Learning for Cognitive Wireless Users

Short Paper

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 quantified.**

Keywords- learning, feedback delay, cognitive wireless users

I. Introduction

A promising way of improving 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]. While conceptually simple, the realization of cognitive wireless devices is highly challenging. Several problems must be solved: sensing over wide frequency bands; identifying available spectrum opportunities; exploiting the identified transmission opportunities etc. In particular, 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"[1].

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 [3]-[5]. As opposed to the previous works, which focus on studying either the long-term convergence behavior of certain learning algorithms [3][4] or determine the operational shorter-term performance without providing any performance guarantees [5], this paper aims to characterize and analytically quantify the achievable performance that can be obtained by cognitive wireless users with learning capabilities. We study how much a cognitive device with no prior knowledge should learn about its environment, e.g. time-varying interference, 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 gathered through the information feedback from the receiver to the transmitter and this information can be delayed during this process, we study how a cognitive device can improve its performance if the feedback delay is accurately known. We also quantify the impact of imperfect delay measurements on the achieved performance.

The rest of the paper is organized as follows. Section II presents the system model and formulates the problem of learning and adapting to the spectrum usage pattern. Section III analytically derives 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

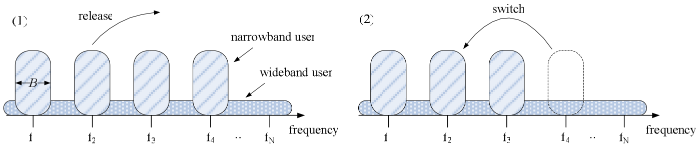

Fig. 1. Investigated cognitive wireless networks.

We assume a cognitive wireless system similar with the one studied in [2] (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 [2]. An example is given in Fig. 1. If one device releases f_2, the device occupying f_4 will switch to f_2. The system state is defined as the number of occupied frequency channels n_{nb}. Narrowband radios enter and exit the system independently following Poisson distributions. The spectrum usage pattern can be captured as a continuous time Markov chain with infinitesimal generator [2][6][7]

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

The Markov chain model and its corresponding infinitesimal generator Q can take various forms based on the configuration

This work was sponsored by NSF CAREER award no. 0541867 and an ONR grant.

978-1-4244-2017-9/08/$25.00 ©2008 IEEE

of the considered wireless network. Denote the steady state probability vector of the spectrum usage pattern as $\boldsymbol{\pi} = [\pi_0, \pi_1, \cdots, \pi_N]$, in which π_i represents the probability of having active i narrowband devices in the system. No matter what form the infinitesimal generator Q takes, we always have $\boldsymbol{\pi} Q = \mathbf{0}$. 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 $\boldsymbol{P} = [P_1, \cdots, P_N]^{\mathrm{T}}$, in which P_i is the power allocated in frequency band i. Hence, the achievable rate is given by

$$R(\boldsymbol{\pi}, \boldsymbol{P}) = \sum_{i=1}^{N} \left(\sum_{n \geq i}^{N} \pi_n B \log\left(1 + \frac{h_i P_i}{N_i + I}\right) + \sum_{n=0}^{n<i} \pi_n B \log\left(1 + \frac{h_i P_i}{N_i}\right) \right). \quad (2)$$

B. Learning Duration and Performance

Fig. 2 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 $\boldsymbol{\pi}$ by the observed frequencies of the system states. We define an empirical frequency function

$$\gamma^t(n) = c^t(n) \Big/ \sum\nolimits_{n=0}^{N} c^t(n), \quad (3)$$

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

$$c^t(n) = \begin{cases} c^{t-1}(n) + 1, \, if \; n_{nb}^{t} = n \\ c^{t-1}(n) \qquad , \, otherwise. \end{cases} \quad (4)$$

The wideband user approximates the steady state $\boldsymbol{\pi}$ using the empirical frequency function $\boldsymbol{\gamma}^t$, and takes the best response action $\boldsymbol{P}(\boldsymbol{\gamma}^t)$ that maximizes $R(\boldsymbol{\gamma}^t, \boldsymbol{P})$, 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 $\mathbf{1} = [1, \cdots, 1]^{\mathrm{T}}$. Denote the achievable rate when the wideband user takes the best response to the empirical frequency function $\boldsymbol{\gamma}^t$ as $R_a(\boldsymbol{\gamma}^t) = R(\boldsymbol{\pi}, \boldsymbol{P}(\boldsymbol{\gamma}^t))$.

