Analysis of Reactive Spectrum Handoff in Cognitive Radio Networks

Chung-Wei Wang and Li-Chun Wang, *Fellow, IEEE*

***Abstract*—In this paper, we present an analytical framework to evaluate the effects of multiple spectrum handoffs on channel utilization and latency performances in cognitive radio (CR) networks. During the transmission period of a secondary connection, multiple interruptions from the primary users result in multiple spectrum handoffs. In order to decide the target channel for each spectrum handoff and resume the unfinished transmission, wideband sensing is performed in an on-demand reactive manner. Although spectrum handoff procedure can enhance channel utilization, transmission latency of the secondary users is prolonged due to multiple handoffs. Thus, two fundamental issues in CR networks with multiple spectrum handoffs arise: (1) to what extent the channel utilization can be improved; and (2) how long the transmission latency will be extended for the secondary users. To solve the first problem, we introduce the preemptive resume priority (PRP) M/G/1 queueing network to characterize the channel usage behaviors of CR networks. Based on this queueing network, channel utilization under various traffic arrival rates and service time distributions can be evaluated. Furthermore, on top of the proposed queueing network, a state diagram is developed to characterize the effects of multiple handoff delay on the transmission latency of the secondary users. The analytical results can provide a helpful insight to study the effects of traffic arrival rates and service time on channel utilization and transmission latency and then facilitate the designs of admission control rules for the secondary users subject to their performance requirements.**

***Index Terms*—Cognitive radio, spectrum handoff, spectrum mobility, extended data delivery time, preemptive priority, preemption, queueing theory.**

I. Introduction

COGNITIVE radio (CR) is an emerging technique to promote spectrum efficiency [1]–[6]. In CR networks, the low-priority secondary users can access the unused licensed spectrum of the high-priority primary users to transmit data. However, the secondary users must vacate the occupied channel when the primary user appears because the primary users have the preemptive priority to access channels. In order to return the occupied channel to the primary users and resume the unfinished transmission at the suitable channel, the *spectrum handoff* procedures are initiated for the interrupted secondary user [7]–[9]. Based on this dynamic spectrum access scheme, spectrum efficiency can be improved.

Manuscript received 5 January 2012; revised 16 May 2012. This paper is supported by National Science Council, Taiwan, under the contract NSC-100-3113-P-009-001. This work was presented in part at the IEEE Global Communications Conference (GLOBECOM), 2010.

C.-W. Wang is with MStar Semiconductor, Inc.,Taipei, Taiwan (e-mail: hyper.cm91g@nctu.edu.tw).

L.-C. Wang is with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan (e-mail: lichun@cc.nctu.edu.tw).

Digital Object Identifier 10.1109/JSAC.2012.121116.

In CR networks, a secondary user's connection may encounter multiple interruptions from the primary users during its transmission period. These interruptions result in a series of spectrum handoffs because spectrum handoff procedures are performed whenever an interruption event occurs. Hence, a series of target channels will be selected sequentially for multiple handoffs. Clearly, these handoffs will increase the secondary connections' extended data delivery time although spectrum handoff functionality can enhance channel utilization. Here, the extended data delivery time is defined as the duration from the instant of starting transmitting data until the instant of finishing the whole transmission.

According to the decision timing for selecting target channels, spectrum handoff mechanisms can be categorized as either the proactive-decision spectrum handoff or the reactive-decision spectrum handoff schemes [10].

- In the proactive-decision spectrum handoff, the target channels for future spectrum handoffs are determined according to the long-term traffic statistics *before* data connection is established. The secondary user can immediately change its operating channel to the predetermined target channel whenever it is interrupted [11]–[13].
- In the reactive-decision spectrum handoff, the target channel is searched in an on-demand manner. *After* spectrum handoff request is made, the target channel is determined according to the result from instantaneous spectrum sensing [14], [15]. Then, the interrupted secondary user can resume the unfinished transmission on the selected channel.

In this paper, we focus on the modeling technique and performance analysis for the reactive-decision spectrum handoff scheme, while the related studies on the proactive-decision spectrum handoff has been reported in [16], [17]. Compared to the proactive-decision spectrum handoff scheme that the preselected target channel may no longer be available at the instant that spectrum handoff procedures are initiated, the reactive-decision spectrum handoff may have shorter handoff delay because it can reliably find an idle channel through spectrum sensing. Nevertheless, the reactive-decision spectrum handoff scheme needs the sensing time to search the idle channels. In addition, it also needs the handshaking time to achieve a consensus on the target channel between the transmitter and receiver of a secondary connection. Hence, one important issue for the reactive-decision spectrum handoff scheme is to characterize the effects of the sensing time and the handshaking time on the handoff delay. Obviously, when the sensing time and the handshaking time are too large, the reactive-decision spectrum handoff is worse than the proactive-decision spectrum handoff in terms of the extended data delivery time.

