

# Delay Analysis for Cognitive Radio Networks with Random Access: A Fluid Queue View

Shanshan Wang\*, Junshan Zhang\* and Lang Tong†

\*Department of Electrical Engineering, Arizona State University, Tempe, AZ, 85287

† School of Electrical Engineering, Cornell University, Ithaca, NY, 14853

**Abstract**—We consider a cognitive radio network where multiple secondary users (SUs) contend for spectrum usage, using random access, over available primary user (PU) channels. Our focus is on SUs' queueing delay performance, for which a systematic understanding is lacking. We take a fluid queue approximation approach to study the steady-state delay performance of SUs, for cases with a single PU channel and multiple PU channels. Using stochastic fluid models, we represent the queue dynamics as Poisson driven stochastic differential equations, and characterize the moments of the SUs' queue lengths accordingly. Since in practical systems, a secondary user would have no knowledge of other users' activities, its contention probability has to be set based on local information. With this observation, we develop adaptive algorithms to find the optimal contention probability that minimizes the mean queue lengths. Moreover, we study the impact of multiple channels and multiple interfaces, on SUs' delay performance. As expected, the use of multiple channels and/or multiple interfaces leads to significant delay reduction.

## I. INTRODUCTION

The traditional wisdom of spectrum management is top-down, in the sense that (orthogonal) frequency channels are assigned to users through licensed or designated bands and only licensed users (also called *primary users*, PUs) can carry out communications over the allotted channels. A recent report from FCC reveals that under this static allocation, merely 5% ~ 15% of the spectrum is utilized on average [1]. This significant under-utilization has spurred a surge of interest in studying cognitive radio networks, such as DARPA's NeXt Generation project and the IEEE standards 802.22 — to name a few. Simply put, cognitive radio (CR) is expected to capture temporal and spatial "spectrum holes" in the radio environment, and to exploit these holes and enable *secondary users* (SUs) for spectrum sharing. Cognitive radio should be frequency-agile, i.e., capable of dynamically adapting operating frequency, bandwidth, and other PHY-layer system parameters. Needless to say, the optimal design of cognitive radio networks is very challenging. One grand challenge is to discover "spectrum holes" for secondary users and distribute them efficiently and fairly, especially when the demand for available spectrum nearly outstrips the supply. In particular, a key functionality needed is the capability of sensing the spectrum and opportunistically using it without causing interference to the primary users.

This research was supported in part by the U. S. National Science Foundation under Grants CNS0905603, CNS 0917087, CNS-0721820, CCF-0635070 and DoD MURI project No. FA9550-09-1-0643.

Most existing works focus on either *spectrum sensing* or *dynamic spectrum sharing*. For spectrum sensing, much attention has been paid to detect spectrum holes while restricting potential interference to primary users to be minimized or under a specified level. Cooperation among secondary users is also exploited to improve detection accuracy. For dynamic spectrum sharing, much effort has been put forward on developing spectrum sharing strategies to maximize the throughput. Interesting models based on graph coloring and game theory have been proposed (see, e.g. [13], [15], [30]).

Needless to say, delay is another important QoS metric in wireless networks. We emphasize that delay performance has important engineering implications, e.g., it can be used to characterize the number of secondary users that can be supported under a given delay constraint. Unfortunately, the delay performance is an under-explored area and not well understood in general, partially due to the difficulties in analyzing it. In particular, to the best of our knowledge, little work on delay performance analysis has been done for cognitive radio networks. In this paper, we analyze delay performance in a cognitive radio network where secondary users contend for channel access using random access only when the PU channels are available.

Recent testbed measurement studies (e.g., [14]) show that the WLAN channels exhibit a semi-Markovian property when transiting between the idle and busy states. Accordingly, we use this model to characterize the PU channel activities, i.e., the PU activity is driven by an ON-OFF process such that each primary user generates data only during the ON-periods. For secondary users, we assume that they generate data packets in each slot according to a Poisson distribution. Based on stochastic fluid queue theory, we model the system dynamics by using *Poisson driven stochastic differential equations* (PDSDE), and analyze the steady state queue lengths of secondary users accordingly. To facilitate tractability, we focus on the light traffic regime where the traffic intensity is low, as is often the case for delay analysis of buffered Aloha, e.g., [10], [24]. We consider the homogeneous case where the arrival rates of secondary users are the same, and characterize the moments of the random queue lengths of secondary users, for cases with a single PU channel (SCH) and multiple PU channels (MCH). Clearly, the moments provide critical statistical information about SUs' queueing length distribution. We also examine the impact of the PU traffic on SUs' queue lengths and the gain

of using multiple PU channels. Adaptive algorithms, based on local information only, are developed to find the optimal contention probabilities that achieve the minimum mean queue lengths.

Next, we explore the gain of using two interfaces per secondary user, i.e., each secondary user is equipped with two interfaces (radios). Accordingly, each secondary user can sense two channels at a time and thus transmit on up to two channels, as long as the PU channels are available and no contention collisions occur. Our analysis and numerical examples corroborate the intuition that the usage of two interfaces can greatly improve the delay performance by decreasing the mean queue lengths of secondary users.

Perhaps the most related work to our study here is [18], which studied fluid queues with priority. We emphasize here a few major differences. First, [18] considered a single low priority flow which receives a constant service rate whenever the buffer of the high priority flows is empty. In contrast, in our study, secondary users contend for the channels using random access when primary users are inactive, and the number of backlogged secondary users is time-varying. As a result, the service rate is random. Second, random access to the PU channels gives rise to the coupling across secondary users' queue lengths, which was not the case in [18] since only one low priority flow was considered there. Besides the work in [18], we also note that delay performance of a multi-hop wireless ad hoc network has been studied in [7], where *diffusion approximation* was used to characterize the average end-to-end delay. In [21], WLANs with access points connecting a fixed number of users in the presence of HTTP traffic was considered. A processor sharing queue with state-dependent service rate was used to model the system and analyze the mean session delay. In [27], queueing delay at nodes in an IEEE 802.11 MAC-based network was analyzed, where each node was modeled as a discrete time  $G/G/1$  queue. Delay analysis for buffered Aloha has also been studied (see [4], [10], [24], [26], [28] and references therein). In [4] and [28], the approach named "tagged user" was adopted. Specifically, the interfering users/nodes were modeled as "independent" queues in the sense that the analysis was conducted on one particular user, named the "tagged user", while the interference across users was incorporated into the characterization of the service time distribution of this tagged user. Another approach utilizing Markov chains with reduced state space to approximate delay analysis can be found in [10], [24] and [26]. Two Markov chains, one for the queueing dynamics at one user, and the other for the system status (i.e., the number of busy users, and/or the identities of the users (empty, busy or blocked)), were employed for characterizing the steady state distributions of the system as well as the delay. It is worth noting that the approximation worked well only for the low traffic regime, as has been pointed out in [10] and [24]. In very recent work [17], the stationary queue tail distribution of a single SU in a multi-channel CR network was analyzed using a large deviation approach, where the traffic of the SU was modeled as a constant arrival process and PU traffic as

a Markov chain. For the special case with two PU channels, closed-form expressions were obtained for the tail distribution of the queue length, and upper and lower bounds were obtained for more general case.

