Modeling and Analysis of Opportunistic Spectrum Sharing with Unreliable Spectrum Sensing

Shensheng Tang, *Senior Member, IEEE* and Brian L. Mark, *Member, IEEE*

Abstract— We analyze the performance of a wireless system consisting of a set of secondary users opportunistically sharing bandwidth with a set of primary users over a coverage area. The secondary users employ spectrum sensing to detect channels that are unused by the primary users and hence make use of the idle channels. If an active secondary user detects the presence of a primary user on a given channel, it releases the channel and switches to another idle channel, if one is available. In the event that no channel is available, the call waits in a buffer until either a channel becomes available or a maximum waiting time is reached. Spectrum sensing errors on the part of a secondary user cause false alarm and misdetection events, which can potentially degrade the quality-of-service experienced by primary users. We derive system performance metrics of interest such as blocking probabilities. Our results suggest that opportunistic spectrum sharing can significantly improve spectrum efficiency and system capacity, even under unreliable spectrum detection. The proposed model and analysis method can be used to evaluate the performance of future opportunistic spectrum sharing systems.

Index Terms— Opportunistic spectrum sharing, dynamic spectrum access, spectrum sensing, performance modeling, Markov process, false alarm, misdetection.

I. INTRODUCTION

Studies of wireless spectrum usage [2], [3] have shown that large portions of the allocated spectrum are highly underutilized. Frequency agile radios (FARs) are cognitive radios that are capable of detecting idle frequency channels and making use of them opportunistically without causing harmful interference to the primary users [4]. In a scenario of opportunistic spectrum sharing (OSS), the FARs are called *secondary* users and the owners of the allocated spectrum are the *primary* users. By allowing secondary users to reclaim idle channels, much higher spectrum efficiency can be achieved [1]. More generally, cognitive radios [5] may be capable of opportunistic spectrum access over frequency channels, time slots, or spreading codes.

Secondary users opportunistically make use of channels that are not occupied by primary users. A secondary user senses when a channel is idle and then makes use of such a channel. Conversely, an active secondary user also detects

when a primary user accesses a channel that it is using and then either moves to an idle channel, if one is available, or moves to a waiting pool. In the latter case, the secondary user's call waits in a buffer until either a new channel becomes available or until a timeout occurs after a predefined maximum waiting time. The reliable detection of primary users is a major challenge for the implementation of an OSS system. The spectrum usage of the secondary users is contingent on the requirement that the disruption to the primary users must be limited to a certain threshold.

In this paper, we model an opportunistic spectrum sharing system and evaluate its performance in terms of various performance metrics, including blocking probability, reconnection probability, channel utilization, total carried traffic, mean waiting time in the buffer, and mean peak period time, and the probability of collision for arriving primary calls. We consider a wireless network, which provides a group of channels to a set of primary users. The wireless network may be infrastructured or infrastructureless. Here, we use the term *channel* in a broad sense. A channel could be a frequency channel in an FDMA system [6], a time-slot in a TDMA system [7], a spreading code in a CDMA system [8], or a tone in an OFDM system [9]. Our proposed system model can be applied to all of these scenarios.

A number of papers related to opportunistic spectrum sharing (OSS) have appeared in the literature. In [10], a collaborative spectrum sensing mechanism is proposed and studied as a means to combat the shadowing or fading effects that a user experiences. The results suggest that collaboration may improve sensing performance significantly. In [11], it was shown that by taking advantage of the local oscillator leakage power that all RF receivers emit, sensor nodes could detect the exact channel that a primary user was tuned to and transmit this information to a set of cognitive radios through a control channel. This approach could potentially allow cognitive radios to operate in dense urban environments without interfering with primary receivers.

In [4], a framework was developed for modeling the interference caused by FARs employing spectrum access mechanisms based on the simple Listen-Before-Talk (LBT) scheme. Two variations of LBT were considered: individual LBT and collaborative LBT. In [12], a measurement-based model is proposed to represent the busy and idle periods of a WLAN statistically. Two different sensing strategies, energy-based detection and feature-based detection, are explored to identify spectrum opportunities. In [13], a multi-channel MAC protocol is proposed to enable the interoperation of the primary-secondary overlay network. The protocol detects the

frequency bands utilized by the primary system, and creates and maintains a record of the unutilized resources for the secondary users.

In [14], a sensing-based approach was studied for channel selection in spectrum-agile communication systems. The proposed approach includes two steps. The first step determines whether or not a given channel is idle. The second step determines whether or not an idle channel is a good opportunity. In [15], a multichannel OFDMA technique is proposed for networks allowing opportunistic spectrum access (OSA). The OSA nodes compete amongst themselves and with the primary users by using fast retrials. Mechanisms are proposed to minimize the probability of collision and interference caused to the primary users while maximizing the throughput of the OSA network.

Our focus in this paper is on the modeling and performance analysis of an OSS system at the call level under the assumption of imperfect OSA, i.e., the spectrum sensing performed by a secondary user is imperfect and subject to false alarm and misdetection events, which negatively impact the performance of primary users as well as other secondary users. The remainder of the paper is organized as follows. Section II describes the system model and assumptions in further detail. Section III develops a Markovian model of the system dynamics, while Section IV derives the performance metrics of interest. Section V presents numerical results, illustrating the performance of the OSS system with respect to the different metrics, over a range of parameter settings. Finally, the paper is concluded in Section VI.

II. MODEL AND ASSUMPTIONS

Consider a cellular network operating over a given service area. The network owns the license for spectrum usage in the service area and hence is referred to as the *primary system*. The users of this network are the primary users. Calls generated by primary users constitute the primary traffic (PT) stream. Each cell in the primary system contains a base station or an access point (AP), which provides wireline connectivity to the backbone network. A node within a cell sets up a call through the AP to communicate with other nodes. The AP is in charge of allocating resources, providing access to the fixed network, and other administrative tasks. In addition, the AP maintains information such as channel status, QoS parameters, user profiles, etc.

Next, we introduce another wireless network in the same service area, which opportunistically shares the precious spectrum resource with the existing network. This network is referred to as the *secondary system* and the associated users are called secondary users. Calls generated by secondary users constitute the secondary traffic (ST) stream. The system consisting of the primary and secondary systems is called an opportunistic spectrum sharing (OSS) system. In the OSS system, spectrum availability for the secondary users is subject to the spectrum occupancy of the primary users. Secondary users have the capability to sense channel usage and switch among different channels to make use of idle channels and to avoid interfering with primary users. Such functionality could be realized by cognitive radios.

Secondary users determine channel status by sensing spectrum usage. If an arriving secondary call finds an idle channel, it can make use of the channel. If all channels are busy, the secondary call is blocked and considered lost from the system. On the other hand, if an active secondary user detects the presence of a primary call on a given channel, it either switches to an available channel or moves to a waiting pool. In the ideal case, the quality-of-service experienced by primary users is not affected by the secondary users, since in effect, primary calls are given preemptive priority over secondary calls.

In a practical system, however, a primary call that is actively using a given channel may experience disruption if an arriving ST call searching for a free channel incorrectly determines that the given channel is idle. A second class of disruption events to a primary call may occur when an active secondary user on a given channel fails to detect the presence of an arriving primary user on the given channel. This may incur service degradation on the primary user. We refer to such detection errors as class-A and class-B *misdetection* events, respectively. In this paper, we shall only consider class-A misdetection events; i.e., we shall assume that an active secondary user can perfectly detect the arrival of a new primary call on its current channel such that class-B misdetection never occurs. The model presented in this paper can easily be extended to handle the class-B misdetection events, but we shall omit the details here. In the remainder of the paper, the term "misdetection event" shall refer only to class-A misdetection events.