The goal of this paper is to investigate the effects of spectrum handoffs on the channel utilization and the extended data delivery time of the secondary users' connections with various traffic arrival rates and service time distributions. We consider the three key design features for spectrum hanodff, consisting of (1) heterogeneous arrival rates of the primary users at different channels, where the traffic arrival rates may be different at various channels because these channels may belong to different primary system operators; (2) various arrival rates of the secondary users at different channels, where the arrival rates can be determined by the initial operating channel selection mechanisms [18]; and (3) handoff processing time, resulting from the sensing time, the handshaking time, and the channel switching time. How to model the channel utilization at each channel and the extended data delivery time in the context of multiple handoffs is challenging since the operating channels for multiple handoffs are selected according to the channel occupancy states at the moments of link transitions. To the best of our knowledge, an analytical model for characterizing all the three features for multiple handoffs has rarely been seen in the literature. The contributions of this paper are summarized in the following:

- First, the preemptive resume priority (PRP) M/G/1 queueing network model is proposed to characterize the channel usage behaviors of CR networks. Based on this queueing model, we can evaluate the channel utilizations of each channel under various traffic arrival rates and service time distributions.
- Next, a state diagram is developed to characterize the effect of multiple handoff delay on the extended data delivery time of the secondary connections. Then, we can evaluate how long the extended data delivery time is prolonged due to multiple spectrum hanodffs.

The rest of this paper is organized as follows. Section II reviews the current spectrum usage models for the reactive-decision spectrum handoff schemes in the literature. Section III presents an illustrative example for multiple handoff issue. Next, in order to characterize the spectrum usage behaviors with multiple handoffs, we propose the PRP M/G/1 queueing network model in Section IV. Based on this model, Section V evaluates the channel utilization factors under various traffic arrival rates and service time distributions. Furthermore, we develop a state diagram to characterize the effects of handoff delay on the extended data delivery time of the secondary connections in Section VI. Analytical and simulation results are given in Section VII. Finally, we give our concluding remarks in Section VIII.

II. Related Work

A key property of reactive-decision spectrum handoff is that the interrupted secondary user can actually find the idle channel if at least one idle channel exists at the moment of link transition. In order to characterize the channel usage behaviors with this property, we should consider three key design features, consisting of (1) heterogeneous arrival rates of the primary users (PUs); (2) various arrival rates of the secondary users (SUs); and (3) handoff processing time. Based on the three features, Table I classifies the existing modeling techniques of the reactive-decision spectrum handoff. In this table, the signs " $\circ$ " and " $\times$ " indicate that the proposed model "does" and "does not" consider the corresponding feature, respectively. In the literature, the modeling techniques for reactive-decision spectrum handoff behaviors can be categorized into the following four types: (1) ON/OFF random process; (2) M/M/m queueing model; (3) multi-dimensional Markov chain; and (4) M/G/1 queueing model. One can observe that the current modeling techniques have not considered all the aforementioned three design features. In the following, we briefly discuss the features of these analytical models for spectrum handoff behaviors.

- **ON/OFF random process**: In [19]–[21], the ON/OFF random process was used to characterize the channel usage behaviors of the primary networks at each channel, where the distributions of ON (busy) period- and OFF (idle) period at each channel are geometrical distributed. The OFF state can be regarded as a potential spectrum opportunity for the secondary users. The authors in [19] and [20] calculated the channel utilization factors and the extended data delivery time of the secondary users, respectively. Unlike [19], [20] that did not address the effects of spectrum sensing time, the authors in [21] examined the effects of spectrum sensing time on the extended data delivery time of the secondary users. However, [21] assumed that at least one channel is certainly available after spectrum sensing, and the case that all channels are busy after spectrum sensing did not been considered.
- **M/G/m queueing model**: In [22], the channel usage behaviors of the primary users are characterized by the M/G/m queueing model, where m is the total number of channels in the CR network. When interruption occurs, the secondary user can transfer to the idle channel instantly without delay if idle channel exists. Otherwise, the secondary user must wait until the instant that one channel becomes idle. This waiting time had been investigated in [22]. Note that the sensing time had not been considered in their model.
- **Multiple-dimensional Markov chain**: In [23]–[37], the spectrum usage behaviors of both the primary and secondary networks were modeled by a two-dimensional Markov chain, where the two dimensions represent the total numbers of the primary and the secondary users in a CR network, respectively. The blocking probability and forced termination probability for the secondary users' connections in the CR network without and with queue are studied in [23]–[32] and [33]–[35], respectively. Different from [23]–[35] that considered infinite user population, [36], [37] derived the blocking probability in a CR network with finite user population. Furthermore, the authors in [38]–[41] further extended the two-dimensional Markov chain model to the multiple-dimensional Markov chain, where the new dimension is used to describe the channel state, queue length, or the number of interfered channels. Note that these analytical models are suitable