We have a few more words on fluid models. Fluid approximation is a widely used tool for performance analysis in many fields, including communication networks and control techniques [20]. It can provide a good approximation to the original systems by converting the discrete packets into a continuous fluid and offers greater tractability in analyzing the system performance. We should note that along a different avenue, the *deterministic* fluid model has been developed to analyze queueing systems, where microscopic fluctuations in the original systems are replaced by their mean values (see, e.g., [11], [12]). For a given random process  $G(t)$ , the resulted *fluid scale process*, obtained by using the *Functional Law of Large Numbers*, is defined as  $\tilde{g}^n(t) = G(nt)/n$ , i.e., the time and space are scaled by the same factor  $n$  for  $n$  being large. This deterministic model leads to the application of *ordinary differential equations* (ODE), which is in contrast to the stochastic differential equations we shall use in our context.

The rest of the paper is organized as follows. In Section II, we introduce the system model. Fluid flow approximation and PDSDE-based analysis on the single PU channel case are given in Section III, followed by analysis on the multi-channel case in Section IV. Section V studies the delay performance when each secondary user is equipped with two interfaces. Finally, conclusions and discussions come in Section VI.

## II. SYSTEM MODEL

Consider a time-slotted (with slot duration normalized to be 1) cognitive radio network with  $N$  PU channels and  $M$  secondary users, where secondary users contend for the channel using random access when the primary users are inactive, as illustrated in Fig. 1. This model is of interest to many practical scenarios. For example, in a sensor network equipped with cognitive radios, sensors send out measurement data of the environment sporadically and opportunistically over "empty" PU channels.

Without loss of generality, we associate one primary user with one channel (one can use a virtual primary user to represent the PU activity). The PU activity in channel  $j$  can be represented as an ON-OFF process  $x_j(t)$ ,  $j = 1, 2, \dots, N$ , i.e., the primary user is ON when  $x_j(t) = 1$  (OFF when  $x_j(t) = 0$ ) and generates data traffic at rate  $r_j$ . The transmission rate on each channel is normalized to be 1. We are interested in the case where  $r_j > 1$  (the case with  $r_j \leq 1$  is trivial since the primary users' buffers are always empty). Let  $A_{n,j}$  and  $S_{n,j}$  denote the  $n$ th active and silent period of  $x_j(t)$  respectively. As is standard, we assume that  $\{A_{n,j}\}$  are *i.i.d.* and follow a general distribution with a Laplace-Stieltjes transform  $F(\chi) = E[e^{-\chi A_{n,j}}]$  and  $E[A_{n,j}] = \nu_{H_j}$ , and that  $\{S_{n,j}\}$  are independent from  $\{A_{n,j}\}$  and follow an exponential distribution with  $E[S_{n,j}] = 1/\lambda_{H_j}$ . It is worth noting that since PU's ON/OFF periods are typically much larger than the duration of one slot, we here neglect the edge

effect where collisions between PUs and SUs occur when primary users transit from OFF to ON. That is, the probability that primary users generate new data during the middle of a slot and therefore preempt the transmission of secondary users is negligible.

We assume that in each slot, each secondary user generates data packets according to a Poisson distribution with rate  $\lambda$ . Typically, in a cognitive radio network, primary users have strict priority over secondary users: secondary users can transmit only if the channels are unoccupied by primary users. Simply put, each secondary user with backlogged data chooses a channel independently and uniformly at a time to probe. If the channel is sensed to be unoccupied, it contends for the channel with probability  $p$ . If the contention is successful, the user then transmits its backlogged data. Note that in practical scenarios, a secondary user would not have the knowledge of how many backlogged SUs there are, and accordingly we set the contention probability  $p$  to be oblivious of backlogged SUs.

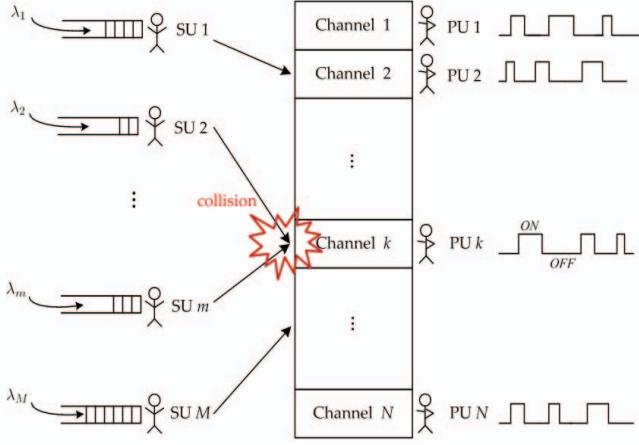


Fig. 1. A cognitive radio network with multiple PUs and SUs

For notational convenience, let  $H_j(t)$  and  $L_i(t)$  denote the queue lengths corresponding to primary user  $j$  and secondary user  $i$  at time  $t$ , respectively, and  $P_{I_j}$  be the probability that primary user  $j$  is idle, i.e.,  $P_{I_j} = \Pr(H_j(t) = 0)$ . In the following, we shall focus on characterizing the queue lengths of secondary users.

### III. MULTIPLE SUs MEET SINGLE PU

#### A. Sample Path Description Using Poisson Driven Stochastic Differential Equations

We first consider the case with a single PU channel. For notational convenience, we drop the subscript  $j$  related to the PU parameters. In order to guarantee system stability, we enforce that

$$\lambda < \min \left\{ \frac{1}{M} \left( 1 - \frac{r\lambda_H\nu_H}{\lambda_H\nu_H + 1} \right), \frac{1}{eM} P_I \right\}. \quad (1)$$

It is worth mentioning that the second term in (1) was established using the idea of “dominant systems”, which has been used in approximating the stability region of interacting queues in random access systems (e.g., [19], [25]). In our context, the “dominant system” is a system where a secondary user continues to probe the PU channel regardless of its buffer state (empty or backlogged). Accordingly, the stability region for this system is given by  $\lambda < P_I p(1-p)^{M-1}$ . Based on [19] and [25], the original system is stable if the dominant system is stable. In other words, the stability region obtained through the dominant system serves as an inner bound to that of the original system.

The queue dynamics of secondary user  $i$  can be written as

$$L_i(n+1) = [L_i(n) + U_i(n) - V_i(n)]^+,$$

where  $U_i(n)$  and  $V_i(n)$  stand for the arrivals and departures to/from SU  $i$ 's queue during slot  $n$ .

