}(t)\right)\right) \cdot \alpha_i'\left(\left|d_t' - d_t\right|\right)}
\end{aligned}
$$

REFERENCES

[1] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.: Int. J. Comput. Telecommun. Netw.*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.

[3] Y. Xing, R. Chandramouli, S. Mangold, and S. Shankar, "Dynamic spectrum access in open spectrum wireless networks," *IEEE J. Sel. Areas Commun.—Special Issue on 4G Wireless Systems*, vol. 24, no. 3, pp. 626–637, Mar. 2006.

[4] E. Friedman and S. Shenker, *Learning and Implementation on the Internet*. New Brunswick, NJ: Dept. Economics, Rutgers Univ., 1997. Manuscript. [Online]. Available: http://citeseer.ist.psu.edu/eric98learning.html

[5] C. Pandana and K. J. R. Liu, "Near-optimal reinforcement learning framework for energy-aware wireless sensor communications," *IEEE J. Sel. Areas Commun.—Special Issue on Self-Organizing Distributed Collaborative Sensor Networks*, vol. 23, no. 4, pp. 788–797, Apr. 2005.

[6] C. Long, Q. Zhang, B. Li, H. Yang, and X. Guan, "Non-cooperative power control for wireless ad hoc networks with repeated games," *IEEE J. Sel. Areas Commun.—Special Issue on Non-Cooperative Behavior in Networking*, vol. 25, no. 6, pp. 1101–1112, Aug. 2007.

[7] F. Fu and M. van der Schaar, "Learning to compete for resources in wireless stochastic games," *IEEE Trans. Veh. Technol.*, to be published.

[8] X. Liu and W. Wang, "On the characteristics of spectrum-agile communication networks," in *Proc. IEEE Symp. DySPAN*, Nov. 2005, pp. 214–223.

[9] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 66–72, May 2007.

[10] R. G. Gallager, *Discrete Stochastic Processes*. New York: Springer-Verlag, 1995.

[11] O. Bousquet, S. Boucheron, and G. Lugosi, "Introduction to statistical learning theory," in *Advanced Lectures on Machine Learning Lecture Notes in Artificial Intelligence*, vol. 3176. New York: Springer-Verlag, 2004, pp. 169–207.

[12] I. Csiszár and P. C. Shields, "Information theory and statistics: A tutorial," *Commun. Inf. Theory*, vol. 1, no. 4, pp. 417–528, Dec. 2004.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

[14] J. B. Conway, *A Course in Functional Analysis*, 2nd ed. New York: Springer-Verlag, 1994.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[16] J. J. Leader, *Numerical Analysis and Scientific Computation*. Reading, MA: Addison-Wesley, 2004.

[17] M. Kazantzidis and M. Gerla, "End-to-end versus explicit feedback measurement in 802.11 networks," in *Proc. 7th ISCC*, 2002, pp. 429–434.

[18] D. Kliazovicha and F. Granelli, "Cross-layer congestion control in ad hoc wireless networks," *Ad Hoc Netw.*, vol. 4, no. 6, pp. 687–708, Nov. 2006.

[19] J. S. Rosethal, "Markov chain convergence: From finite to infinite," *Stoch. Process. Appl.*, vol. 62, no. 1, pp. 55–72, 1996.

[20] D. Ĕabriá and R. Brodersen, "Physical layer design issues unique to cognitive radio systems," in *Proc. 16th IEEE Int. Symp. PIMRC*, Sep. 2005, pp. 759–763.

[21] A. Sahai, R. Tandra, M. Mishra, and N. Hoven, "Fundamental design tradeoffs in cognitive radio systems," in *Proc. TAPAS*, Article 2, Aug. 2006.

[22] A. Goldsmith and P. Varaiya, "Capacity of fading channel with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[23] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

Yi Su (S'08) received the B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2004 and 2006, respectively. He is currently working toward the Ph.D. degree in the Electrical Engineering Department, University of California, Los Angeles.

Mihaela van der Schaar (SM'04) received the Ph.D. degree from Eindhoven University of Technology, Eindhoven, The Netherlands, in 2001.

She is currently an Associate Professor with the Electrical Engineering Department, University of California, Los Angeles.