Effective Bandwidth for Delay Tolerant Secondary User Traffic in Multi-PU, Multi-SU Dynamic Spectrum Access Networks

Ebrahim Safavi and K.P. Subbalakshmi
Department of Electrical and Computer Engineering, Stevens Institute of Technology
Email: {ssafavi,ksubbala}@stevens.edu

***Abstract*—In this paper we study two important quality of service (QoS) parameters of dynamic spectrum access (DSA) networks with multiple primary and secondary users. We assume that the SUs are delay tolerant and model the PU and SU activity as Markov chains with ON and OFF states. We derive a closed form expression for the probability distribution of buffer occupancy for secondary networks under steady state conditions and formulate an expression for the effective bandwidth under buffer overflow constraints. Extensive simulations are used to validate our analysis.**

***Index Terms*—Dynamic Spectrum Access, Multi-Primary and Multi-Secondary Users, Effective Bandwidth, Buffer occupancy, Asymptotic Tail Behavior.**

I. Introduction

In this paper we address two important QoS parameters of DSA networks: (1) the effective bandwidth and, (2) the tail behavior of the transmission buffer. Effective bandwidth is defined as the maximum reliable bandwidth that a network can provide under some predefined QoS constraints [1] and reflects the efficiency of resource allocation in the network [2]. Here, we investigate the effective bandwidth available to secondary users (SUs) subject to some maximum buffer overflow probability. We assume delay tolerant secondary network traffic, where the traffic generated by SUs is buffered for future service when no networks are available (due to primary network activity or otherwise). We also study the tail behavior of this queue, since it can be an indicator of network performance. For eg., a heavy tailed distribution of the queue length, can lead to network performance degradation [3].

Our work proposes an analytically tractable finite-state Markov model to jointly describe the PU and SU network. We model the cognitive radio network (CRN) consisting of an arbitrary number of SUs and PUs as a continuous time Markov model, where network users can switch ON and OFF at any time [4]–[6]. Also, when a PU turns ON, the channel associated with that PU (primarily licensed to it) is unavailable to SUs. We assume that SUs can dynamically move to another band upon the return of the PU. The buffer behavior for this system is formulated as a stochastic fluid flow model [7]–[16].

A. Related Work

Although DSA networks have been well studied in the past decade, the performance of these networks under QoS constraints are just beginning to be studied [4]–[7], [17]–[20]. Examples of studies on effective bandwidth for wireless networks can be found in [5], [17], [21]–[23]. Studies on QoS parameters of CRN include [5], [7], [10], [24]. Some of these works use a priority model for the PU and SU activity; however, these models don't capture some nuances of DSA networks. For eg., *they do not allow for the fact that a lower priority SU can still transmit when a PU switches ON, by switching to another unoccupied channel. Our current model is able to address this deficiency.*

The effective bandwidth and asymptotic tail behavior of delay distribution for SUs in single-channel DSA networks was analyzed using the law of large numbers in [6]. A large deviation approximation of the queue length distribution as a function of the PU and SU traffic was investigated in [4]. Both concluded that the delay and buffer occupancy distributions for the single-PU networks are light-tailed if the busy period is light-tailed. The asymptotic analysis of the steady-state queue length distribution of SUs for a single PU channel under heavy-tailed network environment is considered in [18]. Unlike [4] and [6] which derive *approximations* for the *tail* end of the distribution in *single-PU* networks, we derive the *exact closed-form expression of the buffer occupancy for a general secondary network comprising multiple channels over the entire range.* We also analyze the asymptotic tail behavior of buffer occupancy distribution of general multi-channel DSA networks and show that *if the busy period distribution for PUs are light-tailed, then the buffer occupancy for multi-SU, multi-PU network is also a light-tailed distribution*.

In [25], the effective capacity for pre-assumed light-tailed distribution of the queue occupancy is analyzed where a *single* channel is shared among PUs and the SU. In contrast to [25], we analytically derive the distribution of the buffer occupancy for the CRN and also analyze the effective bandwidth available to cognitive users that have access to an arbitrary number of channels. The effective capacity for the coexisting CRNs in Nakagami wireless fading channels with respect to some delay constraints is addressed in [26]. In [27], the capacity of opportunistic secondary communication for a network of two independent channels is explored. The achievable rate region for CRNs is studied in [28]. The Gaussian throughput for a CRN coexisting with a primary network is investigated in [22]. The effective capacity for an underlay CRN with a single primary and secondary link subject to the average interference

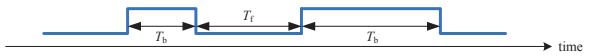


Fig. 1: Each channel is modeled as an ON-OFF process. The busy and free period random variables are denoted by T_b and T_f, respectively.

and delay constraints is addressed in [23], which calculates some bounds on the effective capacity of the model.

Effective bandwidth for a general multi-SU multi-PU DSA networks has not yet been fully explored for dynamic spectrum enabled CRNs. This work studies the effective bandwidth performance for such networks under certain QoS provisioning. More specifically, the contributions of this paper are (1) closed-form expression for the probability distribution of the buffer occupancy for secondary networks under steady state conditions, (2) formulation and expression for the effective bandwidth under buffer overflow constraints, (3) analysis of sufficient conditions for light-tailed buffer occupancy distributions for CRNs with multiple SUs and multiple PUs.

II. System Model

Our system model comprises of an arbitrary number, N_s, of homogeneous SUs that independently switch between active and idle states. The SU generates data at rate c_s bits per second (bps) during the active periods, whether or not it has access to a channel for transmission. The transmission activities of SUs are modeled as independent identically-distributed (i.i.d.) Poisson processes, where the intervals between consecutive events (active and idle states) are i.i.d. exponentially distributed. This model captures the burstiness of the data stream [14].

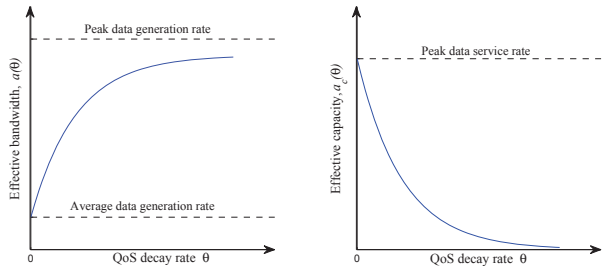
Since a given channel is essentially unavailable to the CRN upon the return of the PU, channels can also be modeled as ON-OFF processes. We model the PU network, comprising N_p homogenous PUs as a continuous time Markov chain (CTMC). Each channel can exist in one of two states: free (when the associated PU is OFF) or busy (when the associated PU is ON). Each channel has c_p bps capacity. The channel free/busy time is shown in Fig. 1. To make the model analytically tractable and still capture the characteristics of real applications in typical environments, the channel busy and free periods are modeled as i.i.d. exponential random variables [29]–[31]. Other measured PU traffic duration distributions on different channel bands are possible [32]; analysis of effective bandwidth for these distributions is deferred for future work.

The channel is busy for a random time period $T_b \propto \exp(\lambda_p)$ and is free for a random time period $T_f \propto \exp(\mu_p)$. The summary of system parameters are identified in Table I.

We assume that there exists a spectrum availability database through which the spectrum availability is perfectly known. This assumption is reasonable since as per the US President Council of Advisors on Science and Technology's report, there is a shift towards a database approach for deciding spectrum availability (away from the spectrum sensing approach) [33]. When sensing based dynamic spectrum access is the predominant alternative, there can be imperfections in the sensing mechanisms leading to imperfect knowledge about spectrum bands. We leave this case for future work.

TABLE I: Notations

N_s	Number of homogeneous SUs
N_p	Number of homogenous PUs
T_f	Channel free period random variable
T_b	Channel busy period random variable
λ_s	Active time exp. dis. parameter for SUs
μ_s	Idle time exp. dis. parameter for SUs
λ_p	Active / busy period exp. dis. parameter for PUs / channels
μ_p	Idle / free period exp. dis. parameter for PUs / channels
c_s	Traffic generation rate for each SU
c_p	Channel capacity for each PU
θ^*	Dominant decay rate for buffer over flow probability distribution
a^*	Effective bandwidth available to each SU
$P_{i,j}(\cdot)$	Cumulative distribution function of buffer occupancy



(a) (b)

Fig. 2: *a*) Typical effective bandwidth, *b*) Typical effective capacity.