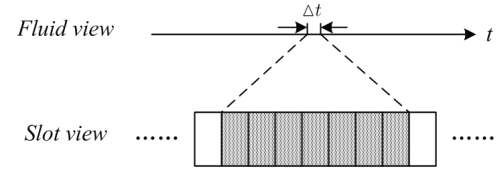


Fig. 2. Fluid approximation of a slotted system

To facilitate analysis, in the following, we take a macroscopic view on the queue evolution of secondary users across multiple slots and use continuous approximation to characterize the dynamics in secondary users' activities (as illustrated in Fig. 2). Let  $\zeta_i(t)$  be the indicator random variable for the contention of secondary user  $i$  at time  $t$  (i.e., when it contends,  $\zeta_i(t) = 1$ ; otherwise  $\zeta_i(t) = 0$ ). The following stochastic differential equation is thus obtained:

$$dL_i(t) = dN_i(t) - (1 - \mathcal{I}_{H(t)})\mathcal{I}(\zeta_i(t) = 1)\mathcal{I}_{L_i(t)} \prod_{\substack{k=1 \\ k \neq i}}^{M-1} [1 - \mathcal{I}_{L_k(t)}\mathcal{I}(\zeta_k(t) = 1)] dt, \quad (2)$$

where  $\{N_i(t)\}$  are a set of Poisson counters with rate  $\lambda$ ; and  $\mathcal{I}_{f(t)}$  stands for the indicator function  $\mathbf{1}(f(t) > 0)$ .

Furthermore, it is clear that for the primary user, its dynamics can be characterized as follows:

$$dH(t) = rx(t)dt - \mathcal{I}_{H(t)}dt. \quad (3)$$

Observe that (2) forms a set of Poisson driven stochastic differential equations (PDSDE) [8], [9]. Simply put, in a PDSDE, Poisson processes are the driving sources capturing the system dynamics, and this is in contrast to the conventional SDE where the Brownian Motion is used to describe the dynamics in the trajectory of a stochastic differential equation. In general, a PDSDE can be given as

$$x(t) = x(0) + \int_0^t f(x(\sigma), \sigma) d\sigma + \int_0^t g(x(\sigma), \sigma) dN_\sigma, \quad (4)$$

where  $N_\sigma$  is a Poisson counter. For the sake of completeness, we restate the definition of the solution to the above PDSDE [8].

*Definition 1:* A function  $x(\cdot)$  is a *solution* to (4), in Itô's sense, if on an interval where  $N_\sigma$  is constant,  $x$  satisfies  $\dot{x} = f(x, t)$  and if  $N_\sigma$  jumps at  $t_1$ ,  $x$  behaves in a neighborhood of  $t$  according to the rule

$$\lim_{\substack{t \rightarrow t_1 \\ t > t_1}} x(t) = g(\lim_{\substack{t \rightarrow t_1 \\ t < t_1}} x(t), t_1) + \lim_{\substack{t \rightarrow t_1 \\ t < t_1}} x(t),$$

and  $x(\cdot)$  is taken to be continuous from the left. When this definition is adopted, we can rewrite (4) as

$$dx(t) = f(x, t)dt + g(x, t)dN_\sigma(t).$$

Based on the properties of PDSDE [8], it can be shown that for  $n \geq 2$ ,

$$dL_i^n(t) = nL_i^{n-1}(t)dL_i(t) + \sum_{k=2}^n \binom{n}{k} L_i^{n-k}(t)dN_i(t).$$

It follows that the moments of  $L_i(t)$  in the steady state satisfy the following recursive equation:

$$nE[L_i^{n-1}f] - \sum_{k=1}^n \binom{n}{k} E[L_i^{n-k}] \lambda = 0, \quad (5)$$

where

$$f = (1 - \mathcal{I}_{H(t)})\mathcal{I}(\zeta_i(t) = 1)\mathcal{I}_{L_i(t)} \prod_{\substack{k=1 \\ k \neq i}}^{M-1} [1 - \mathcal{I}_{L_k(t)}\mathcal{I}(\zeta_k(t) = 1)].$$

### B. Moments of SU Queue Lengths

We now start to study in more detail the moments of the queue lengths of secondary users based on the above PDSDEs. Recall that secondary users can access the channel only when the buffer of the primary user is empty. With this observation, we first examine the idle period of the primary user. Note that the primary user generates data at rate  $r$  only during an ON-period, and that the buffer is depleted at rate 1 as long as the queue is nonempty. Based on [2] and [16], the distribution of the queue length of a single class fluid queue is closely related to the distribution of the stationary workload of an  $M/G/1$  queue. Let  $W$  denote the stationary workload of the  $M/G/1$  queue. It follows that

$$P_I = 1 - \gamma_h \Pr(W > 0) = 1 - \gamma_h \eta \lambda_H, \quad (6)$$

where  $\gamma_h = r/((r-1)(1+\nu_H \lambda_H))$  and  $\eta$  is the mean service time for the  $M/G/1$  queue.

The expression in (6) gives the idle probability of PU activity with generally distributed ON-periods. To obtain a more concrete characterization of  $P_I$ , in the following, we focus on the special case where the ON-periods of  $x(t)$

are exponentially distributed, i.e.,  $\{A_n\}$  are exponentially distributed with  $E[A_n] = 1/\mu_H$ . The sample path description of the PU traffic then boils down to the following PDSDE:

$$\begin{aligned} dx(t) &= (1 - x(t))dN_{H_1}(t) - x(t)dN_{H_2}(t), \\ dH(t) &= r x(t)dt - \mathcal{I}_{H(t)}dt, \end{aligned} \quad (7)$$

where  $N_{H_1}(t)$  and  $N_{H_2}(t)$  are a pair of Poisson counters driving  $x(t)$ , with rate  $\lambda_H$  and  $\mu_H$  respectively. Based on the properties of PDSDE [8], it is not difficult to show that in the steady state

$$P_I = 1 - \frac{r\lambda_H}{\lambda_H + \mu_H}. \quad (8)$$

We then start characterizing the moments of the SU queue lengths. Based on (2), we observe that the  $M$  SU queues interact with each other through channel contention. In other words, besides the impact from PU activities, the service time of one secondary user also depends on other SUs' activities, and it turns out to be a quantity that follows a general distribution which is difficult to determine.