TABLE I
COMPARISON OF VARIOUS CHANNEL USAGE MODELS.

Model Name	Heterogeneous Arrival Rates of PUs	Various Arrival Rates of SUs	Handoff Processing Time
ON/OFF Random process [19], [20]	×	×	×
ON/OFF Random Process [21]	×	×	○
M/G/m Queueing Model [22]	○	×	×
Multi-Dimensional Markov Chain [23]–[41]	×	×	×
M/G/1 Queueing Model [42], [43]	○	○	×
Proposed Unifying Model	○	○	○

for the CR network with homogeneous traffic loads, and the issues of heterogeneous arrival rates of the primary and the secondary users has not been addressed.

- **M/G/1 queueing model**: [42], [43] used the M/G/1 queueing model to characterize the channel usage behaviors of a secondary network, where each secondary user can simultaneously use all idle channels to transmit its data. Because the total number of idle channels depends on how many channels are occupied by the primary users, the service rates of the secondary users are related to the traffic statistics of the primary users, resulting in a non-trivially distributed service time. Thus, the authors suggested using the M/G/1 queueing system to characterize this system. However, the proposed model did not consider the effect of handoff processing time.

In this paper, we propose a unifying model, which consists of PRP M/G/1 queueing network and state transition model. This model takes into account of all the design features, consisting of heterogeneous arrival rates of the primary users, various arrival rates of the secondary users, and the handoff processing time. In the following, we will discuss the proposed unifying model in more details.

III. SYSTEM MODEL

A. Assumptions

In this paper, we assume that the considered CR network is a time-slotted system as [13], [44]–[47], where each secondary user must perform spectrum sensing at the beginning of each time slot to detect the presence of primary users. The secondary user can transmit or receive data in the remaining duration of this time slot if the current operating channel is assessed as idle. Otherwise, the secondary user will perform spectrum handoff procedures to find the idle channels and then resume its unfinished transmission at one of the idle channels.

We consider a CR network with M independent channels, where each channel has a virtual high-priority and a virtual low-priority queues as discussed in [48]. The traffic of the primary and secondary users is connected to the high-priority and the low-priority queues before transmission. Then, the *primary connections* and the *secondary connections* are respectively established for the primary and secondary users according to the time that traffic arrives at queues. Here, for the same priority users' connections, we assume that they are established by following the first-come-first-served (FCFS) scheduling principle.

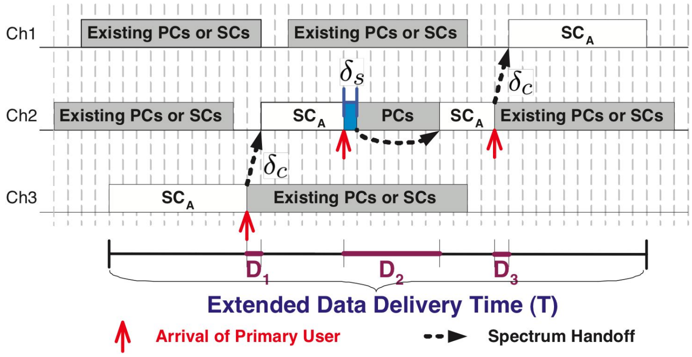


Fig. 1. An example of transmission process for the secondary connection SC_A, where T is the extended data delivery time of SC_A and D_i is the handoff delay of the i^{th} interruption. The gray areas indicate that the channels are occupied by the existing primary connections (PCs) or secondary connections (SCs). Because SC_A is interrupted three times in total, the overall data connection is divided into four segments.