For a fixed access policy π, let $Q^\pi(t)$ denote the queue size of the tagged SU at time t. Then, the buffer content of the SU for the next Δt seconds can be modeled as a Lindley process [34] satisfying the following equation [4]:

$$Q^\pi(t+\Delta t) = (Q^\pi(t) - d^\pi(t+\Delta t))^+,\ Q^\pi(0) = 0 \qquad (1)$$

where $d^\pi(\cdot)$ shows the buffer depletion rate for the tagged SU at a given time given by:

$$d^\pi = \begin{cases} c^\pi(t+\Delta t) - c_s\Delta t & \text{active SU} \\ c^\pi(t+\Delta t) & \text{idle SU} \end{cases} \qquad (2)$$

and $(\cdot)^+$ shows the positive portion, $\max(\cdot,0)$. At given time t, the aggregate buffer occupancy, $Q^\pi(t)$ is a measure of the performance of the secondary network.

In our study, we are interested in characterizing the queue performance of the first-in first-out (FIFO) channel access policy when there is perfect knowledge about spectrum bands for SUs determined by the spectrum availability database. From now on, we shall omit the superscript π.

A. Effective Bandwidth and Capacity

In the multi-user dynamic spectrum access network literature, the stochastic behavior of a data arrival process is asymptotically modeled by its effective bandwidth[1]. The effective

[1]The effective bandwidth is referred to as the effective capacity as well in [4], [35]. In this paper we use effective bandwidth for the arrival service and effective capacity for the service process.

bandwidth is defined as the minimum constant service rate required to guarantee a given QoS, θ, which characterizes the decay rate of the buffer overflow probability for the arrival process. We can use large deviations to analyze the asymptotic probability distribution of the buffer and decay rate [36], [37].

For the accumulated arrival process, $\{A(t) : t \geq 0\}$, which shows the amount of data until time t, the asymptotic log-moment generating function of the arrival process is defined as, $\Psi_B(\theta) = \lim_{t\to\infty} \frac{1}{t} \log \mathbf{E}[e^{\theta A(t)}], \theta > 0$, where $\mathbf{E}[\cdot]$ stands for the expected value. Then, the effective bandwidth function of the arrival process is, $a(\theta) = \frac{\Psi_B(\theta)}{\theta}, \forall \theta > 0$.

Effective capacity is the dual of effective bandwidth and for a given cumulative service process, $\{S(t) : t \geq 0\}$, the effective capacity of the service process can be expressed as rate function with respect to θ defined as, $a_c(\theta) = -\frac{\Psi_C(-\theta)}{\theta}, \forall \theta > 0$, where $\Psi_C(\cdot)$ is the Gärtner-Ellis limit of the service process, $S(t)$ defined as, $\Psi_C(\theta) = \lim_{t\to\infty} \frac{1}{t} \log \mathbf{E}[e^{\theta S(t)}], \theta > 0$. The effective capacity, $a_C(\theta)$ is the maximum *constant* arrival rate that can be supported for the QoS decay rate θ [1].

The effective bandwidth theory is a powerful approach to evaluate the capability of a wireless networks to support data traffic with diverse statistical QoS guarantees [35], [38]–[40].

Typical effective bandwidth and effective capacity functions for a queueing system of infinite buffer size is shown in Fig 2. In Fig. 2a, the average data generation rate of the source is $a(0)$, while $a(\infty)$ denotes the peak data generation rate. Fig. 2b shows that increasing QoS decay rate decreases the effective capacity that is, as the QoS requirement becomes more stringent, the source rate that a network can support with this QoS guarantee, decreases. Fig. 2 illustrates the duality between the effective bandwidth and effective capacity [35].

Practically, the QoS decay rate, θ depends on the statistical characterization of the arrival and service processes establishing QoS constraint on the buffer overflow probability. The following lemma characterizes the QoS decay rate.

Lemma 2.1: When the Gärtner-Ellis limit for the stable queue exists, and there is a unique positive solution, θ^*, for:

$$a = -\frac{\Psi_C(-\theta)}{\theta}, \tag{3}$$

then the buffer overflow probability of a given buffer value q is independent of time index, t, and has the following form:

$$\Pr(Q > q) = \mathcal{O}(e^{-\theta^* q}), \tag{4}$$

where $\mathcal{O}(\cdot)$ denotes the Big-O asymptotic notation, a is the maximum sustainable arrival rate (the effective bandwidth), and $\theta^* = \theta^*(a)$ is the decay rate of the overflow probability which is the inverse function of the effective bandwidth of the arrival process (i.e. $a_c(\theta^*) = a$) [4].

Eq. (3) in Lemma 2.1 indicates the equilibrium state of a stable system when the effective bandwidth for the arrival process equals to the effective capacity of the service process. From Eq. (4), we can see that a larger θ^* implies lower buffer overflow probability, which means a more strict QoS constraint for the arrival process whereas lower values of θ^* represent looser QoS constraint for the arrival process constraint.

Throughout this paper, we model the CRN as a queue system where the channel assignment is modeled as service process where the data generated by SUs is served by assigning one channel for transmission (see Fig. 3 and Sec. II-C for more detail on the system model). Therefore, the effective bandwidth for such model indicate the maximum data rate generated by SUs that can be served for the given channel assignment policy. For example, in the CRN with a single primary channel that we analyze in the Section II-B, $A(t)$ is the arrival process for a data source with fix data generation rate which is buffered in the queue and $S(t)$ is the service process for an ON-OFF server illustrated in Fig. 1 (See Section II-B for the detailed analysis).

In CRNs, due to the dependence between arrival and service processes through the sensing mechanism, the effective bandwidth is a function of both the arrival and service processes. In the following, we use the buffer overflow probability of the dependent system as the QoS indicator to evaluate the effective bandwidth. The key point is that for the service process modeled by $\theta^* = \theta^*(a)$ at given buffer bound $q \gg 1$, and QoS constraint $\epsilon \ll 1$, the effective bandwidth a should satisfy the approximation $\epsilon \approx e^{-q\theta^*(a)}$.

The objective of this paper is to obtain the maximum sustainable arrival rate $a^*(\epsilon)$ which, by definition, is the effective bandwidth of the system, calculated for a given value of buffer size, q, and ϵ QoS constraint:

$$a^*(\epsilon) = \arg \sup_a \{\Pr(Q > q) \leq \epsilon\}. \tag{5}$$

The effective bandwidths of a network can determine a CRN's spare capacity to accept more SUs at any time. For instance, say we want to determine if a SU with a new QoS constraint can be accommodated in a CRN that is currently being used by N_s SUs with the network inherent QoS constraint. Through effective bandwidth analysis, we can determine if by accepting the new SU to the CRN, the new QoS constraint is met or not. The concept of effective bandwidth has been used for resource allocation problems in LTE CRNs in [41].

In the following section, we model the arrival and service processes in the CRN. Later, we find the closed-form expression for the probability distribution function (pdf) of the buffer occupancy, to evaluate the effective bandwidth.