For ease of exposition, we shall focus on the light traffic regime and approximate the SU activities as if they were "weakly coupled" in the sense that the event that one secondary user is idle (i.e., with no backlogged data) is independent from other secondary users being idle. Similar approximations to "decouple" the interacting queues have been made in [4] and [28], among other works. According to the homogeneity assumption, this *idle probability* would be the same across all secondary users. Let  $p_0$  be this probability. It is clear that the number of backlogged SUs follows a Binomial distribution with its PMF (probability mass function) given by

$$P_m = \binom{M}{m} (1 - p_0)^m p_0^{M-m}, \quad (9)$$

where  $p_0$  can be shown to satisfy [28]

$$p_0 = 1 - \rho,$$

with  $\rho = \frac{\lambda}{\mu}$  and  $\mu$  being the mean service rate. In the case with a single PU channel,  $\mu$  can be calculated as

$$\begin{aligned} \mu &= \frac{1}{M} \sum_{m=1}^M m p (1 - p)^{m-1} P_I P_m \\ &= p P_I (1 - p_0) (1 - p + p p_0)^{M-1}. \end{aligned}$$

It follows that

$$p_0 = 1 - \frac{\lambda}{p P_I (1 - p_0) (1 - p + p p_0)^{M-1}}. \quad (10)$$

Now with all related parameters being characterized, we are in a position to calculate the moments of the queue lengths for secondary users. Based on (2) and (5), the first two moments of SU  $i$ 's queue length can be derived as

$$E[L_i] = \frac{\lambda}{-2\lambda + 2\alpha_s}, \quad E[L_i^2] = \frac{\lambda(\lambda + 2\alpha_s)}{6(\lambda - \alpha_s)^2}, \quad (11)$$

where  $\alpha_s$  is given by

$$\alpha_s = \sum_{m=1}^M p(1-p)^{m-1} P_m P_I. \quad (12)$$

### C. Adaptive Algorithm for Optimal Contention Probability

Based on the analysis above, it is clear that the contention probability,  $p$ , and the idle probability of one secondary user,  $p_0$ , are two key parameters to the characterization of the mean queue lengths of SUs, and thus the delay performance. Intuitively speaking, when  $p$  is very small (approaching 0), secondary users contend for the channel sporadically, and  $p_0$  is small. On the other hand, when  $p$  is very large (approaching 1), all SUs with backlogs contend for the channel almost always, leading to a high contention collision among SUs, which makes the queue lengths increase. It is thus indicated that there exists an optimal value of  $p$ , which minimizes the mean queue lengths.

We note that (10) formulates a fixed point equation for the idle probability  $p_0$ , and  $p_0$  is in itself an implicit function of the contention probability  $p$ , i.e.,  $p$  is the argument of  $p_0$ . Therefore, we obtain the optimal value of  $p$  by taking derivative with respect to  $p$  on both sides of (10) and setting  $dp_0/dp = 0$ . After some straightforward calculation, we obtain

$$p = \min \left\{ \frac{1}{M(1-p_0)}, 1 \right\}. \quad (13)$$

Intuitively speaking,  $M(1-p_0)$  corresponds to the average number of backlogged users who would contend for channel access. Recall that  $p_0$  is the probability that one SU's queue is empty. Accordingly, stochastic approximation algorithms, based on local information only, can be easily developed to find the optimal contention probability. Specifically, let  $\Phi_i(t)$  be the indicator random variable denoting the status of the queue of secondary user  $i$  at time  $t$ , i.e., when the queue is empty,  $\Phi_i(t) = 1$ , and otherwise,  $\Phi_i(t) = 0$ . Due to homogeneity, we drop the subscript  $i$ . We note that the adaptation of  $p$  is based on the update of  $p_0$ . It follows that we can devise the following adaptive algorithms to obtain the optimal  $p$ . First, we use stochastic approximation to update  $p_0$  as:

$$p_0(t+1) = \left(1 - \frac{1}{t+1}\right) p_0(t) + \frac{1}{t+1} \Phi(t+1). \quad (14)$$

Based on this adaptation, we next derive the adaptive algorithm for achieving the optimal  $p$  as:

$$p(t+1) = p(t) + \frac{1 - Mp(t)(1 - \Phi(t+1))}{t + Mp(t)(1 - \Phi(t+1))} p(t). \quad (15)$$

We now illustrate by numerical examples the above results, where the contention probability is set to be the optimal value given by (13). As shown in Fig. 3, the queue lengths increase with the arrival rate of secondary users, and as the *duty cycle* of  $x(t)$ , defined as  $\tau \equiv \frac{\lambda_H}{\lambda_H + \mu_H}$ , increases, the average queue lengths of secondary users increase, indicating the impact of the PU traffic on SUs' delay performance.

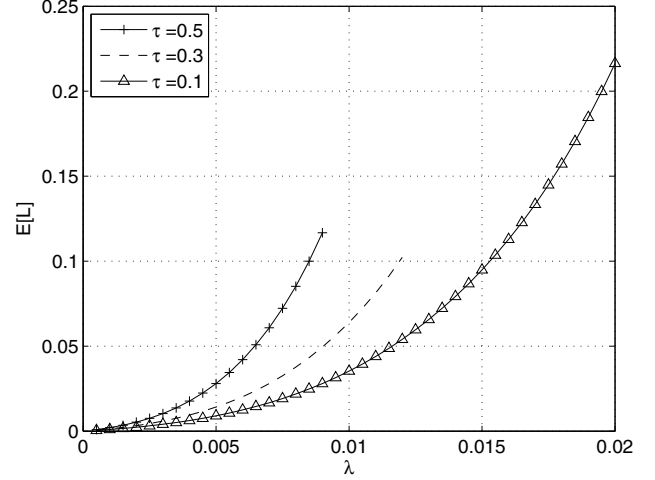


Fig. 3. Mean queue length of SUs for the case with a single PU channel

## IV. MULTIPLE SUs MEET MULTIPLE PUS

We next consider the case where there are multiple PU channels, and examine the performance gain therein.

### A. Sample Path Description Using Poisson Driven Stochastic Differential Equations

In this case, to keep the system stable, we enforce that

$$\lambda < \min \left\{ \frac{1}{M} \left( N - \sum_{j=1}^N \frac{r_j \lambda_{H_j} \nu_{H_j}}{\lambda_{H_j} \nu_{H_j} + 1} \right), \frac{N}{eM} P_I \right\}, \quad (16)$$

where, again, the second term was obtained along the same line as in the previous section using the idea of “dominant system”.

Let  $\xi_{il}(t)$  be an indicator random variable denoting that secondary user  $i$  chooses channel  $l$  at time  $t$ . As in the single PU channel case, we do continuous approximation when characterizing the dynamics of SU activities. The system dynamics can then be written as: for  $j = 1, 2, \dots, N$ ,

$$dH_j(t) = r_j x_j(t) dt - \mathcal{I}_{H_j(t)} dt, \quad (17)$$

and for  $i = 1, 2, \dots, M$ ,

$$dL_i(t) = dN_i(t) - \sum_{l=1}^N (1 - \mathcal{I}_{H_l(t)}) \mathcal{I}_{L_i(t)} \mathcal{I}(\xi_{il}(t)=1) \mathcal{I}(\zeta_i(t)=1) \\ \prod_{\substack{k=1 \\ k \neq i}}^{M-1} [1 - \mathcal{I}_{L_k(t)} \mathcal{I}(\xi_{ki}(t)=1) \mathcal{I}(\zeta_k(t)=1)] dt. \quad (18)$$

Again, the coupling across secondary users is observed in (18). We next carry out analysis on the moments of SUs' queue lengths by focusing on the light traffic regime as before.