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 $\boldsymbol{\pi}$. 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 Δ_R with respect to perfectly knowing $\boldsymbol{\pi}$ and the outage probability δ_R, we want to determine the minimum required learning duration:

$$\min t \quad s.t. \, \mathsf{Prob}\left(R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}^t) \geq \Delta_R\right) \leq \delta_R. \quad (5)$$

Understanding this problem 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.

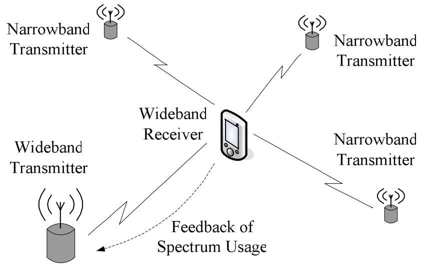

Fig. 2. Spectrum usage feedback of the wideband device.

III. Minimum Required Learning Duration

Although similar bounds exist in statistical learning theory, e.g. Hoeffding's inequality [8], it is still difficult to solve the problem in (5) 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 (5). For this, we adopt tools from large deviations theory [9]. According to the large deviations theory, the empirical frequency function $\gamma^t(n)$ of a random sample of size t drawn from $\boldsymbol{\pi}$ satisfies

$$\mathsf{Prob}\left(D\left(\boldsymbol{\gamma}^t \,||\, \boldsymbol{\pi}\right) \geq \delta\right) \leq \binom{N+t}{N} 2^{-\delta t}, \, \forall \delta > 0, \quad (6)$$

where $D(p \,||\, q)$ is the Kullback-Leibler (KL) distance between $p(x)$ and $q(x)$. 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 $\mathsf{Prob}\left(D\left(\boldsymbol{\gamma}^t \,||\, \boldsymbol{\pi}\right) \geq \delta\right) \geq \mathsf{Prob}\left(R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}^t) \geq \Delta_R\right)$ and this value provides an upper bound for the problem in (5). The whole procedure is divided into the following three steps.

A. Extreme Points with Performance Loss Constraints

First, in the probability simplex $\Omega = \left\{\boldsymbol{\gamma} \,|\, \mathbf{1}^{\mathrm{T}} \boldsymbol{\gamma} = 1, \boldsymbol{\gamma} \succeq 0\right\}$, we construct a convex set $\mathcal{B}$ that contains the actual pmf $\boldsymbol{\pi}$. Let $\mathrm{A} = \left\{\{k, j\} : k, j \in \{0, 1, \cdots, N\} \; and \; k < j\right\}$ which contains a total number of $M = \binom{N+1}{2}$ combinations of any two different integers in $\{0, 1, 2, \cdots, N\}$. Let $(S)_m$ denote the m th element of set S. Based on the tolerable performance loss Δ_R, we choose $2M$ pmfs and view them as "extreme points" of the set $\mathcal{B}$ in which we will derive an upper bound of the minimum required learning duration. For $m = 1, 2, \cdots, 2M$, the $2M$ pmfs that we are interested in satisfy:

[1] Note that here we normalize the feedback period, and Section II and III implicitly assume that this period is sufficiently large such that adjacent samples of spectrum usage pattern is independent of each other. Section V discusses the optimal adaptation strategies for various feedback delays and sampling intervals.