B. Cognitive Radio Network with Single Primary Channel

Before proceeding with a general network to explain the method, we consider a simple network consisting of a single SU and a single PU, where the SU is active and always generates data. Let $P_{\mathrm{f}}(q,t)$ be the cumulative distribution function (CDF) of the buffer occupancy at given time t when the channel is in the free state, that is $P_{\mathrm{f}}(q,t) = \Pr\{Q(t) \leq q, T_{\mathrm{f}}\}$, where $Q(t)$ denotes the buffer content at time t and T_{f} is period for which the channel is free, as defined earlier. The CDF when the channel is busy, $P_{\mathrm{b}}(q,t)$, also can be defined in the same way. The distribution parameters for active and idle periods for the PU are λ_p and μ_p, respectively. Therefore, the probability that the PU switches from the idle state to the busy state (channel state switches from free to busy) in the next infinitesimally small Δt seconds is $\mu_p \Delta t$ [31]. The probability that no transition happens is $1 - \mu_p \Delta t$. In this case, the amount

of depletion of the buffer content is $(c_p - c_s)\Delta t$. Thus, the buffer occupancy CDF for the next Δt sec., $P_{\rm f}(q,t+\Delta t)$, is:

$$\lambda_p \Delta t P_{\rm b}(q,t) + (1-\mu_p\Delta t)P_{\rm f}(q+(c_s-c_p)\Delta t,t) \quad (6)$$

Similar equation can be written for the buffer occupancy CDF when the PU is in the busy state. Here due to the lack of the channel access, the buffer content increase is $c_s\Delta t$, therefore

$$P_{\rm b}(q,t+\Delta t) = \mu_p \Delta t P_{\rm f}(q,t) + (1-\lambda_p\Delta t)P_{\rm b}(q+c_s\Delta t,t) \quad (7)$$

Solving Eq. (6) and (7) for infinitesimally small Δt, we have:

$$(c_p - c_s)\frac{\partial P_{\rm f}}{\partial q} + \frac{\partial P_{\rm f}}{\partial t} = \lambda_p P_{\rm b} - \mu_p P_{\rm f} \quad (8)$$

and

$$-c_s\frac{\partial P_{\rm b}}{\partial q} + \frac{\partial P_{\rm b}}{\partial t} = \mu_p P_{\rm f} - \lambda_p P_{\rm b} \quad (9)$$

In our analysis, we are interested in steady state behavior, hence the probability distribution has no time variation and we can drop time dependency index. To uniquely determine the solution for the set of first order differential equations in Eq. (8) and (9), the knowledge of boundary condition is necessary. When there is no constraint on the buffer content (i.e. $q=\infty$), the probability of being in the free or busy states is $P_{\rm f}(\infty) = \frac{\lambda_p}{\lambda_p+\mu_p}$ and $P_{\rm b}(\infty) = \frac{\mu_p}{\lambda_p+\mu_p}$. In addition, in the busy state, when no channel is available, the buffer should always be non-empty, i.e. $P_{\rm b}(0)=0$.

Using the boundary conditions, the buffer occupancy CDF in free and busy states are the summation of exponentials $P_{\rm f}(q) = \alpha_1 e^{-\theta q} + \alpha_2$ and $P_{\rm b}(q) = \beta_1 e^{-\theta q} + \beta_2$, where:

$$\theta = \frac{\lambda_p}{c_s} - \frac{\mu_p}{c_p - c_s} \quad (10)$$

and $\alpha_1 = \frac{c_s\mu_p}{(c_p-c_s)(\mu_p+\lambda_p)}$ and $\alpha_2 = \frac{\lambda_p}{\lambda_p+\mu_p}$, and also $\beta_1 = -\beta_2 = -\frac{\mu_p}{\mu_p+\lambda_p}$. To have bounded CDF functions, the decay rate, θ, should be positive which results, $c_s < c_p\frac{\lambda_p}{\lambda_p+\mu_p}$, where c_s is the input rate and the RHS is the average service rate. This inequality is the stability condition. Using Bayes' rule, the buffer occupancy CDF is $P(q) = P_{\rm b}(q) + P_{\rm f}(q)$. Eq. (10) shows that when $c_s \to c_p\frac{\lambda_p}{\lambda_p+\mu_p}$, the buffer occupancy CDF decays very slowly with respect to the buffer size.

The buffer overflow probability is $\Pr(Q>q) = 1-P(q)$. Hence, we have:

$$\Pr(Q>q) = \frac{\mu_p}{\mu_p+\lambda_p}\frac{c_p}{c_p-c_s}e^{-\theta q} \quad (11)$$

Following Eq. (5), for the buffer size q and the QoS constraint ϵ, the effective bandwidth, a^* can be calculated as the solution to the nonlinear equation $\Pr(Q>q)=\epsilon$. Finding a closed-form expression for this equation is very difficult, however, we can study some special cases. For example, for small value of buffer size, q, the effective bandwidth is significantly smaller than channel PU's channel capacity, $a^* \ll c_p$. Therefore, the effective bandwidth is given by, $a^*_{a^*\ll c_p} \approx \frac{\lambda_p c_p q}{\mu_p q - c_p \log\epsilon'}$, where $\epsilon' = \epsilon(1+\lambda_p/\mu_p)$. Therefore, the maximum rate that the SU can use to transmit its data with the buffer size q at the maximum buffer overflow probability ϵ is when $c_s = a^*$. However, for a large buffer size, $q \gg 1$, the effective bandwidth is the maximum possible transmit rate for the stable system i.e., $\lim_{q\to\infty} a^* = c_p\frac{\lambda_p}{\lambda_p+\mu_p}$.

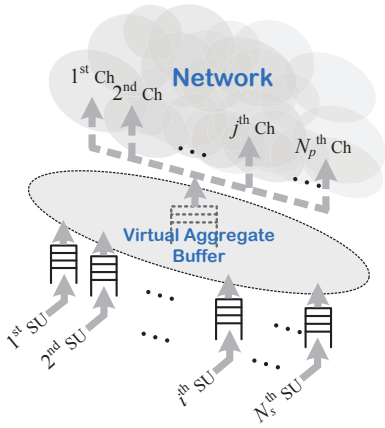


Fig. 3: Each channel is associated with a PU and is available to the SU only when the channel is unused by the corresponding PU. The data generated by the SU is placed in its own buffer. If a PU returns to the system, the SU can switch to the other available channels.

The analysis carried out in the case of single PU can be generalized to the case $N_p > 1$. In what follows, we extend our analysis to the general case when $N_s, N_p \geq 1$ and SUs can also switch between active and idle states.

C. General Cognitive Radio Network

We again model the system as a CTMC. Let the number of SUs in the active state, at time t, be i and the number of idle channels be j; the network system state is represented by the two-tuple $s=(i,j)$. We can write the state space, $\mathcal{S} = \{s=(i,j)|i\in[0,N_s], j\in[0,N_p]\}$. The total number of states is $|\mathcal{S}| = (N_s+1)\times(N_p+1)$. The corresponding distribution parameters for active and idle periods for SUs are λ_s and μ_s, respectively. Therefore, the average time for the active and idle periods for an SU are λ_s^{-1} and μ_s^{-1}, respectively.

When the PU turns active, the corresponding channel (licensed to the PU) is unavailable for use by any SU. However, an SU in that band can look for and occupy any vacant channel upon the return of the PU to its current channel. If any SU generates data when there is no available channel, the data traffic goes into a buffer associated with that SU (see Fig. 3).

Each SU buffers its data stream until it gains access to the transmission channel in accordance with the FIFO rule i.e., each SU should wait until all prior channel requests are processed. Therefore, we can model the entire system of buffers as a single virtual aggregate buffer, $Q(\cdot)$, where the content of the aggregate buffer is the summation over the content of all individual SU buffers (see Fig. 3). The aggregate buffer depletion rate at any state, $d_{i,j}$ depends on the number of free channels and active SUs. The buffer depletion rate at any state and time is:

$$d_{i,j} = jc_p - ic_s. \quad (12)$$

In Eq. (12), we assume that SUs can use more than one channel at a time to use all the available capacity upon their access.

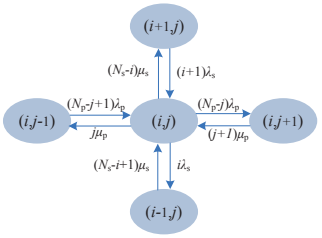


Fig. 4: The state transition diagram in steady state for the CTMC model.

