

# Learning for Cognitive Wireless Users

Yi Su and Mihaela van der Schaar

Dept. of Electrical Engineering, UCLA  
{yisu, mihaela}@ee.ucla.edu

**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 cognitive user which learns the stationary usage pattern of the spectrum occupied by narrowband users and, based on this information adapt its transmission. Specifically, we investigate the trade-off between the learning duration and the achievable performance in stationary environments. We apply optimization and large deviations theory to analytically derive an upper bound of the minimum required learning duration, given the user’s tolerable performance loss and outage probability. Furthermore, noticing that learning techniques require the information feedback of the spectrum usage pattern between the transceivers, we investigate how the cognitive user can further improve its performance by taking account of its feedback delay. The impact of inaccurate delay estimation on the achievable performance is also quantified.

**Keywords**- learning, feedback delay, cognitive wireless users

## I. INTRODUCTION

A promising way of improving the radio spectrum utilization is to build cognitive wireless devices 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 communications systems. They enable the dynamically interacting communications devices to acquire information, build knowledge, and ultimately improve their performance [4]-[7]. 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 to characterize and analytically quantify the achievable performance which 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. time-varying channel condition or interference, in order to reach its performance 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 deviations 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 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 in which cognitive communication devices deploy strategic learning solutions to accumulate knowledge about its environment and improve its performance [5][7]. The rest of the 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 shows the numerical results. Section V quantifies the impact of spectrum usage information feedback delay. Conclusions are drawn in Section VI.

## II. SYSTEM MODEL

### A. System Description

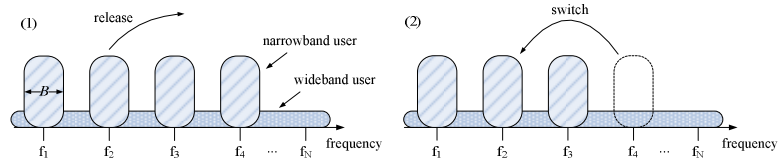


Fig. 1. Investigated cognitive wireless networks.

We assume a cognitive wireless system 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 given 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 rate 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]

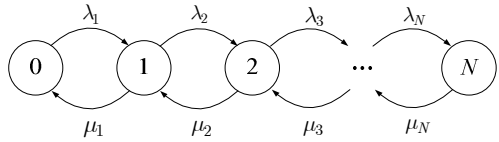
$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 & & & \\ \mu_1 & -(\lambda_2 + \mu_1) & \lambda_2 & & \\ & \mu_2 & -(\lambda_3 + \mu_2) & \ddots & \\ & & \ddots & \ddots & \lambda_N \\ & & & \mu_N & -\mu_N \end{bmatrix}. \quad (1)$$


Fig. 2. An example of continuous time Markov chain model.

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  $\pi = [\pi_0, \pi_1, \dots, \pi_N]$ , in which  $\pi_i$  represents the probability of having active  $i$  narrowband devices in the system. No matter what form the infinitesimal generator  $Q$  takes, we always have

$$\pi Q = \mathbf{0}. \quad (2)$$

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  $h_i$ . 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  $\mathbf{P} = [P_1, \dots, P_N]^T$ , in which  $P_i$  is the power allocated in frequency band  $i$ . Hence, the achievable rate is given by

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

### B. Learning Duration and Performance

As mentioned previously, 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 in which 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 devices it encountered in the past and approximating the stationary spectrum usage pattern  $\pi$  by the observed frequencies of the system states. We define an empirical frequency function

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

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

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

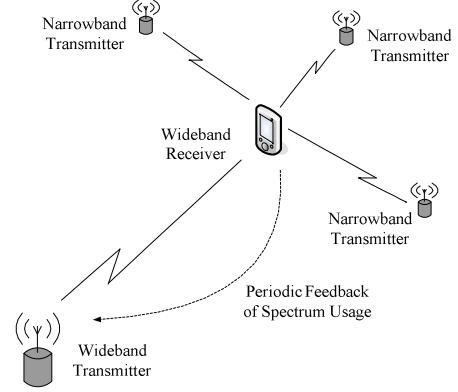


Fig. 3. Spectrum usage feedback of the wideband device.