In this paper, we consider the spectrum handoff protocol presented in [10]. When the spectrum handoff procedures are initiated, the secondary users must spend τ slots on spectrum sensing to find the idle channels. If more than one channel is assessed as idle, the interrupted secondary user will randomly select one idle channel from all idle channels to be its target channel for spectrum handoff. Here, we assume that this random selection follows the uniform distribution. Furthermore, the interrupted secondary user will stay on the current operating channel if all channels are busy. Next, the handshaking time of t_h slots is spent in order to achieve a consensus on the target channel between the transmitter and the receiver of a secondary connection. Hence, when a secondary user changes its operating channel to another channel, the total processing time for executing spectrum handoff procedures is $\delta_c \triangleq \tau + t_h + t_s$ where t_s (slots) is the channel switching time. On the other hand, if the secondary user stays on the current operating channel, the total processing time is $\delta_s \triangleq \tau + t_h$.

B. Spectrum Handoffs with Multiple Interruptions

During the transmission period, a secondary connection may encounter multiple interruption requests from the primary users. Because these interruptions result in multiple handoffs, a series of target channels, called the *target channel sequence*, is selected according to spectrum sensing results. Fig. 1 shows an example that three spectrum handoffs occur during the transmission period of the secondary connection SC_A, where SC_A's initial (default) channel is Ch3. We assume that the transmitter of SC_A wants to establish a connection flow with 30 slots to the intended receiver. Its extended data delivery time is denoted by T. Furthermore, D_i is the handoff delay of the i^{th} interruption. Here, handoff delay is defined as the duration from the instant that transmission is interrupted until the instant that the unfinished transmission is resumed. The transmission process with multiple handoffs is described as follows:

1) In the beginning, SC_A is established at its default channel Ch3. When an interruption event occurs, SC_A must perform spectrum sensing to search the idle channel for spectrum handoff.
2) At the first interruption, SC_A changes its operating channel to the idle channel Ch2 from Ch3. Thus, the handoff delay D_1 is δ_c.
3) At the second interruption, SC_A stays on its current operating channel Ch2 because all channels are busy. SC_A cannot be resumed until all the high-priority primary connections at Ch2 finish their transmissions. In this case, the handoff delay D_2 is the sum of δ_s and the duration from the time instant that Ch2 is used by the primary users' connections until the time instant that the high-priority queue becomes empty. This duration (denoted by $Y_p^{(2)}$) is called the *busy period* resulting from the transmissions of multiple primary connections at Ch2.
4) At the third interruption, SC_A finds both Ch1 and Ch3 are idle. Then, SC_A randomly selects one channel to be the target channel. In this example, SC_A selects Ch1 to be its target channel. Note that the handoff delay D_3 in this case is δ_c.
5) Finally, SC_A is completed on Ch3.

Hence, SC_A's *target channel sequence* is (Ch2, Ch2, Ch1) in this example.

IV. The PRP M/G/1 Queueing Network Model

A. Key Features and Assumptions

In this section, we introduce a PRP M/G/1 queueing network model to characterize the spectrum usage behaviors of the primary and the secondary connections with multiple spectrum handoffs between different channels. Some key features of the PRP M/G/1 queueing network model are listed below:

- Each server (channel) can serve two types of customers (connections): the high-priority connections from the primary users and the low-priority connections from the secondary users.
- The primary user has the preemptive priority to interrupt the transmission of the secondary users. Instead of retransmitting the whole connection [11], the interrupted secondary user can resume the unfinished transmission. Note that the interrupted secondary connection's target channel can be different from its current operating channel. This is a key difference to the traditional PRP M/G/1 queueing theory.
- A secondary connection may experience multiple interruptions from the primary users during its transmission period.

Furthermore, some assumptions are adopted for the ease of analysis.

- In order to balance the overall traffic loads of the secondary users to all channels, each secondary user was preassigned a default channel through spectrum decision algorithm [18]. When a secondary transmitter wants to establish a new connection for data transmission, it must transmit handshaking signal at the default channel of the intended receiver firstly [49]. If the default channel of the intended receiver is busy, the secondary transmitter must wait at this channel until it becomes available [44].
- Each primary connection has a default channel.
- The presence of the primary user can be detected by the secondary user. How to extend the proposed model to consider the effects of false alarm and missed detection were discussed in [18].
- At one channel, only one user can transmit data at each time slot.