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

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

$$\gamma_m(n,\delta_m) = \begin{cases} \pi_n - \delta_m,\ if\ n = \left((\mathrm{A})_m\right)_1 \\ \pi_n + \delta_m,\ if\ n = \left((\mathrm{A})_m\right)_2,\ m = 1,2,\cdots,2M \\ \pi_n, \qquad if\ n \notin (\mathrm{A})_m \end{cases}. \tag{7}$$

Denote $\boldsymbol{\gamma}_m(\delta_m) = \left[\gamma_m(0,\delta_m),\cdots,\gamma_m(N,\delta_m)\right]$. We can choose the extreme points by setting the parameter δ_m based on the tolerable performance loss Δ_R. For $m = 1,2,\cdots,M$,

$$\delta_m = \begin{cases} \pi_l\ with\ l = \left((\mathrm{A})_m\right)_1, if\ S_\delta = \varnothing \\ \min\ \delta \in S_\delta \qquad , otherwise \end{cases}, \tag{8}$$

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

$$\delta_{m+M} = \begin{cases} -\pi_l\ with\ l = \left((\mathrm{A})_m\right)_2, if\ S_{-\delta} = \varnothing \\ \min\ \delta \in S_{-\delta} \qquad , otherwise \end{cases}, \tag{9}$$

in which $S_{-\delta} = \left\{\delta : R_a(\boldsymbol{\pi}) - R_a\left(\boldsymbol{\gamma}_m(-\delta)\right) \geq \Delta_R\ and\ \delta \geq 0\right\}$. Due to the non-negative 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 the performance loss is as close to Δ_R as possible. 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 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 (5), i.e

$$\mathcal{B} = \left\{\boldsymbol{\gamma} : \boldsymbol{\gamma} = \sum_{m=1}^{2M} \alpha_m \boldsymbol{\gamma}_m(\delta_m),\ \alpha_m \geq 0, and \sum_{m=1}^{2M} \alpha_m = 1\right\}. \tag{10}$$

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

Proof: The proof is given in [14]. 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

We apply large deviations 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

$$\min_{\boldsymbol{\gamma}} D(\boldsymbol{\gamma} \,||\, \boldsymbol{\pi})\ \ s.t.\ \boldsymbol{\gamma} \in \mathcal{S}(\mathcal{B}), \tag{11}$$

where $\mathcal{S}(\mathcal{B})$ is the surface of set $\mathcal{B}$, i.e. $\mathcal{S}(\mathcal{B}) = \mathcal{B} \setminus \mathrm{int}(\mathcal{B})$. Here we denote the interior of set $\mathcal{B}$ as $\mathrm{int}(\mathcal{B})$ [10].

Note that the KL distance $D(\boldsymbol{\gamma} \,||\, \boldsymbol{\pi})$ is convex in the pair $(\boldsymbol{\gamma}, \boldsymbol{\pi})$, and $\boldsymbol{\gamma} \in \mathcal{S}(\mathcal{B})$ is a linear constraint. Therefore, the problem in (11) essentially belongs to convex programming, and the optimal solution can be obtained efficiently by solving the optimization problem for each polyhedron on the boundary $\mathcal{S}(\mathcal{B})$ [11]. Because the convex combinations of the extreme points in $\mathcal{B}$ cover the adjacent region of $\boldsymbol{\pi}$, $\delta_{D_{\min}}$ is sufficient to ensure $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}) \leq \Delta_R$.

C. Minimum Learning Duration Calculation

The second step shows that $D(\boldsymbol{\gamma} \,||\, \boldsymbol{\pi}) \leq \delta_{D_{\min}}$ leads to $R_a(\boldsymbol{\pi}) - R_a(\boldsymbol{\gamma}) \leq \Delta_R$. Hence, an upper bound of the solution to the problem in (5) can be obtained by solving

$$\min t\ \ s.t.\ \mathsf{Prob}\left(D\left(\boldsymbol{\gamma}^t \,||\, \boldsymbol{\pi}\right) \geq \delta_{D_{\min}}\right) \leq \delta_R. \tag{12}$$

Applying formula (6) from large deviations theory, we have the following proposition:

Proposition 2 (An Upper Bound): 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 (5) is

$$T = Min\_t\left(\delta_{D_{\min}}, N, \delta_R\right), \tag{13}$$

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

Proof: It can be proved by combining (6) and (12). ■

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 $\boldsymbol{\pi} = \left[0.25\ 0.5\ 0.25\right]$. We set the parameters in the problem (5) to be $\Delta_R = 10^{-2.5}$ and $\delta_R = 10^{-2}$. Fig. 3 illustrates the procedure of obtaining the upper bound. Noting that $N = 2$, we choose six extreme points $\gamma_1,\cdots,\gamma_6$ in total, which are determined based on rate-pmf curves 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 $\gamma_1,\cdots,\gamma_6$ is the extreme point set $\mathcal{B}$. The dashed hexagon in Fig. 3 is the surface $\mathcal{S}(\mathcal{B})$ on which we minimize the KL distance. Solving the convex optimization problem (11) leads to $\delta_{D_{\min}} = 0.1265$. Using (13), we obtain that $T = Min\_t\left(0.1265, 2, 10^{-2}\right) = 161$. As shown in Fig. 3, if the learning duration is larger than T, the KL distance between the actual stationary distribution $\boldsymbol{\pi}$ and observed empirical frequency function $\boldsymbol{\gamma}^t$ will lie within the solid circle with an outage probability less than δ_R.

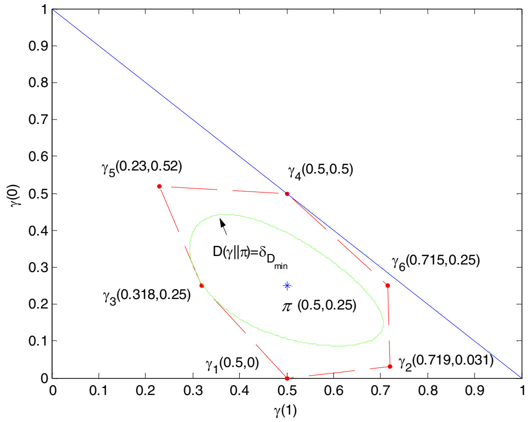

Fig. 3. KL distance minimization in $\mathcal{S}(\mathcal{B})$.

IV. Impact of Feedback Delay

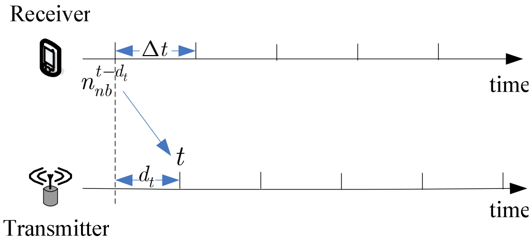

Fig. 4. Feedback delay of the spectrum usage.

This section discusses 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. 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. 4, the spectrum usage pattern that the transmitter receives at time t is the usage pattern the receiver experienced at time $t-d_t$.

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 [7], we know that $\boldsymbol{S}(t)$ is the solution of the Kolmogorov equation, which takes the form of

$$\boldsymbol{S}(t)=\sum_{i=1}^{N+1}\boldsymbol{v_i}e^{t\xi_i}\boldsymbol{\omega_i}\,, \tag{14}$$

in which $\xi_1,\xi_2,\cdots,\xi_{N+1}$ are the $N+1$ distinct eigenvalues of matrix Q, and $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_{N+1}$ and $\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,$ $\boldsymbol{\omega}_{N+1}$ 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 $\boldsymbol{v}_1=[1,1,\cdots,1]^{\mathrm{T}}$ and $\boldsymbol{\omega}_1=\boldsymbol{\pi}$. All the other eigenvalues $\xi_2,\cdots,\xi_{N+1}$ 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

$$\max_{\boldsymbol{P}^{\mathrm{T}}\boldsymbol{1}\le P^{\max}} R\left(\boldsymbol{\pi}^t,\boldsymbol{P}\mid n_{nb}^{t-d_t}\right), \tag{15}$$

in which $\boldsymbol{\pi}^t=\left[\pi_0^t,\pi_1^t,\cdots,\pi_N^t\right]$ is the probability vector of the spectrum usage pattern n_{nb}^t with $\pi_n^t\mid n_{nb}^{t-d_t}=S_{n_{nb}^{t-d_t},n}(d_t)$ $=\Pr\left(n_{nb}^t=n\mid n_{nb}^{t-d_t}\right)$. From (14), we have

$$\lim_{t\to+\infty}\boldsymbol{S}(t)=\boldsymbol{v}_1\boldsymbol{\omega}_1\,. \tag{16}$$