### B. Moments of SU Queue Lengths

Along the same line as in the single PU channel case, we first characterize the idle period of primary users. Again, we consider the case where the ON-periods of the PU traffic sources,  $\{A_{n,j}\}$ , follow an exponential distribution with  $E[A_{n,j}] = 1/\mu_{H_j}$ . The sample path description for primary user  $j$  is given by the following PDSDE:

$$\begin{aligned} dx_j(t) &= (1 - x_j(t))dN_{H_1}^{(j)}(t) - x_j(t)dN_{H_2}^{(j)}(t), \\ dH_j(t) &= r_j x_j(t)dt - \mathcal{I}_{H_j}(t)dt, \end{aligned} \quad (19)$$

where  $N_{H_1}^{(j)}(t)$  (with rate  $\lambda_{H_j}$ ) and  $N_{H_2}^{(j)}(t)$  (with rate  $\mu_{H_j}$ ) are a pair of Poisson counters driving  $x_j(t)$ . For better tractability, we consider the case where the PU channels are *i.i.d.*. It follows that  $P_{I_j} = P_{I_l}, \forall j \neq l$ . Denote by  $P_I = P_{I_j}$  for simplicity. It is easy to show that in the steady state,  $P_I$  can be calculated as is given in (8).

Next, we turn our attention to study the moments of secondary users' queue lengths.

Applying the PDSDE tools, we obtain for secondary user  $i$ ,

$$E[L_i] = \frac{\lambda}{-2\lambda + 2\alpha_{\mathcal{M}}}, \quad E[L_i^2] = \frac{\lambda(\lambda + 2\alpha_{\mathcal{M}})}{6(\lambda - \alpha_{\mathcal{M}})^2}, \quad (20)$$

where

$$\begin{aligned} \alpha_{\mathcal{M}} &= \sum_{l=1}^N \sum_{m=1}^M \sum_{k=0}^{m-1} \binom{m-1}{k} p(1-p)^k \left(\frac{1}{N}\right)^{k+1} \left(1 - \frac{1}{N}\right)^{m-(k+1)} P_m P_I \\ &= \sum_{m=1}^M p P_I \left(1 - \frac{p}{N}\right)^{m-1} P_m, \end{aligned} \quad (21)$$

with  $P_m$  being given by (9).

In order to characterize the moments of the queue length of secondary users, we first need to find the mean service rate  $\mu$ , which, in this case, can be calculated as:

$$\begin{aligned} \mu &= \frac{1}{M} \sum_{m=1}^M m p \sum_{l=1}^N \frac{1}{N} \sum_{k=0}^{m-1} \binom{m-1}{k} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{(m-1)-k} (1-p)^k P_I P_m \\ &= p P_I (1 - p_0) \left(1 - \frac{p}{N} + \frac{p p_0}{N}\right)^{M-1}. \end{aligned}$$

It follows that

$$p_0 = 1 - \frac{\lambda}{p P_I (1 - p_0) \left(1 - \frac{p}{N} + \frac{p p_0}{N}\right)^{M-1}}. \quad (22)$$

The characterization of  $P_m$  and  $E[L_i]$  then follows.

### C. Adaptive Algorithm for Optimal Contention Probability

Similar as in the single PU channel case, taking derivative with respect to  $p$  and setting  $dp_0/dp = 0$  yields that

$$p = \min \left\{ \frac{N}{M(1 - p_0)}, 1 \right\}. \quad (23)$$

It is not difficult to see that  $M(1 - p_0)/N$  is the average number of backlogged secondary users per PU channel. Based on (23), similar adaptive algorithms for obtaining optimal  $p$  can be developed as in the single PU channel case.

Meanwhile, we note that  $p = \frac{N}{M(1 - p_0)}$  holds when  $N < M(1 - p_0)$ . In fact, this is the regime of interest when we characterize the gain of using multiple PU channels. The other case with  $p = 1$  will be discussed in more detail in Section V-C. Here we present numerical examples to illustrate the above analysis. The contention probabilities are set to be their optimal values. As illustrated in Fig. 4, the mean queue lengths of secondary users decrease sharply when multiple PU channels are present, pointing to a *multi-channel gain* therein. A more straightforward illustration of such a gain was depicted in Fig. 5, where the gain was defined as the ratio  $E[L_i]^{(\mathcal{S})}/E[L_i]^{(\mathcal{M})}$  with the superscripts  $\mathcal{S}$  and  $\mathcal{M}$  denoting the cases with a single PU channel and multiple PU channels respectively. It is clear from the figure that as the arrival rate of SUs increases, or the duty cycle of PUs increases, the multi-channel gain increases as well. In other words, with a given delay requirement, the capability of the system supporting an increased amount of SU traffic improves when more PU channels are present, and such improvement becomes more evident as the PU activities grow heavier.

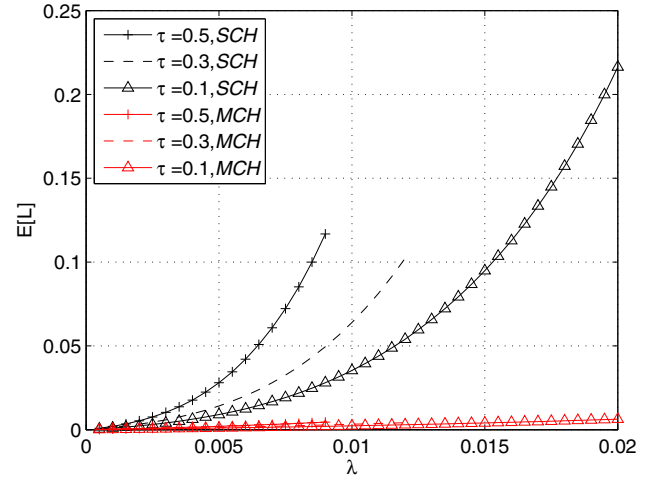


Fig. 4. Comparison of mean queue lengths for cases with a single PU channel and multiple PU channels

## V. POWER OF TWO INTERFACES

Intrigued by the celebrated results in [5] and [23], we next explore the impact of using two interfaces (radios) by each secondary user on the delay performance in a cognitive radio network.