Misdetection events can negatively impact the performance of the primary system. On the other hand, a secondary user may incorrectly determine that a channel is busy when in fact the channel is idle. We refer to this type of error as a *false alarm* event. A false alarm event does not incur performance degradation on the primary system, but lowers the potential spectrum utilization of the OSS system. We denote the probabilities of misdetection and false alarm by p_m and p_f, respectively.

We consider two possible scenarios that may occur as a consequence of a misdetection event, which we call type-I misdetection and type-II misdetection. When a secondary user incorrectly determines a given channel to be idle, while in fact a primary user is using the channel, the secondary user will transmit on the channel and as a result, cause interference to the primary user. In type-I misdetection, both the primary and secondary users drop the channel due to the collision event (large "noise" incurred). In type-II misdetection, the primary user drops the channel, but the secondary user remains on the channel because it had initially determined the channel to be idle. Both misdetection events lower the performance of the primary system, which is not consistent with the principle of designing an OSS system. Hence, a critical issue in the design of an OSS system is to keep the probability of a spectrum sensing error small. That is, the false alarm probability p_f and the misdetection probability p_m should be kept as small as possible so that the negative performance impact to the potential spectrum utilization and to the primary users is minimized. In the analysis of Section III, we consider both type-I and type-II misdetections.
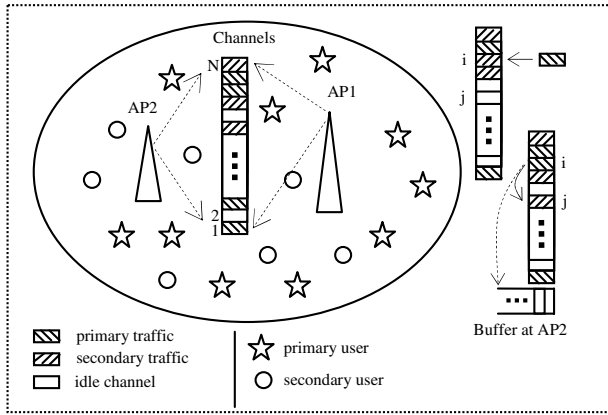
Fig. 1.   Opportunistic spectrum sharing (OSS) system model.

Fig. 2.   State diagram at (i, j) under type-I misdetection with pre-full channel occupancy $(i + j < N)$.

In the OSS system model depicted in Fig. 1, the primary and secondary systems are both represented as infrastructured wireless networks. However, the performance model discussed in this paper applies to more general scenarios. For example, one or both of the primary and secondary systems may be infrastureless ad hoc networks. Suppose there are a total of N channels managed by the primary system with access point AP1 in a given cell. The PT calls operate as if there are no ST calls in the system. When a PT call arrives to the system, it occupies a free channel if one is available; otherwise, it will be blocked. Secondary users detect the presence or absence of signals from primary users and maintain records of the channel occupancy status. The detection mechanism may involve collaboration with other secondary users and/or an exchange with an associated access point called AP2, as shown in Fig. 1.

Secondary users opportunistically access the channels that are in idle status. When an ST node detects or is informed (by AP2 or other ST nodes) of an arrival of PT call in its current channel, it immediately leaves the channel and switches to an idle channel, if one is available, to continue the call (see Fig. 1). If at that time all the channels are occupied, the ST call is placed in a buffer located at AP2.[1] For example, in Fig. 1, when an ST call detects the arrival of a PT call at channel i, it immediately leaves that channel and changes to channel j. If all of the N channels are occupied at that time, the ST call will be queued. Queued ST calls are served in first-come first-served (FCFS) order. The head-of-line ST call is reconnected to the system when a channel becomes available before a predefined maximum waiting time expires. We set the maximum waiting time of an ST call equal to its residence time in the considered service area. Thus, an ST call is lost only when it moves out of the service area, statistically speaking.

III. PERFORMANCE ANALYSIS

In this section, we analyze the OSS system performance in a given service area consisting of the primary and secondary systems sharing the same spectrum. The spectrum is divided
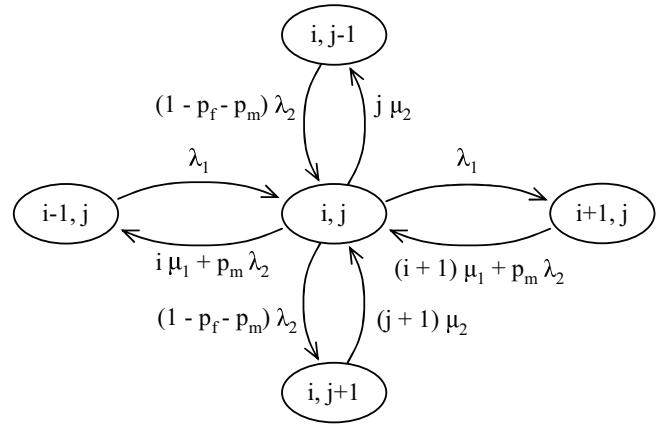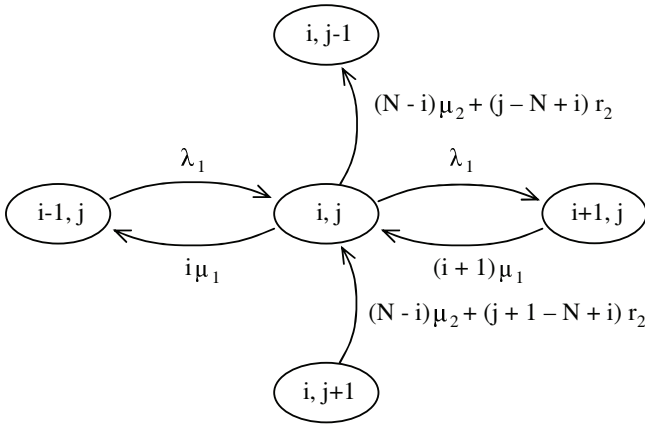
into N channels serving the two types of traffic: primary traffic (PT) and secondary traffic (ST). Arrivals of the PT and ST calls are assumed to form independent Poisson processes with rates λ_1 and λ_2, respectively. The call holding times of the PT and ST calls are assumed to be exponentially distributed with means h_1^{-1} and h_2^{-1}, respectively. The residence times for the PT and ST in the service area are also assumed to be exponentially distributed with means r_1^{-1} and r_2^{-1}, respectively. The channel holding time is the minimum of the call holding and residence times. Hence, the channel holding times for the PT and ST calls are exponentially distributed with means $\mu_1^{-1} = (h_1 + r_1)^{-1}$ and $\mu_2^{-1} = (h_2 + r_2)^{-1}$, respectively. These assumptions have been found to be reasonable as long as the number of users is much more than that of the channels in a service area, and have been widely used in the literature [16]–[20]. We further assume that both types of traffic occupy one channel per call for simplicity. However, the analytical approach used here can be extended to handle variable bandwidth requests (cf. [21]).

Let $X_1(t)$ denote the number of PT calls in the OSS system at time t. Similarly, let $X_2(t)$ be the number of ST calls in the system at time t, including the ST calls being served and those waiting in the buffer at AP2. The process $(X_1(t), X_2(t))$ is a two-dimensional Markov process with state space $\mathcal{S} = \{(n_1, n_2) | 0 \leq n_1, n_2 \leq N\}$. We classify the channel occupancy of the system in state (n_1, n_2) as *pre-full* if $n_1 + n_2 < N$, *just-full* if $n_1 + n_2 = N$, and *post-full* if $n_1 + n_2 > N$.

Due to unreliable spectrum detection, false alarm and misdetection events are considered in our analysis. As mentioned earlier, type-I misdetection and type-II misdetection can cause different channel occupancy behaviors and lead to different system state transitions. In the following, we analyze the OSS system first with false alarm and type-I misdetection events, and then with false alarm and type-II misdetection events.