This can be done by carrier aggregation as discussed in [42]–[44]. For the extension of our system model and analysis to the case where aggregation and fragmentation of spectrum bands is not allowed, we can generalize the buffer depletion rate at a given state as, $d_s = f^{(1)}(s)c_p - f^{(2)}(s)c_s$, where, $f^{(1)}(\cdot)$ and $f^{(2)}(\cdot)$ are functions of the network state $s = (i,j)$. For eg., the special case of no channel aggregation can be studied by assuming $f^{(1)}(s) = \min(i,j)$ and $f^{(2)}(s) = i$. Throughout this paper, we use Eq. (12) for our analysis. The study of the generalized buffer depletion rate when there is no channel fragmentation or aggregation is left for future work.

Let $P_{i,j}(q,t)$, $i \in [0,N_s]$, $j \in [0,N_p]$ and $t,q \geq 0$, be the CDF of the probability that the total content of all SU buffers does not exceed q at the time t when i SUs are in active and j PUs are in idle states i.e., $P_{i,j}(q,t) = \Pr\{Q(t) < q \text{ and } s = (i,j)\}$. As long as the buffer is not empty, the aggregate buffer is depleted at the instantaneous rate $d_{i,j}$. Once all buffers are empty, they stay in the empty state as long as the combined data rate of all SUs is less than the available channel capacity, that is $ic_s < jc_p$. The CDF of aggregate buffer occupancy, $P(\cdot)$, can be expressed as a summation of $P_{i,j}$'s:

$$P(q,t) = \sum_{i \leq N_s, j \leq N_p} P_{i,j}(q,t) \tag{13}$$

We assume an infinite buffer for all SUs. For such a system, the stability condition is satisfied if the long-time average arrival rate, D_{in}, for the entire secondary network is less than the long-time average service rate, D_{out}. It is easy to show that the network average arrival rate is $D_{\text{in}} = N_s c_s \frac{\mu_s}{\lambda_s+\mu_s}$, and the average service rate is $D_{\text{out}} = N_p c_p \frac{\lambda_p}{\lambda_p+\mu_p}$. We define the traffic load factor, $\rho = \frac{D_{\text{in}}}{D_{\text{out}}}$. The system is said to be stable if the traffic load factor, ρ is less than unity.

The assumption of the CTMC system model allows us to characterize the aggregate buffer occupancy probability distribution using Kolmogorov's backward equations. Since the periods of idle and active states for the SUs are exponentially distributed, the probability that for the next Δt time slot, an SU turns ON or OFF can be written as $(N_s - i)\mu_s \Delta t + O(\Delta t^2)$ and $i\lambda_s \Delta t + O(\Delta t^2)$, respectively, where $O(\cdot)$ is the Big-O asymptotic notation. Similarly, for the PUs, these probabilities are $j\mu_p \Delta t + O(\Delta t^2)$ and $(N_p - j)\lambda_p \Delta t + O(\Delta t^2)$, respectively. The probability of any compound event is $O(\Delta t^2)$. The probability that there is no change in the states of the PU or SU is given by $1 - ((N_s - i)\mu_s + i\lambda_s + j\mu_p + (N_p - j)\lambda_p)\Delta t + O(\Delta t^2)$. Assuming Δt is infinitesimally small, we can ignore higher order terms, $O(\Delta t^2)$, and use CTMC to derive the following partial differential equations [9].

$$\frac{\partial P_{i,j}}{\partial t} - d_{i,j}\frac{\partial P_{i,j}}{\partial q} = (N_s - i + 1)\mu_s P_{i-1,j} + \tag{14}$$
$$(i+1)\lambda_s P_{i+1,j} + (j+1)\mu_p P_{i,j+1} + (N_p - j + 1)\lambda_p P_{i,j-1}$$
$$-((N_s - i)\mu_s + i\lambda_s) + j\mu_p + (N_p - j)\lambda_p)P_{i,j}$$

where $P_{i,j} = P_{i,j}(q,t)$. At steady state, the probability distribution does not change with respect to time ($\partial P_{i,j}/\partial t \to 0$). Therefore, the first LHS term in Eq. (14) will vanish at steady state, and hereafter, we can drop the time index t, for probability functions. The state transition diagram is shown in Fig. 4. Using matrix notation, we can reformulate Eq. (14) as:

$$\mathbf{A}\frac{d}{dq}\mathbf{P}(q) = \mathbf{B}\mathbf{P}(q) \tag{15}$$

where vector $\mathbf{P}(\cdot) = [P_{0,0}(\cdot), \cdots, P_{i,j}(\cdot), \cdots, P_{N_s,N_p}(\cdot)]'$, and its dimension is the same as the state-space dimension, $(N_s+1) \times (N_p+1)$. The matrix $\mathbf{A} = \text{diag}\{d_{i,j}\}_{s=(i,j)\in\mathcal{S}}$ is the depletion diagonal matrix, and $d_{i,j}$'s are given in Eq. (12), and $\mathbf{B} = [b_{s_1 \to s_2}]_{s_1,s_2 \in \mathcal{S}}$ denotes the transition rate matrix of the CTMC model where non-zero transition rates from state $s_1 = (i,j)$ to $s_2 = (i',j')$, $b_{s_1 \to s_2}$ can be derived based on the rate transition diagram in Fig. 4.

Let z be an eigenvalue of $\mathbf{A}^{-1}\mathbf{B}$ and $\boldsymbol{\phi}$ be the associated right eigenvector i.e., $z\mathbf{A}\boldsymbol{\phi} = \mathbf{B}\boldsymbol{\phi}$. Since the primary and the secondary network are independent of each other, $\boldsymbol{\phi}$, the system eigenvector, has Kronecker decomposition of the form $\boldsymbol{\phi}_s \otimes \boldsymbol{\phi}_p$, where $\otimes$ denotes the Kronecker-product, and $\boldsymbol{\phi}_s$ and $\boldsymbol{\phi}_p$ are eigenvectors of the corresponding independent secondary and primary networks and can be explicitly found using generating polynomials [9]. In general, the transition rate matrix can be written as, $\mathbf{B} = \mathbf{B}_s \oplus \mathbf{B}_p$, where $\mathbf{B}_s$ and $\mathbf{B}_p$ are transition rate matrix of the corresponding independent secondary and primary networks, respectively and $\oplus$ denotes the Kronecker sum [9]. Although we assume homogeneous SUs and PUs in this paper, the analysis presented here can easily be extended to non-homogenous networks as well by categorizing SUs to K_s and PUs to K_p different classes. In such cases the transition rate matrix for SUs and PUs can be substituted with $\mathbf{B}_s = \oplus_{k=1}^{K_s}\mathbf{B}_s^{(k)}$ and $\mathbf{B}_p = \oplus_{k=1}^{K_p}\mathbf{B}_p^{(k)}$, where $\mathbf{B}_s^{(k)}$ or $\mathbf{B}_p^{(k)}$ denote the transition rate matrix for a given class of SUs or PUs, respectively. For example, a class-based approach has been used to analyze network capacity and resource management of wireless home networks in [14].