### A. System Model

We first outline a few new ingredients in the system model. We assume that each secondary user is equipped with two interfaces (this approach can be easily generalized to cases

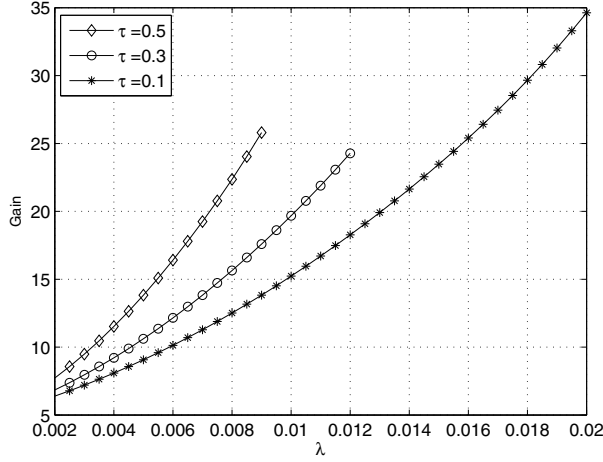


Fig. 5. Gain of using multiple PU channels

with more radios), and randomly chooses two channels independently and uniformly at a time. If the chosen PU channels are unoccupied, the secondary user contends for them with probability  $p$ . If no collisions occur, it starts transmission on the channels. Clearly, in this case, each secondary user can access up to two channels for transmission at a time, thus decreasing the delay.

For notational convenience, let  $c_1(t)$  and  $c_2(t)$  denote the two channels secondary user  $i$  chooses at time  $t$  (we drop the time index  $t$  in the following as the meaning is clear). The system dynamics, after fluid approximation, can be written as: for  $j = 1, 2, \dots, N$ ,

$$dH_j(t) = r_j x_j(t) dt - \mathcal{I}_{H_j(t)} dt, \quad (24)$$

and for  $i = 1, 2, \dots, M$ ,

$$dL_i(t) = dN_i(t) - \sum_{\{c_1, c_2\}} \mathcal{I}(\xi_{ic_1}(t) = 1, \xi_{ic_2}(t) = 1) \mathcal{I}_{L_i(t)} \mathcal{I}(\zeta_i(t) = 1) (D_1 + D_2) dt, \quad (25)$$

where

$$D_1 = (1 - \mathcal{I}_{H_{c_1}(t)}) \prod_{\substack{k=1 \\ k \neq i}}^{M-1} [1 - \mathcal{I}_{L_k(t)} \mathcal{I}(\xi_{kc_1}(t) = 1) \mathcal{I}(\zeta_k(t) = 1)]$$

and

$$D_2 = (1 - \mathcal{I}_{H_{c_2}(t)}) \prod_{\substack{k=1 \\ k \neq i}}^{M-1} [1 - \mathcal{I}_{L_k(t)} \mathcal{I}(\xi_{kc_2}(t) = 1) \mathcal{I}(\zeta_k(t) = 1)]$$

denote the events that channel  $c_l, l = 1, 2$ , is available.

### B. Moments of SU Queue Lengths

With the setting of *i.i.d.* PU channels unchanged, the idle probability of primary users,  $P_I$ , would be still given by (8) for the case where the ON-periods of  $x_j(t)$ 's are exponentially distributed.

As in the case with a single interface, we obtain that for secondary user  $i$ ,

$$E[L_i] = \frac{\lambda}{-2\lambda + 2\alpha_c}, \quad E[L_i^2] = \frac{\lambda(\lambda + 2\alpha_c)}{6(\lambda - \alpha_c)^2}, \quad (26)$$

where

$$\alpha_c = \sum_{\{c_1, c_2\}} \sum_{m=1}^M \frac{1}{\binom{N}{2}} p \{2Z_2 + Z_1\} P_m,$$

with  $Z_1$  and  $Z_2$  being the conditional probabilities that one out of the two chosen channels is available, and both of them are available, given that the total number of backlogged secondary users is  $m$ , respectively. To get a more concrete characterization of  $\alpha_c$ , we denote by  $\mathcal{A}_l$  and  $\mathcal{B}_l$  the events that channel  $c_l, l = 1, 2$ , is unoccupied by primary users and un-contended by other secondary users, given that the total number of backlogged SUs equals  $m$ . It is clear that  $\{\mathcal{A}_l\}$  are independent from  $\{\mathcal{B}_l\}$ . It follows that

$$\begin{aligned} Z_2 &= \Pr(\mathcal{A}_1 \mathcal{A}_2 \mathcal{B}_1 \mathcal{B}_2) = P_I^2 \Pr(\mathcal{B}_1 \mathcal{B}_2), \\ Z_1 &= \Pr(\mathcal{A}_1 \mathcal{A}_2 \bar{\mathcal{B}}_1 \mathcal{B}_2) + \Pr(\mathcal{A}_1 \mathcal{A}_2 \mathcal{B}_1 \bar{\mathcal{B}}_2) + \Pr(\mathcal{A}_1 \bar{\mathcal{A}}_2 \mathcal{B}_1) + \Pr(\bar{\mathcal{A}}_1 \mathcal{A}_2 \mathcal{B}_2) \\ &= 2P_I \Pr(\mathcal{B}_1) - 2P_I^2 \Pr(\mathcal{B}_1 \mathcal{B}_2), \end{aligned}$$

where  $\Pr(\mathcal{B}_1)$  and  $\Pr(\mathcal{B}_1 \mathcal{B}_2)$  can be derived as

$$\Pr(\mathcal{B}_1) = \sum_{k=0}^{m-1} \binom{m-1}{k} \left( \frac{\binom{N-1}{2}}{\binom{N}{2}} \right)^k \left( 1 - \frac{\binom{N-1}{2}}{\binom{N}{2}} \right)^{(m-1)-k} (1-p)^k,$$

and

$$\Pr(\mathcal{B}_1 \mathcal{B}_2) = \sum_{k=0}^{m-1} \binom{m-1}{k} \left( 1 - \frac{\binom{N-2}{2}}{\binom{N}{2}} \right)^k \left( \frac{\binom{N-2}{2}}{\binom{N}{2}} \right)^{(m-1)-k} (1-p)^k.$$

After some algebra, we obtain that

$$\alpha_c = \sum_{m=1}^M 2P_I p \left( 1 - \frac{2p}{N} \right)^{m-1} P_m. \quad (27)$$

Again, we have that  $p_0 = 1 - \frac{\lambda}{\mu}$  where the mean service rate can be calculated as

$$\begin{aligned} \mu &= \frac{1}{M} \sum_{m=1}^M 2mp P_I \sum_{k=0}^{m-1} \binom{m-1}{k} \left( \frac{\binom{N-1}{2}}{\binom{N}{2}} \right)^k \\ &= 2p P_I (1 - p_0) \left( 1 - \frac{2p}{N} + \frac{2pp_0}{N} \right)^{M-1}. \end{aligned}$$

Therefore, the idle probability of one secondary user is given by

$$p_0 = 1 - \frac{\lambda}{2p P_I (1 - p_0) \left( 1 - \frac{2p}{N} + \frac{2pp_0}{N} \right)^{M-1}}. \quad (28)$$

We note that the calculation of  $\mu$  here provides an upper bound on the per user throughput for secondary users, since we have neglected the situation where the SUs have only one packet in the queue and can therefore transmit on only one of

the two channels even if both of them are available. However, such cases would be of little interest and are not considered here.