A. Analysis Under Type-I Misdetection

The state transition diagram of the OSS system under type-I misdetection is shown in Fig. 2 with pre-full channel occupancy and in Fig. 3 with post-full channel occupancy. In Fig. 2, due to the impact of false alarm and misdetection

---

[1]The secondary users may also operate in ad hoc mode, in which case AP2 is not needed, but a *virtual queue* of ST calls could still be maintained.

Fig. 3. State diagram at (i,j) with post-full channel occupancy $(i+j > N)$.

Fig. 4. State diagram at (i,j) under type-II misdetection with pre-full channel occupancy $(i+j < N)$.

events, the system moves from state (i,j) to $(i,j+1)$ with transition rate $(1-p_f-p_m)\lambda_2$ and moves to $(i-1,j)$ with rate $i\mu_1 + p_m\lambda_2$, where $i\mu_1$ is the transition rate due to service completion and $p_m\lambda_2$ is the additional transition rate due to type-I misdetection. In the boundary case of just-full occupancy with $i+j = N$, the state transition diagram is the same as Fig. 2 except with the following three modifications: 1) the transition from (i,j) to $(i,j+1)$ is removed; 2) the term $p_m\lambda_2$ is removed in the transition rate from $(i+1,j)$ to (i,j); 3) the transition rate from $(i,j+1)$ to (i,j) is changed from $(j+1)\mu_2$ to $(N-i)\mu_2+r_2$. In Fig. 3, since the system is in the post-full channel occupancy status, there is no possibility for a new ST call to enter the system. Of course, it is possible for an ongoing ST call to leave the system due to service completion and for a waiting ST call in the buffer to reconnect due to a completion of a PT or ST call.

The state (i,j) in Fig. 3 corresponds to the case in which there are i PT calls, $N - i$ ongoing ST calls in the system and $j - (N - i)$ waiting ST calls in the buffer. The transition rate from state (i,j) to $(i,j-1)$ is given by $(N-i)\mu_2+(j-N+i)r_2$. Note that the false alarm and misdetection events only happen in the pre-full and just-full channel occupancy states. In the post-full channel occupancy situation there is at least one ST call waiting in the buffer, which means all the channels are occupied and no misdetection or false alarm events can occur. Note also that when $i = 0$, we have $p_m = 0$. The transition rate from state (n_1, n_2) to (n'_1, n'_2), denoted by $T^{n'_1,n'_2}_{n_1,n_2}$, is given by

$$T^{n_1+1,n_2}_{n_1,n_2} = \lambda_1 1_{\{0 \le n_1 < N, \ 0 \le n_2 \le N\}},$$
$$T^{n_1-1,n_2}_{n_1,n_2} = n_1\mu_1 1_{\{1 \le n_1 \le N\}} + p_m\lambda_2 1_{\{0 \le n_2 \le N-n_1\}},$$
$$T^{n_1,n_2+1}_{n_1,n_2} = [1-p_f-\bar{\delta}(n_1)p_m]\lambda_2 1_{\{0 \le n_1 \le N-1, \ 0 \le n_2 < N-n_1\}},$$
$$T^{n_1,n_2-1}_{n_1,n_2} = n_2\mu_2 1_{\{0 \le n_1 \le N-1, \ 1 \le n_2 \le N-n_1\}} +$$
$$[(n_2-N+n_1)r_2 + (N-n_1)\mu_2]1_{\{1 \le n_1 \le N, \ N-n_1 < n_2 \le N\}},$$

where[2] $\bar{\delta}(i) = 0$ if $i = 0$ and $\bar{\delta}(i) = 1$ if $i \ne 0$; and $1_{\{x\}}$ is an indicator function defined as 1 if x is true and 0 otherwise.

Let $\pi(n_1, n_2)$ denote the steady-state probability that the OSS system is in state (n_1, n_2). The steady-state system

---

[2] $\bar{\delta}(i) = 1 - \delta(i)$ where $\delta(i)$ is the Kronecker delta function defined by $\delta(i) = 1$ if $i = 0$ and $\delta(i) = 0$ if $i \ne 0$.
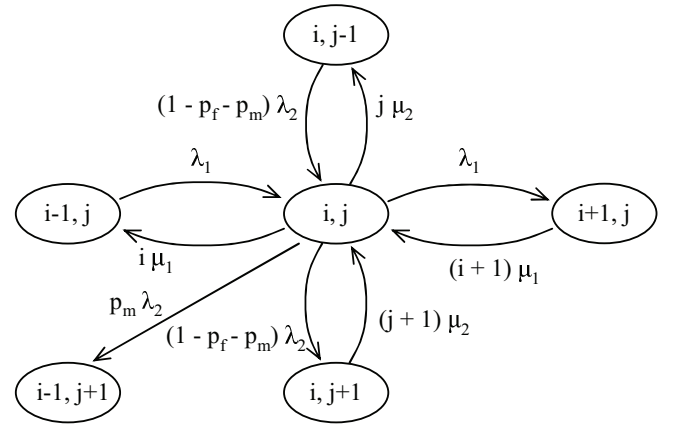
---

probability vector, with states ordered lexicographically, can be represented as $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_N)$, where

$$\boldsymbol{\pi}_n = (\pi(n,0), \pi(n,1), \cdots, \pi(n,N)), \quad 0 \le n \le N.$$

The vector $\boldsymbol{\pi}$ is the solution of the following equations:

$$\boldsymbol{\pi}Q = \mathbf{0} \text{ and } \boldsymbol{\pi}\mathbf{e} = \mathbf{1},$$

where $\mathbf{e}$ and $\mathbf{0}$ are column vectors of all ones and zeros, respectively. The infinitesimal generator, Q, of the two-dimensional Markov process, is given by

$$Q = \begin{bmatrix} E_0 & B_0 & 0 & \cdots & 0 & 0 & 0 \\ D_1 & E_1 & B_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & D_{N-1} & E_{N-1} & B_{N-1} \\ 0 & 0 & 0 & \cdots & 0 & D_N & E_N \end{bmatrix}$$

where each submatrix is of size $N + 1$ by $N + 1$ and defined by

$$B_i = \lambda_1 I_{N+1}, \quad 0 \le i < N;$$
$$E_i = A_i - \bar{\delta}(i)D_i - \bar{\delta}(N-i)B_i, \quad 0 \le i \le N;$$
$$D_i = i\mu_1 I_{N+1} + p_m\lambda_2 I^{(i)}_{N+1}, \quad 1 \le i \le N.$$

where I_n denotes an n-by-n identity matrix,

$$I^{(i)}_n \triangleq \begin{bmatrix} I_{n-i} & 0_{(n-i)\times i} \\ 0_{i\times(n-i)} & 0_{i\times i} \end{bmatrix},$$

and $I^{(0)}_n \triangleq I_n$. The matrix A_i has the same dimensions as E_i. The (j,k) element of the matrix A_i is given by $A_i(j,k) =$

$$\begin{cases} (1-p_f-\bar{\delta}(i)p_m)\lambda_2, & \begin{aligned}&0 \le i \le N-1,\\&0 \le j < N-i, k=j+1,\end{aligned} \\ j\mu_2, & \begin{aligned}&0 \le i \le N-1,\\&1 \le j \le N-i, k=j-1,\end{aligned} \\ (N-i)\mu_2 + (j-N+i)r_2, & \begin{aligned}&1 \le i \le N,\\&N-i < j \le N, k=j-1,\end{aligned} \\ -[A_i(j,j-1) + A_i(j,j+1)], & \begin{aligned}&0 \le i \le N,\\&0 \le j \le N, k=j,\end{aligned} \\ 0, & \text{otherwise.} \end{cases}$$

(1)