The bounded solution of the stable system characterized by the equation in Eq. (15) can be expressed as:

$$\mathbf{P}(q) = a_0\boldsymbol{\phi}_0 + \sum_{z_l<0} a_l e^{z_l q}\boldsymbol{\phi}_l \tag{16}$$

in which z_l's and $\boldsymbol{\phi}_l$'s are eigenvalues and eigenvectors of matrix $\mathbf{A}^{-1}\mathbf{B}$. The corresponding coefficient is denoted by a_l. To uniquely determine the solution of Eq. (15), initial conditions should be specified. If $d_{i,j} \geq 0$ then the depletion rate in aggregate buffer content is positive and therefore, at steady state, the buffer should be empty. In other words, the conditional probability of empty buffer content, given

$s = (i, j)$ is $\Pr\{q = 0 | s = (i, j)\} = 1$. Therefore, the stationary probability is:

$$P_{i,j}(0) = \pi_i^{\text{su}} \pi_j^{\text{pu}}, \quad d_{i,j} \geq 0 \tag{17}$$

where π_i^{su} is the probability i out of N_s SUs are simultaneously in the active state and π_j^{pu} is the probability that j out of N_p channels are free and can be calculated as:

$$\pi_i^{\text{su}} = \binom{N_s}{i} \frac{\lambda_s^i \mu_s^{N_s - i}}{(\lambda_s + \mu_s)^{N_s}}, \; \pi_j^{\text{pu}} = \binom{N_p}{j} \frac{\mu_p^j \lambda_p^{N_p - j}}{(\lambda_p + \mu_p)^{N_p}} \tag{18}$$

where $\binom{N_s}{i}$ is the total number of i-combinations of N_s objects. On the other hand, if the data rate of the active SUs exceeds the available channel capacity, under steady state, the buffer content increases and the buffer cannot stay empty i.e.,

$$P_{i,j}(0) = 0, \quad d_{i,j} < 0. \tag{19}$$

The other boundary condition that can be used to determine the solution to Eq. (15) is infinitely large buffer size ($q = \infty$) where, there is no constraint on the buffer content. In this case, the only non-zero term in Eq. (16) is $\mathbf{P}(\infty) = a_0 \boldsymbol{\phi}_0$. For this asymptotic case, $P_{i,j}(\infty)$'s can easily be determined as, $P_{i,j}(\infty) = \pi_i^{\text{su}} \pi_j^{\text{pu}}$. Therefore, the total number of unknown equals the number of equations in Eq. (17) and (19), which will enable us to uniquely determine coefficients, a_l's.

Solving Eq. (16), the aggregate buffer occupancy CDF can be found from Eq. (13) and expressed as, $P(q) = 1 - \sum_l a_l' e^{z_l q}$, where $a_l' = -a_l(\mathbf{1} \cdot \boldsymbol{\phi}_l)$, $\mathbf{1}$ denotes the unit vector and $\cdot$ is the inner product operator.

Result 2.2: The aggregate buffer overflow probability is a weighted summation of exponentials for different systems and can be expressed as:

$$\Pr\{Q > q\} = \sum_l a_l' e^{z_l q} \tag{20}$$

Eq. (20) shows that the dominant decay rates, θ^*, can be determined using the system eigenvalues as:

$$\theta^* = -\max_l \{z_l : z_l < 0\} \tag{21}$$

Formulation in Eq. (20) is one of the main contributions of this paper and gives the closed-form expression for the pdf of the secondary network buffer occupancy.

III. Asymptotic analysis of buffer occupancy distribution

In this section, we provide some insight into the asymptotic behavior of the secondary network buffer occupancy and analyze the tail of distribution.

Result 3.1: The buffer occupancy distribution is light-tailed[2] if *the channel busy period distribution is light-tailed*.

This can be seen by using the following argument. For the secondary network, we consider the worst case scenario when all SUs are always active by letting $\lambda_s \to 0$. The network buffer overflow probability for any other type of SU activity is smaller for any buffer content value, q.

[2]A random variable, X, is light-tailed if there is some $\theta > 0$ such that $\lim_{x \to \infty} e^{\theta x} \Pr\{X > x\} = 0$, and if there is no such θ, it is called heavy-tailed random variable.

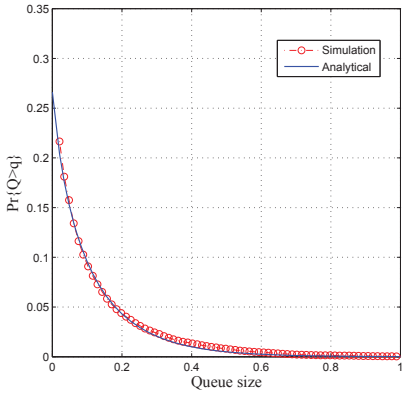


Fig. 5: A comparison of simulation and analytic results, for 5 SUs and 3PUs. Activities are Poisson distributed.

For a system with the same busy state distribution, the probability that the channel is free for a longer time, is greater for the system with a heavy tailed distribution (T_{f}^{HT}) for the free periods compared to a system with a light tailed distribution (T_{f}^{LT}) for the free periods. Therefore, the system with T_{f}^{HT} can transmit more often and has less content in the buffer compared to the system with T_{f}^{LT}. That is, when the means are the same:

$$\Pr\{Q > q, T_{\text{f}}^{\text{HT}}\} \leq \Pr\{Q > q, T_{\text{f}}^{\text{LT}}\} \tag{22}$$

Based on the definition, for any light-tailed free period distribution, T_{f}^{LT}, we have:

$$\Pr\{T_{\text{f}}^{\text{LT}} > t\} \leq e^{-\theta_{\text{f}} t}, \quad \theta_{\text{f}} > 0. \tag{23}$$

For any light-tailed busy period distribution, we can also find an exponential distribution which satisfies:

$$\Pr\{T_{\text{b}}^{\text{LT}} > t\} \leq e^{-\theta_{\text{b}} t}, \quad \theta_{\text{b}} > 0. \tag{24}$$

Therefore, from Eq. (22), (23) and (24), for any queue system with light-tailed busy period distribution and a given free period distribution, there is a queue system with exponentially distributed busy and free periods such that the former queue system outperforms the latter system. Then, we have:

$$\Pr\{Q > q, T_{\text{b}}^{\text{LT}}\} \leq \Pr\{Q > q, T_{\text{b}} \propto \exp(\theta_{\text{b}}), T_{\text{f}} \propto \exp(\theta_{\text{f}})\} \tag{25}$$

We already showed in Sec. II-C that for any system with free and busy periods distributed exponentially, the network buffer overflow probability is a summation of exponentials which is a light-tailed distribution. Substituting Eq. (20) in Eq. (25), we have the following inequality:

$$\lim_{q \to \infty} e^{\theta q} \Pr\{Q > q, T_{\text{b}}^{\text{LT}}\} \leq \lim_{q \to \infty} e^{\theta q} \sum_l a_l' e^{z_l q} \tag{26}$$

The RHS expression in Eq. (26) is zero for any $\theta < \theta^*$ (where θ^* is given in Eq. (21)), which implies that *the buffer occupancy has a light-tailed distribution as long as the channel busy state is light-tailed distributed*.

IV. Effective Bandwidth

In general, finding closed-form expression for the effective bandwidth in Eq. (5) is difficult and the overflow probability in Eq. (20) should be numerically evaluated. However, in some special cases good approximations can be made.

A. Effective Bandwidth for Single Channel and Multiple SUs

In the following, we consider the single channel case when the channel is always available ($\mu_p \to 0$). The dominant decay power in this case can be calculated analytically and is:

$$\theta^*_{N_s} = \frac{N_s(N_s c_s \mu_s - c_p(\lambda_s + \mu_s))}{c_p(N_s c_s - c_p)} \tag{27}$$

This can be seen from the following argument. Any eigenvalue z and eigenvector $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_{N_s})$ for the system described in Sec. II should satisfy the following equation [8]:

$$z(ic_s - c_p)\phi_i = \mu_s(N_s + 1 - i)\phi_{i-1} - ((N-i)\mu_s + i\lambda_s)\phi_i \tag{28}$$
$$+(i+1)\lambda_s\phi_{i+1}$$

The eigenvector generating power series can be defined as $\phi(x) = \sum_i \phi_i x^i$. Eq. (28) can be written in differential form,

$$\phi(x)(zc_p - N_s\mu_s + N\mu_s x) = \phi'(x)(\mu_s x^2 + (zc_s + \lambda_s - \mu_s)x - \lambda_s). \tag{29}$$

The solution to the differential Eq. (29) can be expressed as:

$$\phi(x) = (x - x_1)^k (x - x_2)^{N_s - k} \tag{30}$$

where $k \leq N_s$ is an integer, and

$$x_{1,2} = \frac{-(c_s z + \lambda_s - \mu_s) \pm \sqrt{(c_s z + \lambda_s - \mu_s)^2 + 4\mu_s\lambda_s}}{2\mu_s}. \tag{31}$$