### C. Adaptive Algorithm for Optimal Contention Probability

Following similar steps as in the previous cases, we obtain the optimal contention probability to be

$$p = \min \left\{ \frac{N}{2M(1-p_0)}, 1 \right\}, \quad (29)$$

from where we note that  $\frac{M(1-p_0)}{N/2}$  is intuitively the average number of backlogged secondary users per PU channel. Adaptive algorithms similar to that described by (14) and (15) can be devised to find the optimal  $p$ .

Different from the multi-channel gain, the power of two choices is typically analyzed in the regime where  $N \gg M$ . With this insight, we next focus on the case where  $N \geq 2M$  and characterize the gain of using two interfaces per SU. It is clear that when  $N \geq 2M$ , the optimal contention probability is given by  $p = 1$ . It follows that  $\alpha_{\mathcal{M}}$  and  $\alpha_c$  can be rewritten as

$$\alpha_{\mathcal{M}} = \sum_{m=1}^M P_I \left(1 - \frac{1}{N}\right)^{m-1} P_m, \quad (30)$$

and

$$\begin{aligned} \alpha_c &= \sum_{\{c_1, c_2\}} \sum_{m=1}^M \frac{1}{\binom{N}{2}} \left\{ 2\Pr(\mathcal{A}_1 \mathcal{A}_2 \mathcal{C}_1 \mathcal{C}_2) + [\Pr(\mathcal{A}_1 \mathcal{A}_2 \bar{\mathcal{C}}_1 \mathcal{C}_2) \right. \\ &\quad \left. + \Pr(\mathcal{A}_1 \mathcal{A}_2 \mathcal{C}_1 \bar{\mathcal{C}}_2) + \Pr(\mathcal{A}_1 \bar{\mathcal{A}}_2 \mathcal{C}_1) + \Pr(\bar{\mathcal{A}}_1 \mathcal{A}_2 \mathcal{C}_2) \right\} P_m \\ &= \sum_{m=1}^M 2P_I \left(1 - \frac{2}{N}\right)^{m-1} P_m, \end{aligned} \quad (31)$$

where  $\mathcal{C}_l$  stands for the event that channel  $c_l, l = 1, 2$ , is not chosen by any other SUs, given that the total number of backlogged secondary users is  $m$ .

Also, the mean service rate (denoted as  $\mu_d, d = 1, 2$ ) and the idle probability (denoted as  $p_{0,d}, d = 1, 2$ ) of secondary users can be recalculated as follows: for the case with a single interface,

$$\begin{aligned} \mu_1 &= \frac{1}{M} \sum_{m=1}^M m P_I \left(1 - \frac{1}{N}\right)^{m-1} P_m \\ &= P_I (1 - p_{0,1}) \left(1 - \frac{1}{N} + \frac{p_{0,1}}{N}\right)^{M-1}, \\ p_{0,1} &= 1 - \frac{\lambda}{P_I (1 - p_{0,1}) \left(1 - \frac{1}{N} + \frac{p_{0,1}}{N}\right)^{M-1}}, \end{aligned} \quad (32)$$

and for the case with two interfaces,

$$\begin{aligned} \mu_2 &= \frac{2}{M} \sum_{m=1}^M m P_I \left(1 - \frac{2}{N}\right)^{m-1} P_m \\ &= 2P_I (1 - p_{0,2}) \left(1 - \frac{2}{N} + \frac{2p_{0,2}}{N}\right)^{M-1}, \\ p_{0,2} &= 1 - \frac{\lambda}{2P_I (1 - p_{0,2}) \left(1 - \frac{2}{N} + \frac{2p_{0,2}}{N}\right)^{M-1}}. \end{aligned} \quad (33)$$

By solving the fixed-point equations for  $p_{0,d}, d = 1, 2$ , as given above, the mean queue lengths of SUs for the two cases can be readily derived.

We next characterize the gain of using two interfaces under this regime. It is worth noting here that the stability conditions should be modified to  $\lambda < P_I$  (for the case with a single interface) and  $\lambda < 2P_I$  (for the case with two interfaces) respectively. Let  $\psi \triangleq \frac{E[L_i]^{(\mathcal{M})}}{E[L_i]^{(c)}}$ , where  $E[L_i]^{(\mathcal{M})}$  and  $E[L_i]^{(c)}$  denote the mean queue lengths of SUs for the cases with a single interface and two interfaces respectively. When  $M$  is fixed and  $N \rightarrow \infty$ , we have that

$$\begin{aligned} \lim_{N \rightarrow \infty} \psi &= \lim_{N \rightarrow \infty} \frac{\alpha_c - \lambda}{\alpha_{\mathcal{M}} - \lambda} \\ &= \lim_{N \rightarrow \infty} \frac{2P_I (1 - p_{0,2}^M) - \lambda}{P_I (1 - p_{0,1}^M) - \lambda} \\ &= \frac{2(1 - (1 - \sqrt{\rho/2})^M) - \rho}{(1 - (1 - \sqrt{\rho})^M) - \rho}, \end{aligned} \quad (34)$$

where  $\rho = \lambda/P_I$  is the traffic intensity.

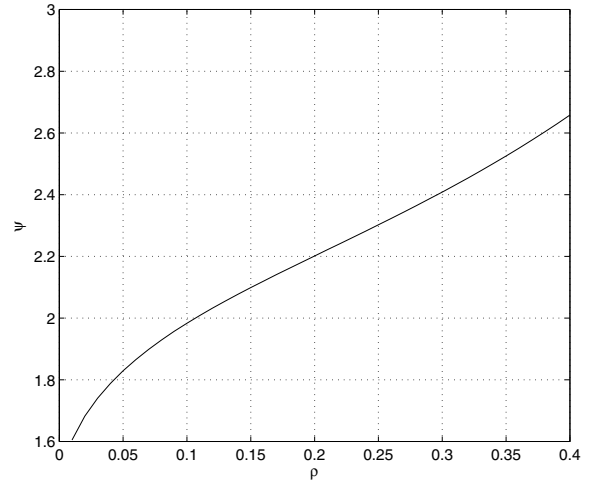


Fig. 6. Gain of using two interfaces

Fig.6 depicts the gain as a function of the traffic intensity. As expected, the application of two interfaces provides significant gain by decreasing the mean queue lengths, and as the traffic intensity grows larger, the gain increases as well.