The wideband user approximates the steady state  $\pi$  using the empirical frequency function  $\gamma^t$ <sup>1</sup>, and takes the best response action  $\mathbf{P}(\gamma^t)$  that maximizes  $R(\gamma^t, \mathbf{P})$ , i.e.  $\mathbf{P}(\gamma^t) = \arg \max_{\mathbf{P} \in \mathcal{P}^{\max}} R(\gamma^t, \mathbf{P})$  with  $\mathbf{1} = [1, \dots, 1]^T$ . Denote the achievable rate when the wideband user takes the best response to the empirical frequency function  $\gamma^t$  as  $R_a(\gamma^t) = R(\pi, \mathbf{P}(\gamma^t))$ .

Throughout this paper, the 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  $\pi$ . Intuitively, the performance of learning is expected to improve if more observations are available. This paper aims to determine how many observations are sufficient for a learning user to reach a certain desirable performance guarantee. Specifically, given the tolerable performance loss  $\Delta_R$  with respect to perfectly knowing  $\pi$  and the outage probability  $\delta_R$ , we want to determine the minimum required learning duration:

$$\begin{aligned} & \min t \\ & \text{s.t. } \text{Prob}(R_a(\pi) - R_a(\gamma^t) \geq \Delta_R) \leq \delta_R. \end{aligned} \quad (6)$$

### III. MINIMUM REQUIRED LEARNING DURATION

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

<sup>1</sup> 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.

for the solution of the problem in (6). Having such a bound is important from both theoretical and practical perspectives, because, due to the real-time adaptation requirement of cognitive networks [1], only limited observations are usually available to cognitive users and it is also necessary for them to understand the basic trade-off of performance vs. learning duration. For this, we adopt tools from large deviations theory, which quantifies the exponential decay of probability measures for certain kinds of tail events [12]. According to the large deviations theory, the empirical frequency function  $\gamma^t(n)$  of a random sample of size  $t$  drawn from  $\pi$  satisfies

$$\text{Prob}(D(\gamma^t \parallel \pi) \geq \delta) \leq \binom{N+t}{N} 2^{-\delta t}, \quad \forall \delta > 0, \quad (7)$$

where  $D(p \parallel q)$  is the Kullback-Leibler (KL) distance between two pmfs  $p(x)$  and  $q(x)$  [13]. Then, we need to convert the performance loss  $R_a(\pi) - R_a(\gamma^t)$  into the KL distance  $D(\gamma^t \parallel \pi)$ . 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  $\delta$  such that  $D(\gamma^t \parallel \pi) \leq \delta$  always leads to  $R_a(\pi) - R_a(\gamma^t) \leq \Delta_R$ . By setting  $t$  to satisfy  $\binom{N+t}{N} 2^{-\delta t} \leq \delta_R$ , we have  $\text{Prob}(D(\gamma^t \parallel \pi) \geq \delta) \leq \text{Prob}(R_a(\pi) - R_a(\gamma^t) \geq \Delta_R)$  and this value provides an upper bound for the problem in (6). As illustrated in Fig. 4, we divide this procedure into three steps. First, we construct a convex set  $\mathcal{B}$  in the standard probability simplex  $\Omega = \{\gamma \mid \mathbf{1}^T \gamma = 1, \gamma \succeq 0\}$  such that, for all  $\gamma \in \mathcal{B}$ , it satisfies  $R_a(\pi) - R_a(\gamma) \leq \Delta_R$ . Second, by solving convex optimization problems that minimize the KL distance between  $\pi$  and the pmfs that lie on the boundary of  $\mathcal{B}$ , we obtain the desired value of  $\delta$ , which is denoted as  $\delta_{D_{\min}}$  in Fig. 4. Third, we apply large deviations theory and derive an upper bound of the minimum required observations. In the following subsections, we will explain each step in detail.

#### A. Extreme Points with Performance Loss Constraints

First, in the probability simplex  $\Omega$ , we construct a convex set  $\mathcal{B}$  that contains the actual pmf. Let  $A = \{\{k, j\} : k, j \in \{0, 1, \dots, N\} \text{ and } k < j\}$  such that  $A$  contains a total number of  $M = \binom{N+1}{2}$  combinations of any two different integers in  $\{0, 1, 2, \dots, N\}$ . Let  $(S)_m$  denote the  $m$ th element of set  $S$ . Based on the tolerable performance loss  $\Delta_R$ , 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  $m = 1, 2, \dots, 2M$ , the  $2M$  pmfs that we are interested in satisfy:

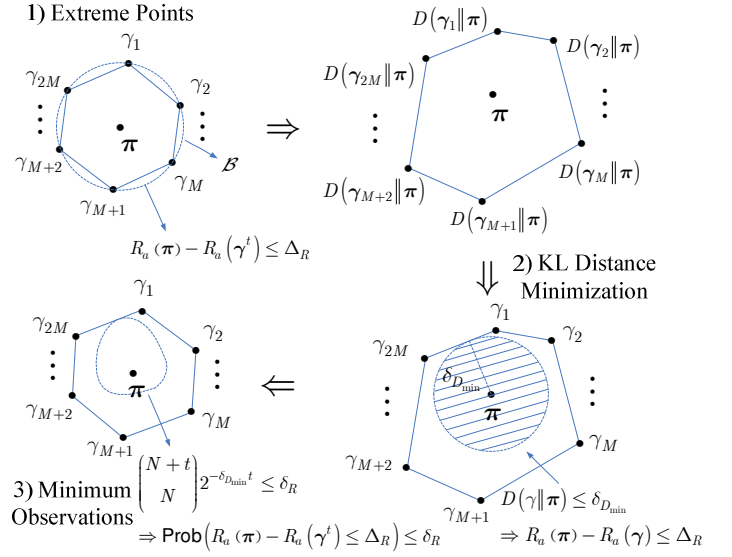


Fig. 4. Performance loss and KL distance.

(P1)  $\gamma_m \in \Omega$ ;

(P2)  $\gamma_m(n) = \pi_n$ , if  $n \notin (A)_m$ .

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

$$\gamma_m(n, \delta_m) = \begin{cases} \pi_n - \delta_m, & \text{if } n = ((A)_m)_1 \\ \pi_n + \delta_m, & \text{if } n = ((A)_m)_2, m = 1, 2, \dots, 2M. \end{cases} \quad (8)$$

Denoting  $\gamma_m(\delta_m) = [\gamma_m(0, \delta_m), \dots, \gamma_m(N, \delta_m)]$ , now we can determine the extreme points by setting the parameter  $\delta_m$  based on the tolerable performance loss  $\Delta_R$ . For  $m = 1, 2, \dots, M$ ,

$$\delta_m = \begin{cases} \pi_l \text{ with } l = ((A)_m)_1, & \text{if } S_\delta = \emptyset \\ \min \delta \in S_\delta, & \text{otherwise} \end{cases} \quad (9)$$

in which  $S_\delta = \{\delta : R_a(\pi) - R_a(\gamma_m(\delta)) \geq \Delta_R \text{ and } \delta \geq 0\}$ , and

$$\delta_{m+M} = \begin{cases} -\pi_l \text{ with } l = ((A)_m)_2, & \text{if } S_{-\delta} = \emptyset \\ \min \delta \in S_{-\delta}, & \text{otherwise} \end{cases} \quad (10)$$

in which  $S_{-\delta} = \{\delta : R_a(\pi) - R_a(\gamma_m(-\delta)) \geq \Delta_R \text{ and } \delta \geq 0\}$ .

Due to the non-negative property in (P1), when  $n \in (A)_m$ , if  $S_\delta = \emptyset$  or  $S_{-\delta} = \emptyset$ , we set  $\gamma_m(n, \delta_m)$  to be zero to ensure the performance loss is as close to  $\Delta_R$  as possible. On the

other hand, if  $S_\delta \neq \emptyset$  or  $S_{-\delta} \neq \emptyset$ , the “extreme points” are the pmfs that cause an exact performance loss of  $\Delta_R$ .

Using the convex hull of the above  $2M$  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

$$\mathcal{B} = \left\{ \gamma : \gamma = \sum_{m=1}^{2M} \alpha_m \gamma_m(\delta_m), \alpha_m \geq 0, \text{ and } \sum_{m=1}^{2M} \alpha_m = 1 \right\}. \quad (11)$$

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

The proof is given in [20]. 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  $\Delta_R$  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  $\Delta_R$ . Next, we apply large deviations theory to translate the performance loss  $\Delta_R$  into another metric, the KL distance  $\delta_D$ . The basic idea is to solve an optimization problem to find the minimum KL distance  $\delta_{D_{\min}}$  such that, for any  $\gamma$  that satisfies  $D(\gamma \parallel \pi) \leq \delta_{D_{\min}}$ , we have  $R_a(\pi) - R_a(\gamma) \leq \Delta_R$ . Particularly, the optimization problem can be formulated as

$$\begin{aligned} & \min_{\gamma} D(\gamma \parallel \pi) \\ & \text{s.t. } \gamma \in \mathcal{S}(\mathcal{B}), \end{aligned} \quad (12)$$

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  $\text{int}(\mathcal{B})$  [14].