The dominant decay power is the solution to $zc_p - N_s\mu_s + N_s\mu_s x_1 = 0$, which results in the following equation:

$$z(4c_p^2 - 4N_s c_p c_s) + (4N_s^2 c_s\mu_s - 4N_s c_p\lambda_s - 4N_s c_p\mu_s) = 0 \tag{32}$$

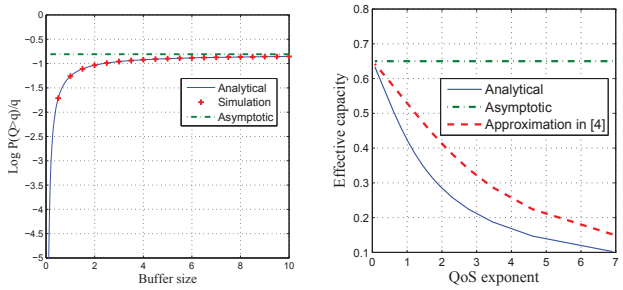
The solution to the Eq. (32) gives the smallest eigenvalue. From this, the dominant decay power can be found is $z = \frac{N_s(N_s c_s \mu_s - c_p(\lambda_s + \mu_s))}{c_p(N_s c_s - c_p)}$. The probability distribution in Eq. (20) can be reformulated as:

$$\Pr\{Q > q\} = a_{\theta^*} e^{-\theta^* q} + \sum_{z_l \neq -\theta^*} a_l' e^{z_l q} \tag{33}$$

where a_{θ^*} is the coefficient corresponding to the dominant decay rate. To have a closed-form expression for the effective bandwidth, we can approximate this coefficient as $a_{\theta^*} \approx 1$ [45]. When $q \gg 1$, we can also ignore the RHS summation in Eq. (33). Using Eq. (4), for the QoS constraint ϵ and buffer size q, the effective bandwidth can be approximated as:

$$a^*_{N_s}(\epsilon) \approx \frac{N_s c_p(\mu_s + \lambda_s) - c_p^2 \log\epsilon/q}{N_s(N_s\mu_s - c_p \log\epsilon/q)} \tag{34}$$

Eq. (34) shows that the effective bandwidth is $a^*_{N_s} = c_p \frac{\mu_s + \lambda_s}{N_s \mu_s}$ when $q \to \infty$, which is the maximum data generation rate for the stable system when the load factor is unity, $\rho = 1$.



(a) (b)

Fig. 6: Single SU, single PU network, *a*) Decay rate for buffer occupancy probability, *b*) Numerical evaluation of Eq. (11) for overflow probability fixed at $\epsilon = 10^{-3}$.

B. Effective Bandwidth for Single SU and Multiple Channel

This case is the dual of the problem considered in Sec. IV-A. When there are N_p channels and the SU is always active ($\lambda_s \to 0$), using the same approach as Sec. IV-A, the dominant decay rate is:

$$\theta^*_{N_p} = \frac{N_p(N_p c_p \lambda_p - c_s(\lambda_p + \mu_p))}{c_s(N_p c_p - c_s)} \tag{35}$$

This can be seen from the following argument. Using the same method as Sec. IV-A, the generating power series can be written in the form of Eq. (30), where $x_{1,2}$ are

$$\frac{(c_p z + \lambda_s - \mu_s) \mp \sqrt{(c_p z + \lambda_p - \mu_p)^2 + 4\mu_p\lambda_p}}{2\lambda_s}. \tag{36}$$

In this case, the dominant decay power is the solution to $zc_s - N_p\lambda_p + N_p\lambda_p x_1 = 0$. By substitution x_1, the dominant decay power can be found as:

$$z = \frac{N_p(N_p c_p \lambda_p - c_s(\lambda_p + \mu_p))}{c_s(N_p c_p - c_s)} \tag{37}$$

For $N_p = 1$, the decay rate can be simplified to $\theta^*_{N_p=1} = \frac{\lambda_p}{c_s} - \frac{\mu_p}{c_p - c_s}$, which was also calculated in Eq. (10). Using the Eq. (33), the effective bandwidth can be approximated as:

$$a^*_{N_p}(\epsilon) \approx \frac{N_p}{2}(c_p + (\lambda_p + \mu_p)q/\log\epsilon) \tag{38}$$
$$+\frac{N_p}{2}\sqrt{(c_p + (\lambda_p + \mu_p)q/\log\epsilon)^2 - 4c_p\lambda_p q/\log\epsilon)}$$

It can be easily verified that for large buffer size ($q \to \infty$) the effective bandwidth will be:

$$a^*_{N_p} \to N_p c_p \frac{\lambda_p}{\lambda_p + \mu_P} \tag{39}$$

Note that the asymptotic effective bandwidth for the network with single SU and single PU with the channel capacity $c_p' = N_p c_p$ is equal to Eq. (39); however, the decay rate is N_p times slower and using Eq. (35), it can be expressed as

$$\theta^*_{N_p}{}' = \frac{N_p c_p \lambda_p - c_s(\lambda_p + \mu_p)}{c_s(N_p c_p - c_s)} = \frac{1}{N_p}\theta^*_{N_p} \tag{40}$$

This is also the case for the decay rate of a network comprising a single PU and a single SU with a generation rate $c_s' = N_s c_s$, say $\theta^*_{N_s}{}'$ as we have $\theta^*_{N_s}{}' = \frac{1}{N_s}\theta^*_{N_s}$.

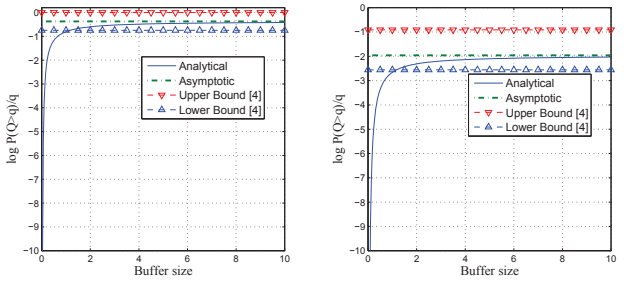


Fig. 7: The dominant decay rate, $\frac{\log P(Q>q)}{q}$ vs Buffer size for the network consisting of a single SU and two PUs for, *a*) $p_{11} = 0.4$ and $p_{01} = 0.6$, *b*) $p_{11} = 0.8$ and $p_{01} = 0.2$.

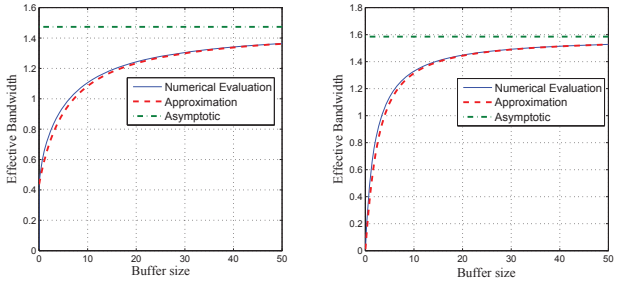


Fig. 8: Effective bandwidth for systems with, *a*) single channel, multiple SUs (approximation error < 1% for $q > 10$), *b*) single SU, multiple PUs (approximation error < 2% for $q > 10$).

C. General Case $N_s, N_p \geq 1$

For the general case, finding a closed-form expression for the effective bandwidth is very difficult. However, the dominant decay rate is a root of equation [9],

$$N_s\sqrt{P_s(\theta)} + N_p\sqrt{P_p(\theta)} - N_p(c_p\theta + \mu_p + \lambda_p) \qquad (41)$$
$$-N_s(-c_s\theta + \mu_s + \lambda_s) = 0,$$

where $P_p(\theta) = (c_p\theta + \mu_p - \lambda_p)^2 + 4\mu_p\lambda_p$ and also $P_s(\theta) = (c_s\theta - \mu_s + \lambda_s)^2 + 4\mu_s\lambda_s$. After finding θ^*, we use the same approximation as Section IV-A and IV-B. Therefore, Eq. (33) and (41) implicitly express the relationship between the effective bandwidth and the overflow probability, which must be evaluated numerically. For the special case of large buffer size ($q \to \infty$), the maximum sustainable rate happens when $\rho = 1$. Then the effective bandwidth is $a^*_{q\to\infty} = \frac{N_p c_p \lambda_p}{N_s \mu_s} \frac{\lambda_s + \mu_s}{\lambda_p + \mu_p}$.