B. Example

Figure 2 shows an example of the PRP M/G/1 queueing network model with three channels. The traffic flows of the primary connections and the secondary connections are directly connected to the high-priority queue and the low-priority queue, respectively. A secondary connection will be interrupted when the primary connection appears at the occupied channel. The interrupted secondary connection can decide its target channel for spectrum handoff according to the instantaneous spectrum sensing outcomes. Note that the required time for spectrum sensing is modeled in $\boxed{\text{S}}$. It can be regarded as a tapped delay line or a server with constant service time, which equals to the handoff processing time. In the proposed queueing network, the interrupted secondary connection can either stay on its current channel or change to another channel through different feedback paths. When all channels are busy, the interrupted secondary connection chooses to stay on its current operating channel. In this case, its remaining data will be connected to the head of the low-priority queue of its current operating channel. On the other hand, when one idle channel exists, the decision is to change its operating channel. Thus, the remaining data of the interrupted secondary connection will be connected to the empty low-priority queue of the selected idle channel. In the figure, $\oplus$ represents that the traffic of the interrupted secondary connection is merged. Furthermore, when the interrupted secondary connection transmits the remaining data on the target channel, it may be interrupted again. Hence, this model can incorporate the effects of multiple interruptions.

C. Traffic Parameters

The traffic parameters required in the proposed PRP M/G/1 queueing network model are summarized as follows. We assume that the arrival processes of the primary and the secondary connections on the virtual queues of each channel are

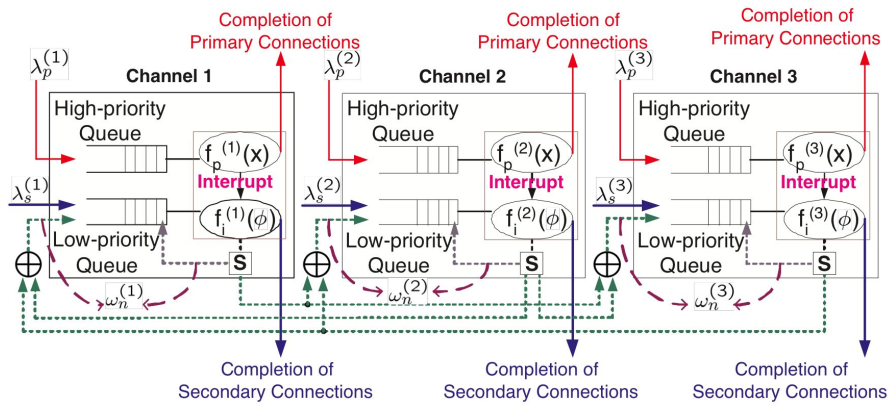


Fig. 2. The PRP M/G/1 queueing network model with three channels where $\lambda_p^{(k)}$, $\lambda_s^{(k)}$, and $\omega_n^{(k)}$ are the arrival rates of the primary connections, the secondary connections, and the type-(n) secondary connections ($n \geq 1$) at channel k. Note that $\lambda_s^{(k)} = \omega_0^{(k)}$. Furthermore, $f_p^{(k)}(x)$ and $f_i^{(k)}(\phi)$ are the probability density functions (PDFs) of $X_p^{(k)}$ and $\Phi_i^{(k)}$, respectively.

Poisson. The traffic arrival rate of the primary connections at channel η is denoted by $\lambda_p^{(\eta)}$ (arrivals/slot), and the secondary connection's initial arrival rate at channel η is denoted by $\lambda_s^{(\eta)}$ (arrivals/slot). Furthermore, denote $X_p^{(\eta)}$ (slots/arrival) and $X_s^{(\eta)}$ (slots/arrival) as the corresponding service time of the primary and the secondary connections, respectively, and let $f_p^{(\eta)}(x)$ and $f_s^{(\eta)}(x)$ be the probability density functions (PDFs) of $X_p^{(\eta)}$ and $X_s^{(\eta)}$, respectively. When the following four traffic parameters $\lambda_p^{(\eta)}$, $\lambda_s^{(\eta)}$, $f_p^{(k)}(\eta)$, and $f_s^{(k)}(\eta)$ are obtained by certain traffic pattern prediction methods [50], we can derive many performance measures for the multi-channel spectrum handoffs with multiple interruptions.