Applying the matrix-analytic method developed in [20], the steady-state probabilities can be determined as follows:

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_{n-1}B_{n-1}(-C_n)^{-1} = \boldsymbol{\pi}_0 \prod_{i=1}^{n}[B_{i-1}(-C_i)^{-1}], \quad (2)$$

where $1 \leq n \leq N$ and $\boldsymbol{\pi}_0$ satisfies $\boldsymbol{\pi}_0 C_0 = \mathbf{0}$ and

$$\boldsymbol{\pi}_0 \left[I + \sum_{n=1}^{N} \prod_{i=1}^{n} [B_{i-1} (-C_i)^{-1}] \right] \mathbf{e} = 1.$$

The C_i are computed recursively by setting $C_N = E_N$ and

$$C_i = E_i + B_i (-C_{i+1})^{-1} D_{i+1}, \quad 0 \leq i \leq N-1. \quad (3)$$

B. Analysis Under Type-II Misdetection

The state transition rate diagram with pre-full channel occupancy under type-II misdetection is shown in Fig. 4. The state diagram for the just-full case is the same as Fig. 4, except that the transition from (i,j) to $(i,j+1)$ is removed. The corresponding diagram with post-full channel occupancy under type-II misdetection is the same as that under type-I misdetection in Fig. 3. The transition rate from state (n_1, n_2) to (n_1', n_2'), denoted by $T_{n_1,n_2}^{n_1',n_2'}$, is given by

$$T_{n_1,n_2}^{n_1+1,n_2} = \lambda_1 1_{\{0 \leq n_1 < N, \ 0 \leq n_2 \leq N\}},$$
$$T_{n_1,n_2}^{n_1-1,n_2} = n_1 \mu_1 1_{\{1 \leq n_1 \leq N, \ 0 \leq n_2 \leq N\}},$$
$$T_{n_1,n_2}^{n_1-1,n_2+1} = p_m \lambda_2 1_{\{1 \leq n_1 \leq N, \ 0 \leq n_2 \leq N-n_1\}},$$
$$T_{n_1,n_2}^{n_1,n_2+1} = (1-p_f - \bar{\delta}(n_1)) p_m \lambda_2 1_{\{0 \leq n_1 \leq N-1, \ 0 \leq n_2 < N-n_1\}},$$
$$T_{n_1,n_2}^{n_1,n_2-1} = n_2 \mu_2 1_{\{0 \leq n_1 \leq N-1, \ 1 \leq n_2 \leq N-n_1\}} +$$
$$[(n_2 - N + n_1)r_2 + (N-n_1)\mu_2]1_{\{1 \leq n_1 \leq N, \ N-n_1 < n_2 \leq N\}}.$$

The steady-state probabilities for the OSS system under type-II misdetection are obtained by following the same approach as that used in the type-I misdetection analysis, except that the matrix D_i is different. The (j,k) element of D_i is given as follows:

$$D_i(j,k) = p_m \lambda_2 1_{\{1 \leq i \leq N, \ 0 \leq j \leq N-i, \ k=j+1\}}$$
$$+ j\mu_1 1_{\{1 \leq i \leq N, \ 0 \leq j \leq N, \ k=j\}}.$$

C. Special Case: Single Primary System

If there are no secondary users in the system, that is, $n_2 = 0$, $\lambda_2 = 0$, and $\mu_2 = 0$, the OSS system reduces to a single primary system. In this case, the performance model simplifies as follows:

$$B_i = \lambda_1, \ 0 \leq i < N; \quad D_i = i\mu_1, \ 1 \leq i \leq N;$$
$$E_i = -i\mu_1 - \bar{\delta}(N-i)\lambda_1, \ 0 \leq i \leq N; \quad A_i = 0, \ 0 \leq i \leq N;$$
$$C_i = -i\mu_1, \ 0 \leq i \leq N.$$

Substituting the above equations into (2), we obtain

$$\pi_0 = \left[\sum_{i=0}^{N} \frac{1}{i!} \left(\frac{\lambda_1}{\mu_1} \right)^i \right]^{-1}, \quad \pi_n = \frac{1}{n!} \left(\frac{\lambda_1}{\mu_1} \right)^n \pi_0, \quad 1 \leq n \leq N,$$
$$(4)$$

which is the well-known M/M/N/N or Erlang loss model [22].

IV. PERFORMANCE METRICS

Next, we obtain various performance measures of interest.

A. Blocking probabilities

- **Blocking probability of the primary traffic**
  The PT call blocking probability, denoted by P_1, is defined as the probability that upon an arrival of a PT call in a service area all the channels are occupied by PT calls and the arrival request has to be blocked. Thus, we have

$$P_1 = \sum_{n_2=0}^{N} \pi(N, n_2) = \boldsymbol{\pi}_0 \prod_{i=1}^{N} [B_{i-1} (-C_i)^{-1}] \mathbf{e}. \quad (5)$$

- **Blocking probability of the secondary traffic**
  The ST call blocking probability, denoted by P_2, is defined as the probability that all the channels in a service area are occupied by either PT calls and/or ST calls and no channel is available for a new ST call request. Thus, we have

$$P_2 = \sum_{n_1=0}^{N} \sum_{n_2=N-n_1}^{N} \pi(n_1, n_2). \quad (6)$$

B. Mean reconnection probability

As mentioned earlier, an ST call that waits in the buffer due to unavailability of a channel could reconnect back to the system if an channel becomes available before the maximum waiting time expires. The mean reconnection probability of an ST call, denoted by γ, is defined as the probability that this ST call reconnects back to the system before its maximum waiting time expires. We obtain

$$\gamma = \frac{\sum_{n_1=1}^{N} \sum_{j=0}^{n_1-1} \pi(n_1, N- n_1 + j + 1)\beta(j)}{\sum_{n_1=1}^{N} \sum_{j=0}^{n_1-1} \pi(n_1, N - n_1 + j + 1)}, \quad (7)$$

where $\beta(j)$ denotes the probability that an ST call arriving at the buffer eventually reconnects back to the system before its maximum queueing time expires, given that the ST call comes to find that there are j ST calls in the buffer ($0 \leq j \leq N-1$).

Recall that the ST calls in the buffer are reconnected to the system when channels become available in first come first served (FCFS) order. If an ST call detects an arrival of a PT call at its channel and there are $N+j$ ($0 \leq j \leq N-1$) calls in the system (i.e., all N channels are serving the PT/ST calls and j ST calls are waiting in the buffer), it releases its channel for the PT call and enters the buffer, which leads to a new system state with all N channels being used and $j+1$ ST calls in the buffer. This ST call reconnects to the system only if $j+1$ calls leave the service area, either releasing a channel or a position in the buffer, before its maximum queueing time expires.

Let τ denote the maximum queueing time that can be tolerated by an ST call in the buffer. The maximum waiting time of an ST call in the buffer is assumed to be statistically the same as the residence time of the ST call. Hence, τ is exponentially distributed with mean r_2^{-1}. To capture the queueing behavior of ST calls, we introduce a process $J(t)$ to represent the number of queued ST calls at time t. Then the OSS system when all channels are occupied can be represented by a 3-dimensional Markov process $(Z_1(t), Z_2(t), J(t))$ with state space

$$S^* = \{(n_1, n_2, j) | n_1 + n_2 = N, 0 \leq j \leq N\}.$$

Let φ_j $(0 \leq j \leq N-1)$ denote the time interval, in steady-state, between a transition to a state $(n_1, N-n_1, j+1) \in \mathcal{S}^*$ until a transition to a new state $(n_1', N-n_1', j)$, such that either a PT/ST call occupying a channel leaves the system, or a queued ST call leaves the system. If a PT call leaves, then $n_1' = n_1 - 1$; otherwise, $n_1' = n_1$. When a PT or ST call leaves the system, the head-of-line ST call in the queue reconnects to the system and the remaining queued ST calls advance by one position in the buffer. Similarly, the dropping of a queued ST call leads to the advancement, by one position, of each of the remaining queued ST calls that were behind it. Hence, φ_j is exponentially distributed with parameter $g_j = n_1 \mu_1 + (N - n_1)\mu_2 + j r_2$, $0 \leq j \leq N-1$.