V. Numerical Results

The simulation parameters are identified in Table II. Fig. 5 shows that the experimental and theoretical values for buffer overflow probability $\Pr\{Q > q\}$, match.

Fig. 6a shows the dominant decay rate of the buffer overflow probability single SU single PU network analyzed in Sec. II-B for fixed packet generation rate ($\rho = 0.7$). It demonstrates that $\log \Pr(Q > q)/q$ converges to $-\theta^*$ as the buffer size increases. It can be seen that the approximation of the asymptotic decay in Eq. (35) is true even for moderate buffer sizes.

The plot in Fig. 6b shows the effective capacity of the stable system Eq. (3) versus the QoS exponent, $-\frac{\log \epsilon}{q}$, which is similar to the typical effective capacity plot in Fig. 2b. In [4], the system of a single (always ON) SU with two-state channels is modeled as Markov chain with transition probabilities, $p_{11} = \Pr[\text{free} \to \text{free}]$ and $p_{10} = \Pr[\text{free} \to \text{busy}]$. For the sake of comparison, we study the special case of $p_{11} = p_{01} = \frac{\lambda_p}{\mu_p + \lambda_p}$ and $\lambda_s \to \infty$ to compare our results with the result in [4]. In Fig. 6b, transition probabilities for the Markov chain are $p_{11} = 0.65$ and $p_{10} = 0.35$. In [4], using the approximation $\Pr(Q > q) \approx e^{-\theta^* q}$, the asymptotic effective bandwidth has been provided. From Fig. 6b, it can be seen that the approximation provided in [4] follows the analytical results however, the approximation error increases as the QoS exponent increases.

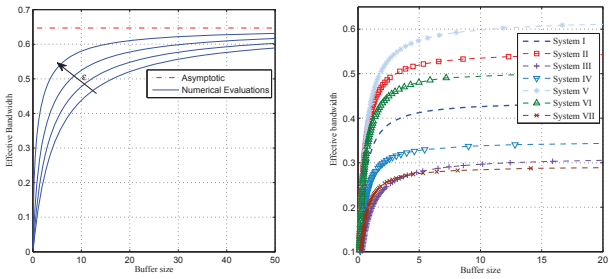
TABLE II: Simulation parameters

	Fig. 5	Fig. 6	Fig. 7a / 7b	Fig. 8a	Fig. 8b	Fig. 9a
N_s	5	1	1	3	1	5
N_p	3	1	2	1	5	3
λ_s	2	-	-	1.2	-	1.2
μ_s	10	-	-	0.5	-	0.5
λ_p	6.7	0.6	0.4 / 0.8	-	0.7	0.7
μ_p	4.7	0.3	0.6 / 0.2	-	1	1
c_s	1	0.4	0.3 / 1.4	-	-	-
c_p	2.5	1	1	1.3	0.7	0.8
ϵ	-	10^{-3}	-	10^{-3}	10^{-3}	-

The dominant decay rate for the single-SU network with different simulation setups are shown in Fig 7. Fig. 7a shows the dominant decay rate for the system of two PUs with active and idle rates $\lambda_p = 0.4$ and $\mu_p = 0.6$, respectively. To compare results with the study in [4], we set $\lambda_s \to \infty$ (always active SU) and set the transition probabilities $p_{11} = 0.4$ and $p_{01} = 0.6$. Fig. 7b shows the decay rate for the network with transition probabilities $p_{11} = 0.8$ and $p_{01} = 0.2$. Fig. 7 verifies our results and shows that our analytical formulations perfectly fit the bounds derived in [4].

In Fig. 8a, the effective bandwidth available to each SU for the system of always available single channel with three SUs is plotted. As it can be seen from Fig. 8a, the available effective bandwidth for the zero-size buffer is $a^*_{q=0} = 0.43$ which is the ratio of the channel capacity to the number of SUs, c_p/N_s. However, in Fig. 8b, for the system of multiple PUs where channels alternate between free and busy states, the zero-size buffer effective bandwidth is zero, implying that to have a constant data rate, a non-zero buffer is necessary. Fig. 8b shows the effective bandwidth for the network of a single SU with five PUs. Fig. 8 shows that effective bandwidth formulation in Eq. (27) and (38) are very good approximations (less than 1% error for large buffer sizes).

The effective bandwidth for a general network consisting of multiple SUs and multiple PUs (5 SUs and 3 PUs) is shown in Fig. 9a which depicts the impact of increasing the QoS constraint ϵ on the effective bandwidth for $\epsilon = 10^{-i}, i = \{1,2,3,4\}$. As we expect, the more strict QoS constraint, the lower effective bandwidth for the same buffer size. However, as it



(a) (b)

Fig. 9: The impact of system parameters on the effective bandwidth for, *a*) the system of multiple SUs, multiple channels for different ϵ value, *b*) given QoS constraint $\epsilon = 10^{-3}$.

can be seen, all curves converge to the asymptotic effective bandwidth value for large buffer sizes.

Fig. 9b shows effective bandwidth for different buffer sizes ($\epsilon = 10^{-3}$, $c_p = 1$). Table III shows the simulation parameters. In System II, the average channel free period ($\overline{T_f} = 1/\mu_p$) is doubled, and as we expect, the effective bandwidth for SUs increases, however, in System III, where the average channel busy period ($\overline{T_b} = 1/\mu_p$) is doubled, the effective bandwidth decreases. The opposite argument holds for the activity periods of SUs i.e., in the System IV where the average SU active time is doubled, the effective bandwidth decreases while, doubling the average SU idle time in System V, increases the effective bandwidth. System VI and VII show the impact of decreasing number of SUs (N_s) and increasing number of PUs (N_p), respectively, on the effective bandwidth. Note rate of convergence to the asymptotic values are different for the different systems. For example, the the fastest System (System II) reaches 95% of its asymptotic value for buffer size $q = 5.7$ while, the slowest System (System III) reaches that value at buffer size ($q = 11.8$).

VI. Conclusion and Future Work

We analyzed the effective bandwidth for a general CRN consisting of multiple PUs and multiple SUs, by defining a stochastic fluid flow model for the SU buffer occupancy and a CTMC to capture the dynamic nature of the PU and SU activities. We obtained the first ever closed-form expression for the effective bandwidth for general DSA networks. We also provided the asymptotic tail distribution analysis for this type of network, where we showed the buffer occupancy is a light-tailed distribution if the busy period for channels are light-tail distributed. This work can be extended to cases where fragmentation and aggregation are not supported as well as different distribution of PU traffic developed recently [32].

TABLE III: Simulation parameters for Fig. 9b

System	I	II	III	IV	V	VI	VII
N_s	7	7	7	7	7	6	7
N_p	3	3	3	3	3	3	2
λ_s	10	10	10	5	10	10	10
μ_s	14	14	14	14	7	14	14
λ_p	6	6	3	6	6	6	6
μ_p	4	2	4	4	4	4	4

References

[1] F. P. Kelly, "Notes on effective bandwidths," in *Stochastic Networks: Theory and Applications*. Oxford University Press, 1996, no. 4.

[2] M. Hassan, M. Krunz, and I. Matta, "Markov-based channel characterization for tractable performance analysis in wireless packet networks," *IEEE Transactions on Wireless Communications*, vol. 3, no. 3, 2004.

[3] K. Park and W. Willinger, *Self-similar network traffic and performance evaluation*. Wiley Online Library, 2000.

[4] A. Laourine, S. Chen, and L. Tong, "Queuing analysis in multichannel cognitive spectrum access: A large deviation approach," in *IEEE Conference on Computer Communications Workshops, INFOCOM*, 2010.

[5] G. Hwang and S. Roy, "Design and analysis of optimal random access policies in cognitive radio networks," *IEEE Transactions on Communications*, vol. 60, no. 1, pp. 121–131, January 2012.