Note that the KL distance  $D(\gamma \parallel \pi)$  is convex in the pair  $(\gamma, \pi)$ , and  $\gamma \in \mathcal{S}(\mathcal{B})$  is a linear constraint [13]. Therefore, the problem in (12) essentially belongs to convex programming, and the optimal solution can be obtained efficiently 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  $\pi$ , the minimum of (12) that ensures  $D(\gamma \parallel \pi) \leq \delta_{D_{\min}}$  is a sufficient condition to ensure that  $R_a(\pi) - R_a(\gamma) \leq \Delta_R$ .

### C. Minimum Learning Duration Calculation

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

$$\begin{aligned} & \min t \\ & \text{s.t. } \text{Prob}\left(D(\gamma^t \parallel \pi) \geq \delta_{D_{\min}}\right) \leq \delta_R. \end{aligned} \quad (13)$$

Applying formula (7) from large deviations 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  $\gamma^t$  and takes the best-response action with respect to  $\gamma^t$ . An upper bound  $T$  of the solution of problem (6) is

$$T = \text{Min}_- t(\delta_{D_{\min}}, N, \delta_R), \quad (14)$$

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

**Proof:** Combining (7) and (13), we know that any  $t$  that satisfies

$$\binom{N+t}{N} 2^{-\delta_{D_{\min}} t} \leq \delta_R \quad (15)$$

is an upper bound of the solution of problem (6). Let  $F(t) = \binom{N+t}{N} 2^{-\delta_{D_{\min}} t}$ . We have  $\frac{F(t+1)}{F(t)} = \left(1 + \frac{N}{t+1}\right) 2^{-\delta_{D_{\min}}}$  and  $\lim_{t \rightarrow \infty} \frac{F(t+1)}{F(t)} = 2^{-\delta_{D_{\min}}} < 1$ . Therefore, we can conclude that  $\lim_{t \rightarrow \infty} F(t) = 0$ . As a result, by choosing  $T = \text{Min}_- t(\delta_{D_{\min}}, N, \delta_R)$  as the minimum integer in the feasible region of inequality (15), we obtain an upper bound of the optimum solution of (6). ■

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

$$T = \text{Min}_- t(f(\Delta_R), N, \delta_R). \quad (16)$$

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

**Remark 1 :** Decreasing the acceptable performance loss  $\Delta_R$  will lead to a larger minimum observation duration  $T$ , which is a direct consequence of the non-increasing property of function  $f$ .

**Remark 2 :** Decreasing the outage probability  $\delta_R$  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 the outage probability is smaller than the threshold of  $\delta_R$ . This argument holds because a larger number

of channels  $N$  will cause  $\binom{N+t}{N}$  increases and  $\delta_{D_{\min}}$  to be smaller than or equal to its original value (given the steady state probability distribution  $\pi$  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 with  $N = 2, \lambda_1 = \mu_2 = 2, \lambda_2 = \mu_1 = 1$  and the power constraint of the wideband device is  $P^{\max} = 1$ . Its channel conditions and the power of noise and interference are given by  $h_1 = 2, h_2 = 1, N_1 = N_2 = I = 1$ . It is easy to solve that the stationary distribution is  $\pi = [0.25 \ 0.5 \ 0.25]$ . The contours of achievable rate  $R_a(\gamma)$  and KL distance  $D(\gamma \| g)$  are shown in Fig. 5 and 6 separately. We can see from Fig. 5 that  $R_a(\pi) - R_a(\gamma) \leq \Delta_R$  is a convex set, which verifies proposition 1.

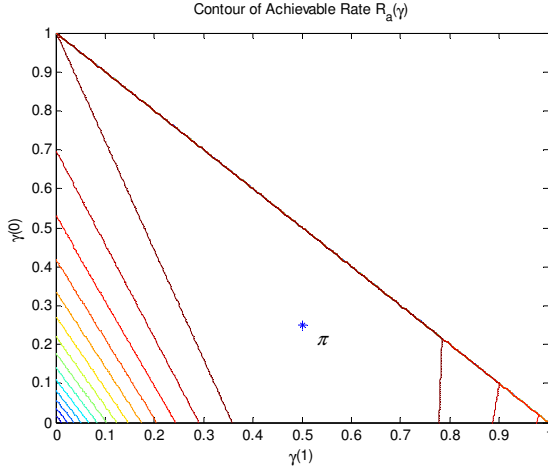


Fig. 5. Contour of achievable rate  $R_a(\gamma)$ .