Let $f_j(\cdot)$ denote the probability density function of φ_j and let $f_j^*(s)$ denote the Laplace transform of $f_j(\cdot)$. By the independence assumption of the random variables φ_j, we can determine $\beta(j)$ as

$$\beta(j) = \Pr(\tau > \varphi_0 + \varphi_1 + \cdots + \varphi_j) = \prod_{i=0}^{j} f_i^*(r_2)$$
$$= \frac{n_1 \mu_1 + (N - n_1)\mu_2}{n_1 \mu_1 + (N - n_1)\mu_2 + (j+1)r_2}, \qquad (8)$$

where the last equation follows from the fact that

$$f_i^*(r_2) = \frac{n_1 \mu_1 + (N - n_1)\mu_2 + i r_2}{n_1 \mu_1 + (N - n_1)\mu_2 + (i+1)r_2}.$$

The reconnection probability can then be calculated by substituting (8) into (7).

C. Total channel utilization and carried traffic

The total channel utilization η is defined as the ratio of the mean number of occupied channels to the total number of channels. We find that

$$\eta = \frac{1}{N}\left\{ \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} (n_1 + n_2)\pi(n_1, n_2) + \right.$$
$$\left. \sum_{n_1=1}^{N} \sum_{n_2=N-n_1+1}^{N} N\pi(n_1, n_2) \right\}. \qquad (9)$$

The total carried traffic (TCT) by the OSS system is defined as the total traffic (both PT and ST) that the OSS system supports in the given service area. We find that

$$\text{TCT} = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} (n_1 + n_2)\pi(n_1, n_2) + $$
$$\sum_{n_1=1}^{N} \sum_{n_2=N-n_1+1}^{N} N\pi(n_1, n_2). \qquad (10)$$

D. Mean waiting time of the ST calls in the buffer

When an ST call in communication detects the arrival of a PT call on its current channel, it has to release the channel. If it cannot find an idle channel to switch to, it is queued in the buffer. The queued ST calls in the buffer reconnect back to the system in first come first served (FCFS) order when channels become available. In steady state, the mean number

of queued ST calls in the buffer $\overline{L_b}$ and the mean arrival rate to the buffer $\overline{\lambda_b}$ can be calculated as follows:

$$\overline{L_b} = \sum_{n_1=1}^{N} \sum_{n_2=N-n_1+1}^{N} (n_2 - N + n_1)\pi(n_1, n_2) \qquad (11)$$

$$\overline{\lambda_b} = \lambda_1 \sum_{n_1=0}^{N-1} \sum_{n_2=N-n_1}^{N} \pi(n_1, n_2). \qquad (12)$$

Using Little's law, the mean waiting time of the ST calls in the buffer is obtained as

$$\overline{W_b} = \frac{\sum_{n_1=1}^{N} \sum_{n_2=N-n_1+1}^{N} (n_2 - N + n_1)\pi(n_1, n_2)}{\lambda_1 \sum_{n_1=0}^{N-1} \sum_{n_2=N-n_1}^{N} \pi(n_1, n_2)}. \qquad (13)$$

E. Peak period time of PT calls

We introduce a new performance metric, peak period time of PT calls, denoted by $B_P(M)$ or simply B_P, defined as the time interval from an epoch when M $(1 \leq M \leq N)$ channels are occupied by PT calls to the next epoch when $M-1$ channels are occupied by PT calls[3]. The peak period time of PT calls is a useful performance metric from the point of view of both the primary and secondary system. The primary system service provider may use this metric to evaluate the cost-effectiveness of the system and to optimize the network design. The secondary system may use this metric to assess the peak traffic distribution of primary users, and to make more efficient use of the shared spectrum resources given knowledge of the related parameters, i.e., arrival rates and service times.

Let $Y_1(t)$ be the number of PT calls being served and $Y_2(t)$ be the number of ST calls being served and in the buffer at time t. Then B_P can be determined by considering a two-dimensional absorbing Markov process $(Y_1(t), Y_2(t))$ defined on a state space similar to that of the process $(X_1(t), X_2(t))$, but in which the states $\{(x_1, x_2)|0 \leq x_1 \leq M-1, 0 \leq x_2 \leq N\}$ are collapsed into a single absorbing state, which we will denote by $(0,0)$. The infinitesimal generator matrix Q_P of the process $(Y_1(t), Y_2(t))$ is given by (cf. [20])

$$Q_P = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ D_M & E_M & B_M & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & D_{N-1} & E_{N-1} & B_{N-1} \\ 0 & 0 & 0 & \cdots & 0 & D_N & E_N \end{bmatrix},$$

where each of the submatrices is defined as in the matrix Q given earlier for the process $(X_1(t), X_2(t))$. The peak period time of PT calls in the service area is just the first absorbing time of the Markov process $(Y_1(t), Y_2(t))$ with initial state probability vector $(0, \boldsymbol{\theta})$, where

$$\boldsymbol{\theta} = \left(\frac{\boldsymbol{\pi}_M}{\boldsymbol{\pi}_M \mathbf{e}}, \mathbf{0}, \cdots, \mathbf{0} \right).$$

If we denote by

$$T_P = \begin{bmatrix} E_M & B_M & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & D_{N-1} & E_{N-1} & B_{N-1} \\ 0 & 0 & \cdots & 0 & D_N & E_N \end{bmatrix},$$

---

[3]If $M = 1$, the peak period time corresponds to the busy period (cf. [20]).

from [23, Lemma 2.2.2] and [20], we can determine the peak period time distribution of B_P as

$$P(B_P \leq x) = 1 - \boldsymbol{\theta}\exp(T_P x)\mathbf{e}, \quad x \geq 0. \tag{14}$$

From equation (14), the non-central moments, $\overline{B_P^k}$, can be obtained as

$$\overline{B_P^k} = (-1)^k k!(\boldsymbol{\theta}T_P^{-k}\mathbf{e}), \quad k \geq 0, \tag{15}$$

when $k = 1$, we obtain the mean peak period time of PT calls, $\overline{B_P}$.

F. Probability of Collision for Arriving Primary Calls

When an incoming PT call is assigned a channel that a secondary user is using currently, the secondary user will detect the interference induced by the primary user, and then in a very short time interval, decide whether or not a PT call has arrived and execute the associated operations. Clearly, during this short time interval, there is a collision between the calls of primary and secondary users. Compared with a single primary system, this additional collision in the short time interval is caused completely by the introduction of the secondary system and can be considered as the probability of collision for arriving primary calls. We wish to study this performance metric quantitatively.

Our analysis considers two scenarios based on the different channel assignment strategies at AP1: uniform channel assignment and quality-based channel assignment.