Therefore, if $d_t\to+\infty$, regardless of $n_{nb}^{t-d_t}$, we always have $\boldsymbol{\pi}^t\to\boldsymbol{\omega}_1=\boldsymbol{\pi}$. As a result, $R\left(\boldsymbol{\pi}^t,\boldsymbol{P}\mid n_{nb}^{t-d_t}\right)$ in (15) is reduced to $R(\boldsymbol{\pi},\boldsymbol{P})$ in equation (2). 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 (15) is a function of the limited feedback delay d_t. In particular, $R\left(\boldsymbol{\pi}^t,\boldsymbol{P}\mid n_{nb}^{t-d_t}\right)$ in the optimal transmission strategy of (15) will become:

$$\begin{aligned}R\left(\boldsymbol{\pi}^t,\boldsymbol{P}\mid n_{nb}^{t-d_t}\right)=R\left(S_{n_{nb}^{t-d_t},:}(d_t),\boldsymbol{P}\right)=\sum_{i=1}^{N}\Bigg(\sum_{n\ge i}^{N}S_{n_{nb}^{t-d_t},n}(d_t)\cdot\\ B\log\left(1+\frac{h_iP_i}{N_i+I}\right)+\sum_{n=0}^{n<i}S_{n_{nb}^{t-d_t},n}(d_t)B\log\left(1+\frac{h_iP_i}{N_i}\right)\Bigg),\end{aligned} \tag{17}$$

where $S_{i,:}(d_t)$ represents the i th 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 $t=\Delta t,2\Delta t,\cdots$ by updating empirical frequency functions[2], and use numerical algorithms, such as curve fitting, to estimate $\boldsymbol{S}(t)$ for non-integer multiples of Δt. As long as the environment is stationary and the sampling data is large enough, the wideband device can estimate $\boldsymbol{S}(d_t)$ 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 [12]. Denote d_t' the estimate that the wideband device has about the feedback delay d_t. The performance degradation $\Delta R(d_t')$ of imperfect estimation d_t' is

$$\Delta R(d_t')=\sum_{i=0}^{N}\pi_i\left[R\left(S_{i,:}(d_t),\boldsymbol{P}\left(S_{i,:}(d_t)\right)\right)-R\left(S_{i,:}(d_t),\boldsymbol{P}\left(S_{i,:}(d_t')\right)\right)\right]\ . \tag{18}$$

We derive an upper bound of this performance degradation based on Markov chain theory and formally state it as follows.

Theorem 1: The performance degradation $\Delta R(d_t')$ depends on

[2] This section assumes the sampling period is much smaller than the mixing time of the Markov chain such that the adjacent samples are not independent of each other.

two terms $\left|d_t' - d_t\right|$ and $\min\left(d_t', d_t\right)$ and it is bounded as

$$0 \le \Delta R\left(d_t'\right) \le \alpha\left(\left|d_t' - d_t\right|\right) e^{-\beta \min\left(d_t', d_t\right)}, \qquad (19)$$

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

The proof can be found in [14]. Two key observations can be made from the above theorem. First, it is straightforward to see that the performance loss is a function of $\left|d_t' - d_t\right|$ and the performance loss is zero if $d_t' = d_t$. This can be interpreted as the short term behavior of the imperfect estimation. More importantly, the theorem indicates that the performance loss decreases at least exponentially with $\min\left(d_t', d_t\right)$, which controls the long-term behavior of the imperfect estimation. This result quantifies the significance of the timeliness of the information feedback. Besides, the existence of $\lim_{t\to+\infty} \alpha(t)$ implies that infinite estimation error of the feedback delay causes bounded performance loss. With the increase of $\min\left(d_t', d_t\right)$, the effect of inaccurate estimation of the delay d_t over the performance diminishes at least exponentially.