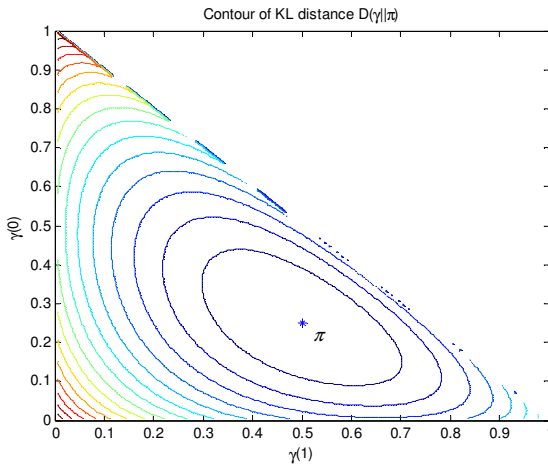


Fig. 6. Contour of KL distance  $D(\gamma \| \pi)$ .

We set the parameters in the problem (6) to be  $\Delta_R = 10^{-2.5}$  and  $\delta_R = 10^{-2}$ . Fig. 7 and 8 illustrate the procedure of obtaining the upper bound in Section III.B. Noting that  $N = 2$ , we choose six extreme points  $\gamma_1, \dots, \gamma_6$  in total, which are determined based on the rate-pmf curves in Fig. 7. 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  $\gamma_1, \dots, \gamma_6$  is the extreme point set  $\mathcal{B}$ . The dashed hexagon in Fig. 8 is the surface  $\mathcal{S}(\mathcal{B})$  on which we minimize the KL distance. Solving the convex optimization problem (12) leads to  $\delta_{D_{\min}} = 0.1265$ . Using (14), we obtain that  $T = \text{Min}_t(0.1265, 2, 10^{-2}) = 161$ . As shown in Fig. 8, if the learning duration is larger than  $T$ , the KL distance between the actual stationary distribution  $\pi$  and observed empirical frequency  $\gamma^t$  will lie within the solid circle with an outage probability less than  $\delta_R$ .

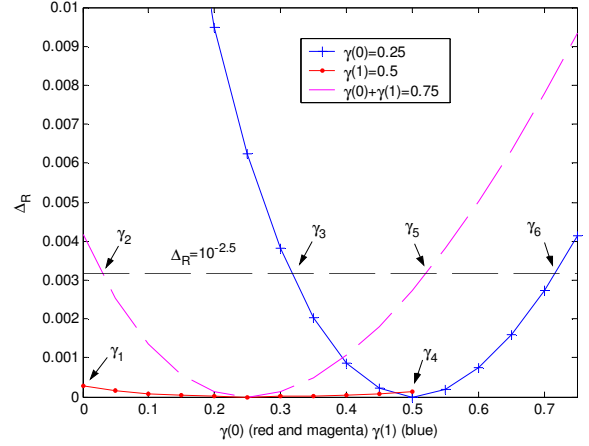


Fig. 7. Constructing the extreme points.

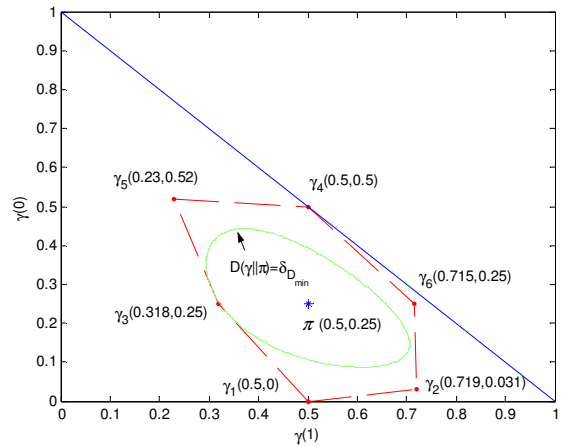


Fig. 8. KL distance minimization in  $\mathcal{S}(\mathcal{B})$ .