1) Uniform channel assignment: In uniform channel assignment, AP1 randomly assigns a channel to an incoming PT call from the set of idle channels, with equal probability. In state (n_1, n_2), there are already n_1 PT calls in the system. Therefore, when a PT call arrives, AP1 will randomly assign a channel from the remaining $N - n_1$ channels. Clearly, the probability that a channel occupied by an ST call is selected is $n_2/(N - n_1)$. Then, the collision probability, denoted by P_C, can be found as

$$P_C = \sum_{n_2=1}^{N} \sum_{n_1=0}^{N-n_2} \frac{n_2}{N - n_1}\pi(n_1, n_2). \tag{16}$$

2) Quality-based channel assignment: In quality-based channel assignment, AP1 assigns an idle channel to an incoming PT call according to the channel quality recorded in a database that it maintains. Channel quality is measured in terms of the channel noise level, and the channel with the smallest noise level is assigned first. That is, AP1 assigns the idle channel with the smallest noise level to an incoming PT call. Intuitively, application of this rule should reduce the total collision probability between the two types of calls and thus reduce the interference, since channels occupied by ST calls are treated by AP1 as idle channels with larger noise and have a smaller probability of being selected. Assuming that each channel has the same level of background noise, channels occupied by ST calls will not be selected for an incoming PT call unless no idle channel remains (i.e., the system is full). Therefore, the collision probability for quality-based channel assignment, denoted by P_C', is determined as

$$P_C' = \sum_{n_2=1}^{N} \pi(N - n_2, n_2). \tag{17}$$

The relation between P_C and P_C' can be easily derived as

$$P_C = P_C' + \sum_{n_2=1}^{N} \sum_{n_1=0}^{N-n_2-1} \frac{n_2}{N - n_1}\pi(n_1, n_2), \tag{18}$$

which suggests that in an OSS system, quality-based channel assignment can indeed greatly reduce the probability of collision.
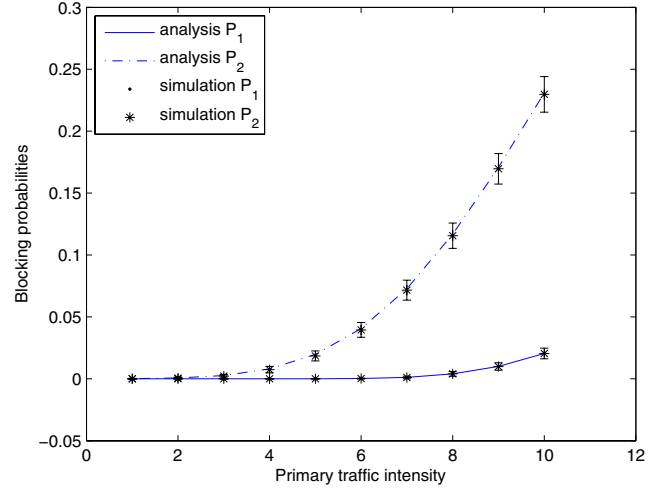
V. NUMERICAL RESULTS

Fig. 5.　Analysis and simulation results for PT and ST call blocking probabilities ($p_f = 0.05$, $p_m = 0.05$, $\rho_2 = 5$).
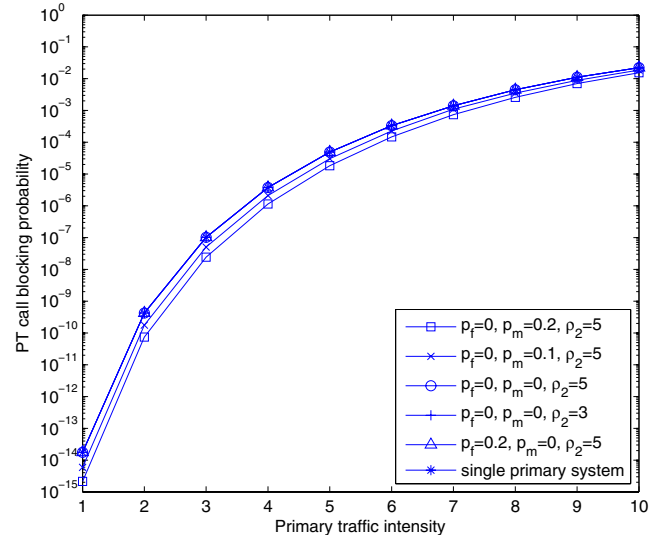

Fig. 6.　PT call blocking probability.

In this section, we present numerical results for the obtained performance measures under the following parameter settings: $N = 16$, $\mu_1 = 10$, $\mu_2 = 10$, $r_2 = 5$. Time is represented in terms of a dimensionless time unit, which can be mapped to a specific unit of time. To show the performance benefit of the OSS system, we include the performance of the original primary system, without the secondary system, as a baseline
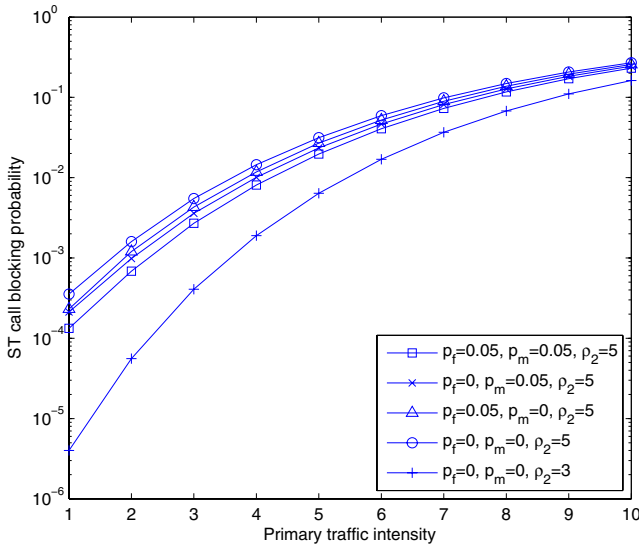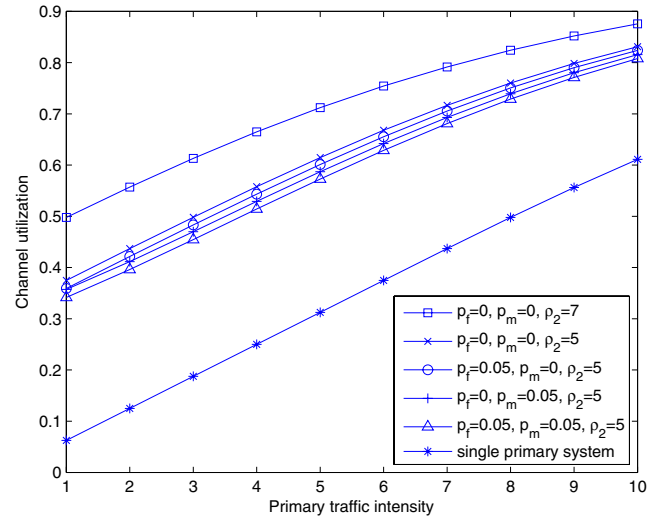
Fig. 7.   ST call blocking probability.

Fig. 8.   Mean reconnection probability of the queued ST calls.
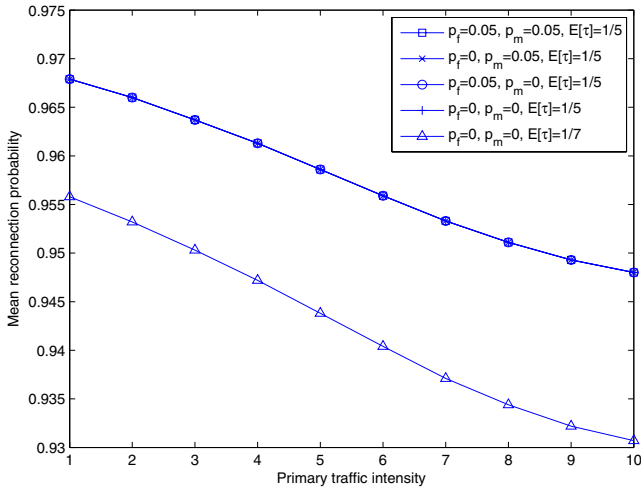

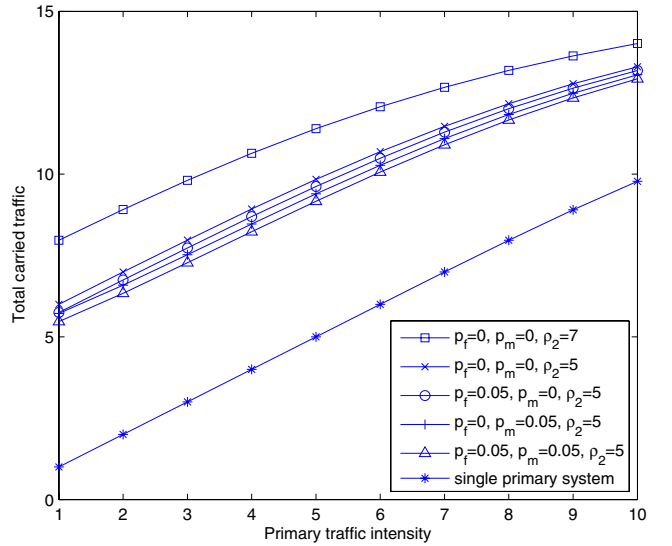Fig. 9.   Channel utilization of the OSS system.