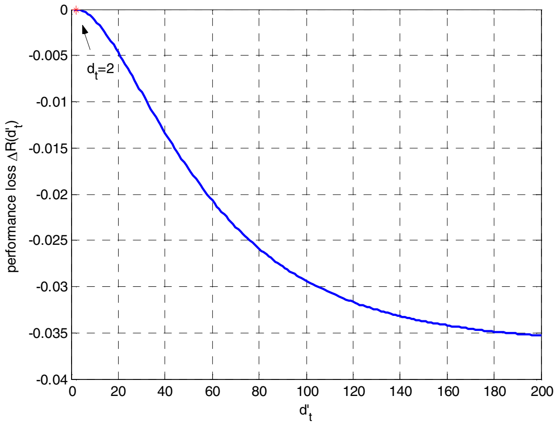

Fig. 5. Performance loss of inaccurate estimate over d_t.

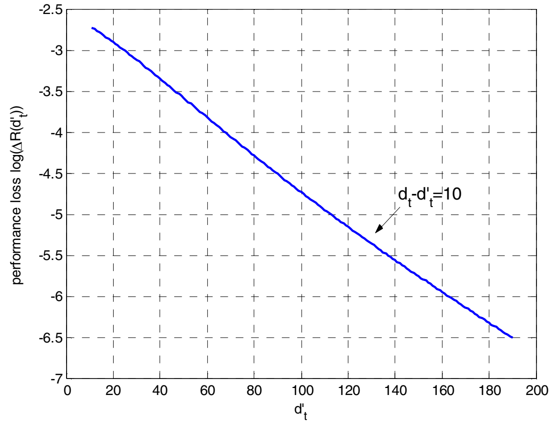

Fig. 6. Performance loss of inaccurate estimate for fixed $d_t - d_t'$.

We numerically show the improvement of measuring the feedback delay d_t. Consider an example with the parameters $N = 2, \lambda_1 = \mu_2 = 0.02,$ and $\lambda_2 = \mu_1 = 0.01$. The feedback delay d_t is assumed to be 2, and the performance loss $\Delta R\left(d_t'\right)$ is plotted in Fig. 5. We can see that it agrees with the argument that $\alpha(0) = 0$ and $\lim_{t\to+\infty} \alpha(t)$ exists in Theorem 1. Compared with taking best response to stationary distribution $\boldsymbol{\pi}$, perfectly knowing the value of feedback delay can increase the achievable rate by 3.5%. We vary d_t' while fixing $d_t - d_t'$ to be 10, and plot the corresponding $\Delta R\left(d_t'\right)$ in Fig. 6. We can see that the performance loss $\Delta R\left(d_t'\right)$ decrease exponentially with d_t', which complies with Theorem 1.

V. 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] 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.

[3] 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

[4] C. Pandana and K.J.R. Liu, "Near Optimal Reinforcement Learning Framework for Energy-Aware Wireless Sensor Communications", *IEEE JSAC*, vol. 23, no 4, pp.788-797, Apr. 2005.

[5] F. Fu and M. van der Schaar, "Dynamic Spectrum Sharing Using Learning for Delay-Sensitive Applications," *Proc. ICC 2008*, to appear

[6] X. R. Zhu, L.F. Shen and T. S. Yum, "Analysis of Cognitive Radio Spectrum Access with Optimal Channel Reservation," *IEEE Commu. Letter*, vol. 11, No. 4, pp. 1-3, April, 2007

[7] R.G. Gallager, *Discrete Stochastic Processes*, Springer, 1995

[8] 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

[9] I. Csiszár and P. C. Shields, “Information theory and statistics: a tutorial,” *Commu. and Inform. Theory*, vol.1, Issue.4, pp. 417-528, 2004

[10] J. B. Conway, *A Course in Functional Analysis*, 2nd edition, Springer-Verlag, 1994.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[12] M. Kazantzidis and M. Gerla, "End-to-end versus Explicit Feedback Measurement in 802.11 Networks", *Proc. ISCC*, pp. 429-434, 2002.

[13] J. S. Rosethal, “Markov chain convergence: From finite to infinite,” *Stochastic Processes and their Applications*, vol. 62, pp.55-72, 1996

[14] Y. Su and M. van der Schaar, “Minimum Required Learning and Impact of Information Feedback Delay for Cognitive Users,” UCLA Technical Report, 2008