We also examine the tightness of the upper bound in different settings. The tolerable performance loss  $\Delta_R$  is varied to be  $10^{-2}$ ,  $10^{-2.5}$ , and  $10^{-3}$ , while the outage probability  $\delta_R$  is set to be a constant of  $10^{-2}$ . In each scenario, we use Monte Carlo method to calculate the actual required learning duration  $T_a$ . The results are summarized in Table I. 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 dashed hexagon is large. At the same time, we also find that the ratio of  $T/T_a$  increases when the performance loss  $\Delta_R$  is decreased, because the mismatch between the contours is also increased as  $\Delta_R$  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  $R_a(\gamma)$  in Fig. 5 and therefore, improve the tightness of the upper bound.

| Performance loss $\Delta_R$ | $10^{-2}$ | $10^{-2.5}$ | $10^{-3}$ |
|-----------------------------|-----------|-------------|-----------|
| Actual value $T_a$          | 38        | 50          | 62        |
| Upper bound $T$             | 101       | 161         | 593       |
| $T/T_a$                     | 2.7       | 3.2         | 9.6       |

Table I. Learning durations for different performance loss requirements

## 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 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.

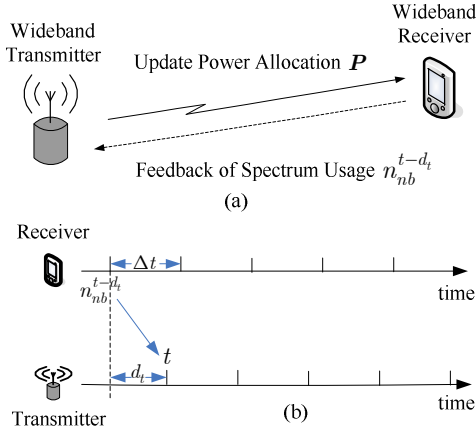


Fig. 9. Feedback delay of the spectrum usage.

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. 9, 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  $\mathbf{S}(t)$  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  $\mathbf{S}(t)$  is the solution of the Kolmogorov equation, which takes the form of

$$\mathbf{S}(t) = \sum_{i=1}^{N+1} \mathbf{v}_i e^{\xi_i t} \boldsymbol{\omega}_i, \quad (17)$$

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

$$\max_{\mathbf{P}^t, \mathbf{1} \leq P \leq P_{\max}} R(\boldsymbol{\pi}^t, \mathbf{P} | n_{nb}^{t-d_t}), \quad (18)$$

in which  $\boldsymbol{\pi}^t = [\pi_0^t, \pi_1^t, \dots, \pi_N^t]$  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 (17), we have

$$\lim_{t \rightarrow +\infty} \mathbf{S}(t) = \mathbf{v}_1 \boldsymbol{\omega}_1. \quad (19)$$

Therefore, when  $d_t \rightarrow +\infty$ ,  $\boldsymbol{\pi}^t \rightarrow \boldsymbol{\omega}_1 = \boldsymbol{\pi}$ , which is independent of  $n_{nb}^{t-d_t}$ . As a result,  $R(\boldsymbol{\pi}^t, \mathbf{P} | n_{nb}^{t-d_t})$  in (18) is reduced to  $R(\boldsymbol{\pi}, \mathbf{P})$  in equation (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 is optimal only when the feedback delay is large.

On the other hand, we also consider the limited feedback delay scenarios. Note that in these cases, 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 (18) is a function of the limited feedback delay  $d_t$ . In particular,  $R(\boldsymbol{\pi}^t, \mathbf{P} | n_{nb}^{t-d_t})$  in the optimal transmission strategy of (18) will become:

$$\begin{aligned} R(\boldsymbol{\pi}^t, \mathbf{P} | n_{nb}^{t-d_t}) &= R(S_{n_{nb}^{t-d_t}, \cdot}(d_t), \mathbf{P}) \\ &= \sum_{i=1}^N \left( \sum_{n \geq i} S_{n_{nb}^{t-d_t}, n}(d_t) B \log \left( 1 + \frac{h_i P_i}{N_i + I} \right) + \right. \\ &\quad \left. \sum_{n=0}^{n < i} S_{n_{nb}^{t-d_t}, n}(d_t) B \log \left( 1 + \frac{h_i P_i}{N_i} \right) \right) \end{aligned} \quad (20)$$

where  $S_{i,\cdot}(d_t)$  represents the  $i$ th row of  $\mathbf{S}(d_t)$ . The problem is converted into how to accurately estimate  $\mathbf{S}(t)$  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  $\mathbf{S}(t)$  at  $t = \Delta t, 2\Delta t, \dots$  by updating empirical frequency functions, and use numerical algorithms [16], such as curve fitting, to estimate  $\mathbf{S}(t)$  for non-integer multiples of  $\Delta t$ . As long as the environment is stationary and the sampling data is large enough, the wideband device can estimate  $\mathbf{S}(d_t)$  accurately.