Now, we define some notations as follows. We call the secondary connection that has experienced i interruptions the type-(i) secondary connection, where $i \geq 0$. Firstly, we consider the type-(i) secondary connections whose default channels are channel η. Two important system parameters $\omega_{i,\eta}^{(k)}$ and $\Phi_{i,\eta}^{(k)}$ are defined as follows:

- $\omega_{i,\eta}^{(k)}$ is the arrival rate of the considered secondary connections at channel k. Note that $\omega_{0,\eta}^{(\eta)} = \lambda_s^{(\eta)}$. How to derive $\omega_{i,\eta}^{(k)}$ from the four traffic parameters is discussed in Section V.
- $\Phi_{i,\eta}^{(k)}$ is the effective service time of the considered secondary connections at channel k. That is, $\Phi_{i,\eta}^{(k)}$ is the considered secondary connection's transmission duration between the i^{th} and the $(i+1)^{th}$ interruptions at channel k. In Section V, we will discuss how to derive $\mathbf{E}[\Phi_{i,\eta}^{(k)}]$ from the four traffic parameters.

Next, let $\omega_i^{(k)}$ and $\Phi_i^{(k)}$ be the arrival rate and the effective service time of the type-(i) secondary connections at channel k, respectively. We can have

$$\omega_i^{(k)} = \sum_{\eta=1}^{M} \omega_{i,\eta}^{(k)} \ , \tag{1}$$

and

$$\mathbf{E}[\Phi_i^{(k)}] = \sum_{\eta=1}^{M} \frac{\omega_{i,\eta}^{(k)}}{\omega_i^{(k)}} \mathbf{E}[\Phi_{i,\eta}^{(k)}] \ , \tag{2}$$

respectively.

Finally, we denote $\rho_p^{(k)}$ and $\rho_i^{(k)}$ as the channel busy probabilities resulting from the transmissions of the primary connections and the type-(i) secondary connections whose current operating channels are channel k, respectively. Then, in an M-channel network, we can have

$$\rho_p^{(k)} = \lambda_p^{(k)} \mathbf{E}[X_p^{(k)}] \ , \tag{3}$$

and

$$\rho_i^{(k)} = \omega_i^{(k)} \mathbf{E}[\Phi_i^{(k)}] = \sum_{\eta=1}^{M} \omega_{i,\eta}^{(k)} \mathbf{E}[\Phi_{i,\eta}^{(k)}] \ , \tag{4}$$

respectively. Furthermore, the busy probability of channel k (denoted by $\rho^{(k)}$) shall satisfy the following constraint:

$$\rho^{(k)} \equiv \rho_p^{(k)} + \sum_{i=0}^{\infty} \rho_i^{(k)} < 1 \ , \tag{5}$$

where $1 \leq k \leq M$. Note that $\rho^{(k)}$ can be also interpreted as the utilization factor of channel k.

The physical meaning of random variable $\Phi_{i,\eta}^{(k)}$ is illustrated in Fig. 3. A two-channel network with the service time $X_s^{(1)}$ and $X_s^{(2)}$ at channels 1 and 2 for the secondary connections is considered. Figure 3(a) shows the three realizations of $X_s^{(1)}$ for channel 1. Similarly, at channel 2, random variable $X_s^{(2)}$ are generated three times in Fig. 3(b). Each secondary connection is divided into many segments because of multiple primary users' interruptions. For example, the first secondary connection in Fig. 3(a) is divided into three segments because it encounters two interruptions in total. The first, second, and third segments are transmitted at channels 1, 1, and 2, respectively. Thus, this secondary connection's default channel is Ch1 and its target channel sequence is (Ch1, Ch2). Note that the first secondary connection in Fig. 3(b) does not have the second segment because it is interrupted only once. In Fig. 3(a), random variables $\Phi_{2,1}^{(2)}$, one of the gray regions, represents the transmission duration between the 2^{nd} and the 3^{rd} interruptions at channel 2 for the secondary connection whose default channel is channel 1. Similarly, random variables $\Phi_{2,2}^{(2)}$, one of the gray regions in Fig. 3(b), represents the transmission duration between the 2^{nd} and the 3^{rd} interruptions at channel 2 for the secondary connection whose default channel is channel

2. Furthermore, random variable $\Phi_2^{(2)}$, one of the gray regions in Fig. 3, represents the transmission duration of a secondary connection between the 2^{nd} and the 3^{rd} interruptions at channel 2. That is, $\Phi_2^{(2)}$ is one of the third segments of the first and the third secondary connections in Fig. 3(a) as well as the third secondary connection in Fig. 3(b).