[6] P. Wang and I. Akyildiz, "Can dynamic spectrum access induce heavy tailed delay?" in *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, 2011, pp. 197–207.

[7] S. Wang, J. Zhang, and L. Tong, "A characterization of delay performance of cognitive medium access," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 800–809, 2012.

[8] D. Anick, D. Mitra, and M. Sondhi, "Stochastic theory of a data handling system with multiple sources," in *Bell Sys. Tech. Jour.*, vol. 61, 1982.

[9] D. Mitra, "Stochastic theory of a fluid model of producers and consumers coupled by a buffer," in *Adv. Appl. Prob.*, vol. 20, 1988, pp. 646–676.

[10] Y. Liu and W. Gong, "On fluid queueing systems with strict priority," *IEEE Transactions on Automatic Control*, vol. 48, no. 12, 2003.

[11] B. Tan and S. Gershwin, "Modeling and analysis of markovian continuous flow systems with a finite buffer," *Annals of Operations Research*, vol. 182, no. 1, pp. 5–30, 2011.

[12] F. Roijers, H. V. D. Berg, and M. Mandjes, "Performance analysis of differentiated resource-sharing in a wireless ad-hoc network," *Performance Evaluation*, vol. 67, no. 7, pp. 528–547, 2010.

[13] N. Antunes, A. Pacheco, and R. Rocha, "An integrated traffic model for multimedia wireless networks," *Computer Networks*, vol. 38, 2002.

[14] F. Wan, L. Cai, E. Shihab, and A. Gulliver, "Admission region of triple-play services in wireless home networks," *Computer Communication*, vol. 33, no. 7, pp. 852–859, May 2010.

[15] V. Arunachalam, V. Gupta, and S. Dharmaraja, "A fluid queue modulated by two independent birth-death processes," *Computers & Mathematics with Applications*, vol. 60, no. 8, pp. 2433–2444, 2010.

[16] M. Krunz and J. G. Kim, "Fluid analysis of delay and packet discard performance for qos support in wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 2, Feb 2001.

[17] Q. Wang, D. Wu, and P. Fan, "Effective capacity of a correlated nakagami-m fading channel," *Wireless Communications and Mobile Computing*, vol. 12, no. 14, pp. 1225–1238, 2012.

[18] P. Wang and I. Akyildiz, "Asymptotic queuing analysis for dynamic spectrum access networks in the presence of heavy tails," *IEEE Journal onSelected Areas in Communications*, vol. 31, no. 3, pp. 514–522, 2013.

[19] F. Kocak, H. Celebi, S. Gezici, K. Qaraqe, H. Arslan, and H. Poor, "Time-delay estimation in dispersed spectrum cognitive radio systems," *EURASIP Journal on Advances in Signal Processing*, vol. 23, Feb 2010.

[20] F. Wang, J. Huang, and Y. Zhao, "Delay sensitive communications over cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1402–1411, April 2012.

[21] Q. Wang, D. Wu, and P. Fan, "Effective capacity of a correlated rayleigh fading channel," *Wireless Communications and Mobile Computing*, vol. 11, no. 11, pp. 1485–1494, 2011.

[22] M. Elalem and L. Zhao, "Effective capacity optimization for cognitive radio network based on underlay scheme in gamma fading channels," in *Int. Con. on Comp., Net. and Comm., ICNC*, 2013.

[23] D. Xu, Z. Feng, Y. Wang, and P. Zhang, "Capacity of cognitive radio under delay quality-of-service constraints with outdated channel feedback," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, 2012, pp. 1704–1709.

[24] I. Suliman and J. Lehtomaki, "Queueing analysis of opportunistic access in cognitive radios," in *Second International Workshop on Cognitive Radio and Advanced Spectrum Management, CogART*, 2009.

[25] S. Akin and M. Gursoy, "Effective capacity analysis of cognitive radio channels for quality of service provisioning," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, November 2010.

[26] L. Musavian and S. Aissa, "Effective capacity of delay-constrained cognitive radio in nakagami fading channels," *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, pp. 1054–1062, March 2010.

[27] S. Srinivasa, S. Jafar, and N. Jindal, "On the capacity of the cognitive tracking channel," in *IEEE Int. Sym. on Info. Theory*, 2006.

[28] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Tran. on Info. Theory*, vol. 52, no. 5, 2006.
[29] Y. Wu, G. Min, and A. Al-Dubai, "A new analytical model for multi-hop cognitive radio networks," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 5, pp. 1643–1648, May 2012.
[30] Z. Liang, S. Feng, D. Zhao, and X. Shen, "Delay performance analysis for supporting real-time traffic in a cognitive radio sensor network," *IEEE Transactions on Wireless Communications*, vol. 10, no. 1, 2011.
[31] W. Gabran, C.-H. Liu, P. Pawelczak, and D. Cabric, "Primary user traffic estimation for dynamic spectrum access," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 544–558, March 2013.
[32] M. Lopez-Benitez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Tran. on Vehicular Tech.*, vol. 62, no. 5, 2013.
[33] The President Council of Advisors on Science and Technology (PCAST). (2012, July) Report to the president: Realizing the full potential of government-held spectrum to spur economic growth. [Online]. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_spec trum_report_final_july_20_2012.pdf
[34] S. Asmussen, *Applied Probability and Queues*. New York: Springer-Verlag, 2003.
[35] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, 2003.
[36] C. Larsson, *Design of Modern Communication Networks: Methods and Applications*. Academic Press, 2014.
[37] P. W. Glynn and W. Whitt, "Logarithmic asymptotics for steady-state tail probabilities in a single-server queue," *Journal of Applied Probability*, pp. 131–155, 1994.
[38] S. Akin and M. Gursoy, "Effective capacity analysis of cognitive radio channels for quality of service provisioning," in *IEEE Global Telecommunications Conference, GLOBECOM*, 2009, pp. 1–6.
[39] S. Shakkottai, "Effective capacity and QoS for wireless scheduling," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, 2008.
[40] A. Anwar, K. Seddik, T. ElBatt, and A. Zahran, "Effective capacity of delay constrained cognitive radio links exploiting primary feedback," in *Inter. Sym. and Workshops on Mod. and Opt. in Mob., Ad Hoc and Wireless Networks (WiOpt)*, May 2013.
[41] S.-Y. Lien, K.-C. Chen, Y.-C. Liang, and Y. Lin, "Cognitive radio resource management for future cellular networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 70–79, February 2014.
[42] K. Johansson, J. Bergman, D. Gerstenberger, M. Blomgren, and A. Wallén, "Multi-carrier HSPA evolution," in *IEEE Vehicular Technology Society Conference*, 2009, pp. 1–5.
[43] H. Shajaiah, A. Khawar, A. Abdel-Hadi, and T. Clancy, "Resource allocation with carrier aggregation in LTE advanced cellular system sharing spectrum with S-band radar," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, 2014, pp. 34–37.
[44] K. Werner, H. Asplund, B. Halvarsson, A. Kathrein, N. Jalden, and D. Figueiredo, "LTE-A field measurements: 8x8 MIMO and carrier aggregation," in *IEEE Vehicular Tech. Con. (VTC Spring)*, 2013.
[45] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 913–931, 1994.

S. Ebrahim Safavi received the B.S. degree in Electrical Engineering from Isfahan University of Technology, Iran and his M.S. degree from Sharif University of Technology, Iran. Since Fall 2010, he has been a Ph.D. student in the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, New Jersey. His research on wireless communications mainly focuses on cognitive radio networks and stochastic network characterization.

Prof. K.P. (Suba) Subbalakshmi is a Professor at Stevens Institute of Technology. Her research interests are in the areas of cognitive radio networks, wireless network security, media forensics as well as social networks. She is a Communications Subject Matter Expert in the National Spectrum Consortium. She is the Founding Chair of the Security Special Interest Group of the IEEE Technical Committee on Cognitive Networks. She is a recipient of the Innovator award instituted by the NJ Inventor's Hall of Fame.