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][18]. Suppose 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

$$\Delta R(d'_t) = \sum_{i=0}^N \pi_i \left[ R(S_{i,:}(d_t), \mathbf{P}(S_{i,:}(d_t))) - R(S_{i,:}(d_t), \mathbf{P}(S_{i,:}(d'_t))) \right]. \quad (21)$$

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 (21) depends on two terms  $|d'_t - d_t|$  and  $\min(d'_t, d_t)$ . Specifically,  $\Delta R(d'_t)$  is bounded as

$$0 \leq \Delta R(d'_t) \leq \alpha(|d'_t - d_t|) e^{-\beta \min(d'_t, d_t)}, \quad (22)$$

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

The proof is given in [20]. Two key observations can be made from the above 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. Besides, the existence of  $\lim_{t \rightarrow +\infty} \alpha(t)$  implies that 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  $N = 2, \lambda_1 = \mu_2 = 0.02$ , and  $\lambda_2 = \mu_1 = 0.01$ . It is easy to show that, for, three eigenvalues of  $Q$  are  $\xi_1 = 0, \xi_2 = -0.02, \xi_3 = -0.04$ , and the transition probability matrix  $\mathbf{S}(t)$  is given by

$$\mathbf{S}(t) = 0.25 \times \begin{bmatrix} 1 + 2e^{-0.02t} + 2e^{-0.04t} & 2 - 2e^{-0.04t} & 1 - 2e^{-0.02t} + e^{-0.04t} \\ 1 - e^{-0.04t} & 2 + 2e^{-0.04t} & 1 - e^{-0.04t} \\ 1 - 2e^{-0.02t} + e^{-0.04t} & 2 - 2e^{-0.04t} & 1 + 2e^{-0.02t} + e^{-0.04t} \end{bmatrix}. \quad (23)$$

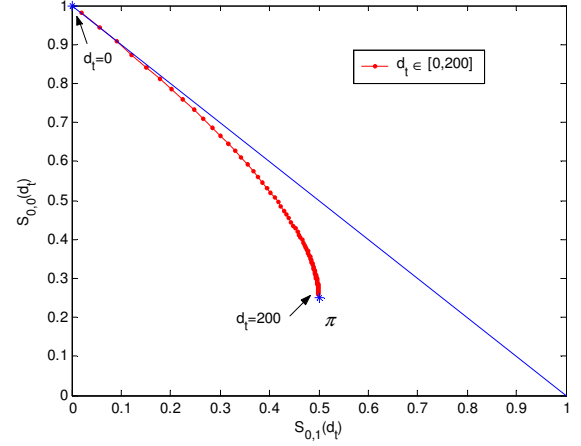


Fig. 10. Transition probability of  $S_{0,0}(d_t)$  and  $S_{0,1}(d_t)$ .

The transition probability  $S_{0,0}$  and  $S_{0,1}$  is plotted as a function of the feedback delay  $d_t$  in Fig. 10. As we expect, if the feedback delay  $d_t \rightarrow 0$ , the spectrum usage pattern  $n_{nb}^t$  has a large possibility to be equal to  $n_{nb}^{t-d_t}$ , i.e. the transmitter knows exactly how many narrowband users are currently active. If  $d_t \rightarrow +\infty$ , the spectrum usage pattern will converge to the stationary distribution  $\pi$ . Therefore, if  $d_t$  is not sufficient large, the wideband transmitter should optimize its power allocation with respect to the transition probability matrix  $\mathbf{S}(d_t)$  rather than the stationary distribution  $\pi$ .

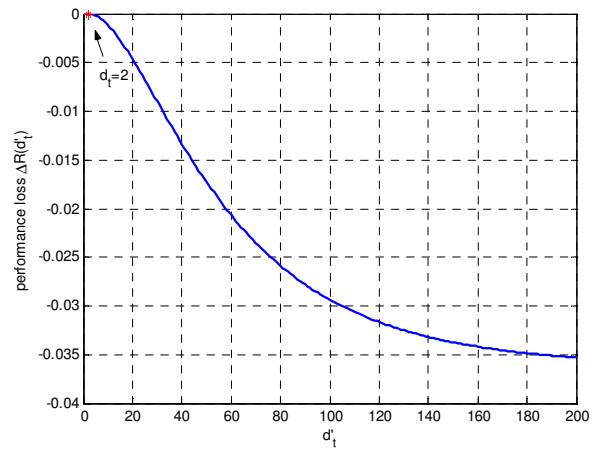


Fig. 11. Performance loss of inaccurate estimate over  $d_t$ .