V. Analysis of Channel Utilization Factor

Based on the proposed PRP M/G/1 queueing network model, we can evaluate many performance measures of CR networks with various traffic parameters. In this section, we show how to evaluate the channel utilization factor $\rho^{(k)}$. Referring to (4) and (5), for each channel k ($1 \le k \le M$), it follows that

$$\rho^{(k)} = \lambda_p^{(k)} \mathbf{E}[X_p^{(k)}] + \sum_{i=0}^{\infty} \left[\sum_{\eta=1}^{M} \omega_{i,\eta}^{(k)} \mathbf{E}[\Phi_{i,\eta}^{(k)}] \right] . \quad (6)$$

Note that $\rho^{(k)}$ is unrelated to channel sensing time τ, channel handshaking time t_h, and channel switching time t_s. In (6), $\lambda_p^{(k)}$ and $\mathbf{E}[X_p^{(k)}]$ are system parameters and can be known in advance. In the following, we will show how to derive $\omega_{i,\eta}^{(k)}$ and $\mathbf{E}[\Phi_{i,\eta}^{(k)}]$.

A. Derivations of $\omega_{i,\eta}^{(k)}$ and $\mathbf{E}[\Phi_{i,\eta}^{(k)}]$

Without loss of generality, we consider a secondary connection whose default channel is channel $\eta \triangleq s_0$ in the following discussions. Its target channel sequence is denoted by $\mathbf{S}^{(\eta)} \triangleq (S_{1,\eta}, S_{2,\eta}, S_{3,\eta}, \cdots)$, where $S_{i,\eta}$ is the target channel at the i^{th} interruption. Note that $S_{i,\eta}$ is a random variable for each $i \ge 1$. It is decided according to the instantaneous sensing results after the i^{th} interruption event occurs. Thus, $\mathbf{S}^{(\eta)}$ is a random sequence. Based on the following discussions, $\omega_{i,\eta}^{(k)}$ and $\mathbf{E}[\Phi_{i,\eta}^{(k)}]$ will be respectively derived in Propositions 4 and 5, and thus we can evaluate the channel utilization $\rho^{(k)}$ in (6).

Lemma 1. *Let $p_{i,\eta}^{(k)}$ be the probability that the considered type-(i) secondary connection is interrupted again at channel k. It follows that*

$$p_{i,\eta}^{(k)} = \lambda_p^{(k)} \boldsymbol{E}[\Phi_{i,\eta}^{(k)}] . \quad (7)$$

Proof of this lemma can be found in [16].

Definition 2. *Let $\boldsymbol{s}_n \triangleq (s_1, s_2, s_3, \cdots, s_n)$ be any target channel sequence which has n elements, where s_i is the selected target channel when a type-($i-1$) secondary connection is interrupted. Note that $\boldsymbol{s}_n \in \Omega^n$, where $\Omega = \{1, 2, \cdots, M\}$.*

Claim 3. *Denote $\boldsymbol{Pr}[S_{i,\eta} = s_i | S_{i-1,\eta} = s_{i-1}]$ as the probability that the considered secondary connection will select channel s_i to be its target channel when an interruption event occurs at channel s_{i-1}. Then, we have (8) in the next page, where $i \ge 1$ and $S_{0,\eta}$ is the default channel $s_0 = \eta$.*

Proof: When an interruption event occurs at channel s_{i-1}, the type-($i-1$) secondary connection must search its target channel s_i for spectrum handoff through spectrum sensing. The probability that one channel is selected to be the target channel is related to the channel busy probabilities of all channels. If all channels are busy, the type-($i-1$) secondary connection will stay on its current operating channel (i.e., $s_i = s_{i-1}$). On the other hand, if these exists one idle channel, the type-($i-1$) secondary connection will change to this idle channel from channel s_{i-1}. Note that this type-($i-1$) secondary connection will randomly and uniformly select one channel from all idle channels to be its target channel if more than one channel is idle. From these observations, we can have (8). ■

Proposition 4. *At channel k, denote $\omega_{i,\eta}^{(k' \to k)}$ as the arrival rate of the type-(i) secondary connections which be redirected from channel k'. Then, we have*

$$\omega_{i,\eta}^{(k)} = \sum_{k'=1}^{M} \omega_{i,\eta}^{(k' \to k)} , \quad (9)$$

where

$$\omega_{i,\eta}^{(k' \to k)} = \omega_{i-1,\eta}^{(k')} \cdot p_{i-1,\eta}^{(k')} \cdot \boldsymbol{Pr}[S_{i,\eta} = k | S_{i-1,\eta} = k'] . \quad (10)$$

Proof: When a type-($i-1$) secondary connection is interrupted at channel k', it will turn into a new arrival of the type-(i) secondary connection at channel k. That is, the traffic loads of the type-(i) secondary connections at channel k come from the remaining traffic loads of the type-($i-1$) secondary connections at any one of M channels. Thus, the arrival rate of the type-(i) secondary connections at channel k can be expressed as (9).