Fig. 10.   Total carried traffic of the OSS system.

case for comparison. We also study the performance impact of unreliable spectrum sensing by choosing different values of the false alarm and misdetection probabilities, p_f and p_m, respectively[4].

We first validate the analysis of blocking probabilities by comparing the analytical results with results from simulations. The PT and ST call blocking probabilities are both calculated and simulated over a wide range of primary traffic intensity ρ_1. One set of results, shown in Fig. 5, illustrates an excellent match between analysis and simulation. The error bars show 95% confidence intervals obtained by running 10,000 simulation trials for each point.

The simulation was implemented in MATLAB. Arrivals of PT and ST calls to the system are generated according to Poisson processes with rates λ_1 and λ_2, respectively. When the system is not in a full state, arrivals of ST calls and completions of PT calls will be affected by the false alarm and misdetection probabilities ($p_f = p_m = 5\%$) according to

the state transition diagram in Fig. 2. The output measures of the simulation are the total number of call arrivals of PT calls (N_1) and ST calls (N_2) to the system, the total number of call arrivals of ongoing ST calls (N_3) to the buffer, the blocked PT calls (N_{01}) and blocked ST calls (N_{02}) from the system, and the number of waiting ST calls eventually lost from the buffer (N_{03}). The above outputs are used to compute the performance metrics of interest, such as the blocking probabilities $P_1 = N_{01}/N_1$, $P_2 = N_{02}/N_2$.

Fig. 6 shows the PT call blocking probability P_1. We observe that P_1 increases with the PT intensity ρ_1 and does not depend on the ST intensity ρ_2 or the false alarm probability p_f. On the other hand, we observe that P_1 decreases a little with an increase in the misdetection probability p_m. The reason is that the type-I misdetection event causes the release of a working channel, leading to an additional opportunity for new incoming PT call requests. Fig. 7 shows the ST call blocking probability P_2. We observe that P_2 increases with increasing ρ_1 or ρ_2, as should be expected. We also observe the impact of unreliable
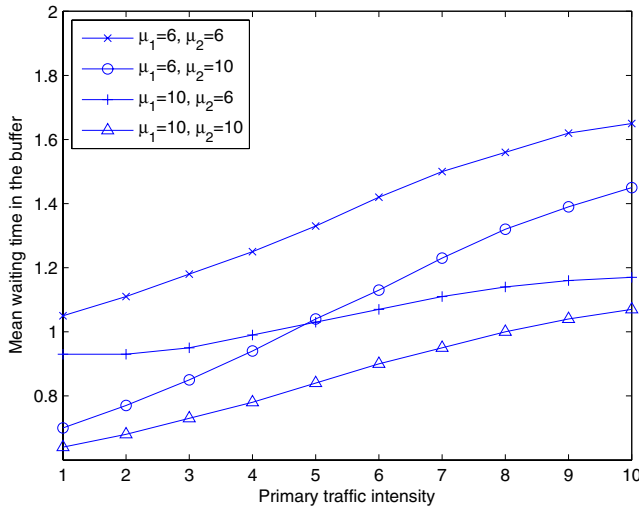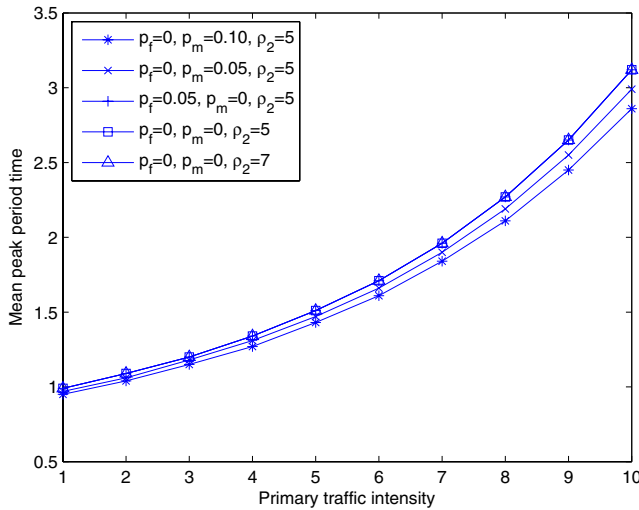
---

[4]Due to space limitations, we only present results related to type-I misdetection and omit those related to type-II misdetection.

Fig. 11. Mean waiting time of the ST calls in the buffer.

Fig. 12. Mean peak period time of PT calls ($M = 11$).

spectrum detection on P_2. When p_f or p_m increases, P_2 decreases slightly. This can be explained as follows. When a false alarm event occurs, the channel remains idle, potentially to be used by other ST call requests. On the other hand, an occurrence of a type-I misdetection event directly results in a release of a channel that was being used by a primary user, leading to an additional opportunity for new incoming ST call requests. However, a misdetection event clearly degrades the performance of the OSS system as seen by primary users. In addition, by comparing Figs. 6 and 7, we observe that the ST call blocking probability is higher than that of PT calls under the same parameter settings, as should be expected.

Fig. 8 shows the impact of various parameters on the mean reconnection probability. We observe that the mean reconnection probability γ of an ST call decreases as ρ_1 increases, and increases as the mean value, $E[\tau]$ of the maximum ST call queueing time τ (see Section IV.B), is increased. The reason is as follows: a higher volume of PT calls results in a smaller chance that a queued ST call reconnects to the system, while a longer maximum queueing time leads to a higher chance of

reconnection. We also note that γ depends on neither p_f nor p_m.

In Fig. 9, we observe that the channel utilization of the OSS system η increases with increasing ρ_1 or ρ_2, and η decreases as the probability of spectrum sensing error increases. Note that the case ($p_f = 0$, $p_m = 0.05$) results in lower channel utilization than the case ($p_f = 0.05$, $p_m = 0$). A false alarm event only wastes an idle channel, while a misdetection event not only wastes an idle channel but also causes an active channel to become an idle channel. Thus, a misdetection event degrades system performance more severely than a false alarm event. It is also observed that the channel utilization of the OSS system is much higher than that of the single primary system, even with unreliable spectrum sensing. Fig. 10 shows that the total carried traffic has a similar performance trend with respect to channel utilization.

Fig. 11 shows the impact of various channel holding times and the primary traffic intensity on the mean waiting time of the ST calls in the buffer. We observe that the mean waiting time decreases as the channel holding time of PT and/or ST calls decreases, and increases as the PT intensity increases. The reason is that the smaller the channel holding time, the faster a PT or ST call leaves the system, and consequently, the waiting time for a queued ST call to reconnect back will be smaller. On the other hand, when the PT traffic intensity is increased, the queued ST calls in the buffer will have less opportunity to reconnect back, since additional PT calls have to be served. By comparing the curves corresponding to ($\mu_1 = 6, \mu_2 = 10$) and ($\mu_1 = 10, \mu_2 = 6$), respectively, with the curve corresponding to ($\mu_1 = 10, \mu_2 = 10$), we see that changing μ_1 has greater impact on the mean waiting time than changing μ_2 at high PT intensity, and has less impact at low PT intensity. This is reasonable since the service rate for primary traffic plays a critical role at high PT intensity and is relatively unimportant at low PT intensity.