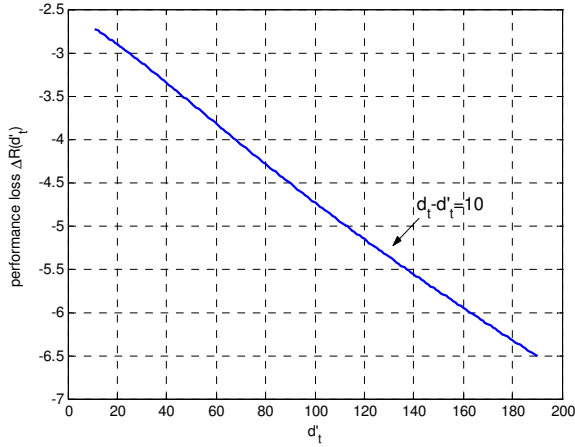


Fig. 12. Performance loss of inaccurate estimate for fixed  $d_t - d_t'$ .

Next, we numerically show the improvement of measuring the feedback delay  $d_t$ . The feedback delay  $d_t$  is assumed to be 2, and the performance loss  $\Delta R(d_t')$  is plotted in Fig. 11. We can see that it agrees with the argument that  $\alpha(0) = 0$  and  $\lim_{t \rightarrow +\infty} \alpha(t)$  exists in Theorem 1. Compared with taking 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 10, and plot the corresponding  $\Delta R(d_t')$  in Fig. 12. We can see that the performance loss  $\Delta R(d_t')$  decrease exponentially with  $d_t'$ , which complies, as expected, with Theorem 1.

## VI. CONCLUSIONS

This paper studies the minimum required observations 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 trade-off 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 communications protocols with learning capabilities such that engineers can build practical systems with desired complexity and performance trade-off.

## REFERENCES

- [1] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE JSAC*, vol. 23, 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", *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 50, pp. 2127-2159, Sep. 2006
- [3] Y. Xing, R. Chandramouli, S. Mangold and S. Shankar, "Dynamic Spectrum Access in Open Spectrum Wireless Networks," *IEEE JSAC Special issue on 4G Wireless Systems*, vol. 24, pp. 626-637, Mar. 2006.
- [4] E. Friedman, and S. Shenker. "Learning and Implementation on the Internet." Manuscript. New Brunswick: Rutgers University, Department of Economics, 1997. <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 JSAC 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 JSAC*, vol. 25, pp. 1101-1112, Aug. 2007
- [7] F. Fu and M. van der Schaar, "Dynamic Spectrum Sharing Using Learning for Delay-Sensitive Applications," *Proc. ICC 2008*, to appear
- [8] X. R. Zhu, L.F. Shen and T. S. Yum, "Analysis of Cognitive Radio Spectrum Access with Optimal Channel Reservation," *IEEE Communications Letter*, vol. 11, No. 4, pp. 1-3, April, 2007
- [9] X. Liu and W. Wang, "On the Characteristics of Spectrum-Agile Communication Networks", *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 214-223, Nov. 2005.
- [10] R.G. Gallager, *Discrete Stochastic Processes*, Springer, 1995
- [11] O. Bousquet, S. Boucheron, and G. Lugosi, "Introduction to Statistical Learning Theory", *Advanced Lectures on Machine Learning Lecture Notes in Artificial Intelligence*, vol. 3176, pp. 169-207. Springer, 2004
- [12] I. Csiszár and P. C. Shields, "Information theory and statistics: a tutorial," *Comm. and Inform. Theory*, vol.1, Issue.4, pp. 417-528, 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 edition, Springer-Verlag, 1994.
- [15] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [16] J. J. Leader, *Numerical Analysis and Scientific Computation*, Addison Wesley, 2004
- [17] M. Kazantzidis and M. Gerla, "End-to-end versus Explicit Feedback Measurement in 802.11 Networks", *Proc. ISCC*, pp. 429-434, 2002.
- [18] D. Kliazovitcha and F. Granelli, "Cross-layer congestion control in ad hoc wireless networks", *Ad Hoc Networks*, vol. 4, pp. 687-708, 2006
- [19] J. S. Rosenthal, "Markov chain convergence: From finite to infinite," *Stochastic Processes and their Applications*, vol. 62, pp.55-72, 1996
- [20] Y. Su and Mihaela van der Schaar, "Minimum Required Learning and Impact of Information Feedback Delay for Cognitive Users," UCLA Technical Report, 2008