## VI. CONCLUSIONS AND DISCUSSIONS

In this paper, we have carried out delay analysis for a cognitive radio network. We took a stochastic fluid queue approach and modeled the system using Poisson driven stochastic differential equations. We characterized the moments of the queue lengths of secondary users, for cases with a single PU channel and multiple PU channels. The impact of the PU traffic on SUs' queue lengths and the gain of using multiple PU channels were examined. Also, we explored the gain of using two interfaces per secondary user. Adaptive algorithms, using local information only, have been developed to find the optimal contention probabilities that achieve the minimum mean queue lengths and thus the minimum queueing delays of secondary users.

Our analysis and numerical examples revealed that the mean queue lengths of secondary users increase as the duty cycle of the primary users' traffic increases, pointing to the impact of PU activity on the delay performance of secondary users. Also, when multiple PU channels were employed, we observed a decrease in the mean queue lengths, indicating a multi-channel gain. Moreover, if each secondary user is equipped with two interfaces, there is a decrease in the mean queue lengths because of the gain of using two choices.

Besides the models we studied in this paper, it is of great interest to characterize the delay performance for cases where SUs contend for the PU channels using queue-based random access, e.g., SU  $i$ 's contention probability is set to be

$$p_i(t) = \min \{ \epsilon L_i(t), 1 \}, \quad (35)$$

where  $\epsilon \in (0, 1)$  is a constant. Clearly, the contention probability is time-varying and correlated across users. The dependency on the queue lengths of the contention probabilities introduces a strong coupling across SUs', making it challenging to analyze SUs' queueing lengths and thus the delay performance. Moreover, admission control is another issue under-explored.

In our ongoing work, we are also investigating a strategy where each secondary user tries to exploit temporal diversity when accessing the PU channels. In this case, similar techniques using Poisson-driven stochastic differential equations can be applied to characterize the mean queue lengths of SUs. Moreover, it would be of interest to exploit the temporal diversity and the spatial diversity simultaneously.

## REFERENCES

- [1] "Report of the spectrum efficiency working group," FCC Spectrum Policy Task Force, Tech. Rep., Nov. 2002.
- [2] S. Aalto and W. R. W. Scheinhardt, "Tandem fluid queues fed by homogeneous on-off sources."
- [3] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of a data-handling system with multiple sources," *The Bell System Technical Journal*, vol. 61, no. 8, pp. 1871–1894, 1982.
- [4] T. K. Apostolopoulos and E. N. Protonotarios, "Queueing analysis of buffered slotted multiple access protocols," *Computer Communications*, vol. 8, no. 1, pp. 9–21, 1985.
- [5] Y. Azar, A. Broder, A. Karlin, and E. Upfal, "Balanced allocations," in *Proc. 26th ACM Symposium on the Theory of Computing*, vol. 2, 1994, pp. 593–602.
- [6] D. Bertsekas and R. Gallager, *Data Networks*, 2nd Edition. Prentice Hall, 1992.
- [7] N. Bisnik and A. Abouzeid, "Queueing network models for delay analysis of multihop wireless ad hoc networks," in *Proc. International Conference on Wireless Communications and Mobile Computing'06*, Vancouver, Canada, 2006, pp. 773–778.
- [8] R. W. Brockett, "Lecture notes: stochastic control," Harvard University, Tech. Rep., 1983.
- [9] R. W. Brockett, W. Gong, and Y. Guo, "Stochastic analysis for fluid queueing systems," in *Proc. IEEE Decision and Control'99*, Phoenix, AZ, 1999, pp. 3077–3082.
- [10] A. Ephremides and R.-Z. Zhu, "Delay analysis of interacting queues with an approximate model," *IEEE Transactions on Communications*, vol. 35, no. 2, pp. 194–201, 1987.
- [11] A. Eryilmaz, P. Marbach, and A. Ozdaglar, "A fluid-flow model for backlog-based csma policies," in *Proc. IEEE WICON'08*.
- [12] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," in *Proc. IEEE INFOCOM'05*, vol. 3, Miami, USA, pp. 1794–1803.
- [13] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 517–528, Apr. 2007.
- [14] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in WLAN channels: empirical model and its stochastic analysis," in *Proc. First International Workshop on Technology and Policy for Accessing Spectrum (TAPAS)*, Aug. 2006.
- [15] L. Grotkop and D. N. C. Tse, "Spectrum sharing between wireless networks," in *Proc. IEEE INFOCOM'08*, Arizona, USA, pp. 201–205.
- [16] O. Kella and W. Whitt, "A storage model with a two-state random environment," *Operations Research*, vol. 40, pp. 257–262, 1991.
- [17] A. Laourine, S. Chen, and L. Tong, "Queueing analysis in multichannel cognitive spectrum access: a large deviation approach," in *Proc. IEEE INFOCOM'10*, San Diego, USA, p. to be appear.
- [18] Y. Liu and W. Gong, "On fluid queueing systems with strict priority," *IEEE Transaction on Automatic Control*, vol. 48, no. 12, pp. 2079–2088, Dec. 2003.
- [19] W. Luo and A. Ephremides, "Stability of N interacting queues in random access systems," *IEEE Transaction on Information Theory*, vol. 45, no. 5, pp. 1579–1587, 1999.
- [20] S. Meyn, *Control Techniques for Complex Networks*. Cambridge University Press, 2008.
- [21] D. Miorandi, A. A. Kherani, and E. Altman, "A queueing model for HTTP traffic over IEEE 802.11 WLANs," in *Proc. 16th ITC Specialist Seminar*, vol. 50, no. 1, 2004, pp. 63–79.
- [22] J. Mitola III, "Cognitive radio: an integrated agent architecture for software defined radio," Ph.D. dissertation, KTH Royal Institute of Technology, 2000.
- [23] M. D. Mitzenmacher, "The power of two choices in randomized load balancing," Ph.D. dissertation, University of California at Berkeley, 1996.
- [24] E. Modiano and A. Ephremides, "A method for delay analysis of interacting queues in multiple access systems," in *Proc. IEEE INFOCOM'93*, vol. 2, San Francisco, CA, 1993, pp. 447–454.
- [25] R. R. Rao and A. Ephremides, "On the stability of interacting queues in a multiple access system," *IEEE Transaction on Information Theory*, vol. 34, no. 5, pp. 918–930, 1988.
- [26] E. D. Sykas, D. E. Karvelas, and E. N. Protonotarios, "Queueing analysis of some buffered random multiple access schemes," *IEEE Transactions on Communications*, vol. 34, no. 8, pp. 790–798, 1986.
- [27] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," in *Proc. IEEE INFOCOM'04*, vol. 2, 2004, pp. 1404–1413.
- [28] T. Wan and A. U. Sheikh, "Performance and stability analysis of buffered slotted aloha protocols using tagged user approach," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 582–593, 2000.
- [29] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: structure, optimality, and performance," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5431–5440, 2004.
- [30] H. Zheng and C. Peng, "Collaboration and fairness in opportunistic spectrum access," in *Proc. IEEE ICC'05*, pp. 3132–3136.