In Fig. 12, we study the performance of the mean peak period time of PT calls $\overline{B_p}$ with $M = 11$. We observe that $\overline{B_p}$ does not change with the change of secondary traffic intensity or false alarm probability p_f. The reason for the former event is that in our system model, PT calls operate as if there are no ST calls in the system; and the reason for the latter is obvious from the definition of p_f. We also observe that $\overline{B_p}$ increases as the PT intensity ρ_1 increases and as the misdetection probability p_m decreases. As ρ_1 increases, the fraction of time for which the number of PT calls remains between M and N increases. When p_m decreases, the probability that channels occupied by PT calls will be released incorrectly reduces, which leads to a longer peak period time of PT calls.

VI. CONCLUSION

We presented an analytical model of a wireless network with opportunistic spectrum sharing (OSS) under unreliable spectrum sensing. In the OSS system, the secondary users opportunistically share a set of spectrum resources with the primary users over a coverage area. The secondary users detect channels that are unused by the primary users and then make use of the idle channels. Unreliable spectrum sensing is modeled using false alarm and misdetection probabilities. The impact of these events on system performance is evaluated.

We derived expressions for the blocking probabilities of primary calls and secondary calls, mean reconnection probability, channel utilization, total carried traffic in the system, and other useful performance metrics. Our results suggest that opportunistic spectrum sharing can significantly improve spectrum efficiency and system capacity, even under unreliable spectrum sensing. The proposed model and analysis method can be used to evaluate the performance of future opportunistic spectrum sharing systems.

APPENDIX

A. Glossary of Notation

N	total number of channels [Sec. II]
p_m, p_f	misdetection, false alarm prob. [Sec. II]
λ_1, λ_2	arrival rates of PT and ST calls [Sec. III]
μ_1^{-1}, μ_2^{-1}	mean channel holding times [Sec. III]
h_1^{-1}, h_2^{-1}	mean call holding times [Sec. III]
r_1^{-1}, r_2^{-1}	mean cell residence times [Sec. III]
P_1, P_2	blocking prob. [Sec. IV.A]
γ	mean reconnection prob. [Sec. IV.B]
τ	max. queueing time in buffer [Sec. IV.B]
$E[\tau]$	mean value of τ [Sec. V]
η	total channel utilization [Sec. IV.C]
TCT	total carried traffic [Sec. IV.C]
$\overline{W_b}$	mean ST call waiting time [Sec. IV.D]
$B_P(M)$	peak period time of PT calls [Sec. IV.E]
P_C	collision prob. [Sec. IV.F]

REFERENCES

[1] S. Tang and B. L. Mark, "Performance analysis of a wireless network with opportunistic spectrum sharing," in *Proc. of IEEE Globecom'07*, (Washington, D.C., USA), Nov. 2007.
[2] M. McHenry, "Frequency agile spectrum access technologies," in *Proc. FCC Workshop on Cognitive Radio*, May 2003.
[3] G. Staple and K. Werbach, "The end of spectrum scarcity," *IEEE Spectrum*, vol. 41, pp. 48–52, March 2004.
[4] A. E. Leu, M. McHenry, and B. L. Mark, "Modeling and analysis of interference in listen-before-talk spectrum access schemes," *Int. J. Network Mgmt*, vol. 16, pp. 131–147, July 2006.
[5] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 201–220, Feb. 2005.
[6] W. R. Young, "AMPS: Introduction, background, and objectives," *Bell System Technical Journal*, vol. 58, pp. 1–14, March/April 2005.
[7] S. M. Redl, M. K. Weber, and M. W. Oliphant, *An Introduction to GSM*. Artech House, March 1995.
[8] V. K. Garg, K. Smolik, and J. E. Wilkes, *Applications of CDMA in Wireless/Personal Communications*. Prentice Hall, Oct. 1996.
[9] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Downlink Scheduling and Resource Allocation for OFDM Systems," in *Proc. 40th Conf. on Info. Sci. and Syst. (CISS)*, pp. 1272–1279, Mar. 2006.
[10] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. 1st IEEE Symp. on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 131–136, Nov. 2005.
[11] B. Wild and K. Ramchandran, "Detecting primary receivers for cognitive radio applications," in *Proc. IEEE DySPAN'05*, pp. 124–130, Nov. 2005.
[12] S. Geirhofer, L. Tong, and B. Sadler, "A measurement-based model for dynamic spectrum access in wlan channels," in *Proc. IEEE Milcom'06*, pp. 1–7, Oct. 2006.
[13] A. Mishra, "A multi-channel MAC for opportunistic spectrum sharing in cognitive networks," in *Proc. IEEE Milcom'06*, pp. 1–6, Oct. 2006.
[14] X. Liu and S. Shankar N., "Sensing-based opportunistic channel access," *Mobile Networks and Applications*, vol. 11, pp. 577–591, Aug. 2006.
[15] P. Pawelczak, R. V. Prasad, and R. Hekmat, "Waterfilling may not good neighbors make," in *Proc. IEEE ICC'07*, June 2007.
[16] Y. Fang, Y.-B. Lin, and I. Chlamtac, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Trans. Comput.*, vol. 47, pp. 679–692, June 1998.
[17] Y.-R. Huang, Y.-B. Lin, and J. M. Ho, "Performance analysis for voice/data integration on a finite mobile systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 367–378, Feb. 2000.
[18] B. Li, L.-Z. Li, B. Li, and X.-R. Cao, "On handoff performance for an integrated voice/data cellular system," *Wireless Networks*, vol. 9, pp. 393–402, Mar.-Apr. 2003.
[19] W. Li and X. Chao, "Modelling and performance evaluation of a cellular mobile network," *IEEE/ACM Trans. Netw.*, vol. 12, pp. 131–145, Feb. 2004.
[20] S. Tang and W. Li, "Performance analysis of a channel allocation scheme for multi-service mobile cellular networks," *Int. J. of Communication Systems (IJCS)*, vol. 20, pp. 177–205, Feb. 2007.
[21] S. Tang and W. Li, "An adaptive bandwidth allocation scheme with preemptive priority for integrated voice/data mobile networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 2874–2886, Oct. 2006.
[22] H. Kobayashi and B. L. Mark, *System Modeling and Analysis: Foundations of System Performance Evaluation*. Upper Saddle River, NJ: Pearson Education, Inc., 2009.
[23] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins University Press: Baltimore, MD, 1981.

Shensheng Tang (SM'07) received his B.S. degree from Tianjin University, Tianjin, China, and M.S. degree from China Academy of Telecommunications Technology (CATT), Beijing, China, both in Electronic Engineering. He received his Ph.D. in Electrical Engineering from University of Toledo, Ohio. He has over 8 years of industrial R&D experience in the field of information technology. Currently, he is a research fellow in the Department of Electrical and Computer Engineering at George Mason University. His research interests focus on wireless networking and mobile computing, modeling and performance evaluation, digital signal processing, and network security.

Brian L. Mark (M'91) received the B.A.Sc. degree in Computer Engineering with an option in Mathematics from the University of Waterloo, Canada, in 1991 and the Ph.D. in Electrical Engineering from Princeton University, Princeton, NJ, in 1995. He is currently an Associate Professor in the Dept. of Electrical and Computer Engineering at George Mason University. His research interests lie in the design, modeling, and analysis of communication systems, computer systems, and communication networks.