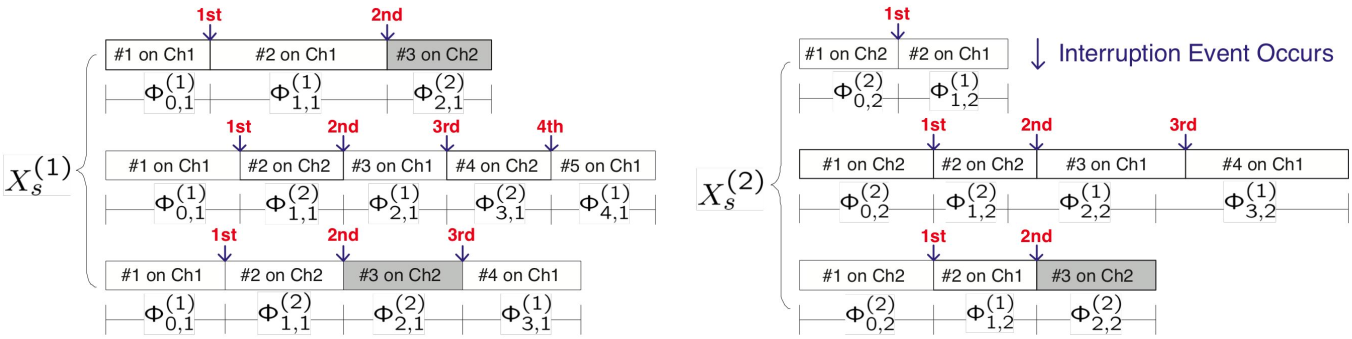
The values of $\omega_{i,\eta}^{(k' \to k)}$ in (9) can be evaluated as follows. For the type-($i-1$) secondary connection at channel k', it will be interrupted again with probability $p_{i-1,\eta}^{(k')}$. When an interruption event occurs at channel k', the type-($i-1$) secondary connection must search its target channel for spectrum handoff through spectrum sensing. Without loss of generality, we assume that channel k is selected to be the target channel. This situation occurs with probability $\mathbf{Pr}[S_{i,\eta} = k | S_{i-1,\eta} = k']$. When channel k is selected, the type-($i-1$) secondary connection will turn into a new arrival of the type-(i) secondary connection at channel k. Hence, we can have (10). ■

Proposition 5. *The closed-form expression of $\boldsymbol{E}[\Phi_{i,\eta}^{(k)}]$ can be derived based on the proposed PRP M/G/1 queueing network model.*

Proof: According to the total probability principle, we have

$$\mathbf{E}[\Phi_{i,\eta}^{(k)}] = \sum_{n=1}^{L} \sum_{\forall \boldsymbol{s}_n \in \Omega^n} \mathbf{Pr}[\mathbf{S}^{(\eta)} = \boldsymbol{s}_n] \mathbf{E}[\Phi_{i,\eta}^{(k)} | \mathbf{S}^{(\eta)} = \boldsymbol{s}_n] , \quad (11)$$

where L is the maximum number of interruptions among all secondary users' connections, i.e., the maximum length of the target channel sequence. Based on the proposed queueing network, $\mathbf{E}[\Phi_{i,\eta}^{(k)} | \mathbf{S}^{(\eta)} = \boldsymbol{s}_n]$ for any given $\boldsymbol{s}_n$ can been derived when $\lambda_p^{(k)}$, $\lambda_s^{(k)}$, $f_p^{(k)}(x)$, and $f_s^{(k)}(x)$ are known. The derivation detail can be found in [16]. As to $\mathbf{Pr}[\mathbf{S}^{(\eta)} = \boldsymbol{s}_n]$, it



(a) Three realizations of $X_s^{(1)}$.

(b) Three realizations of $X_s^{(2)}$.

Fig. 3. Illustration of the physical meaning of random variable $\Phi_i^{(k)}$. For example, $\Phi_2^{(2)}$ is one of the third segments (gray areas) of the first and the third secondary connections in (a) as well as the third secondary connection in (b).

$$\mathbf{Pr}[S_{i,\eta}=s_i|S_{i-1,\eta}=s_{i-1}]=\begin{cases}\displaystyle\prod_{1\le j\le M, j\ne s_{i-1}}\rho^{(j)} & , \quad s_i=s_{i-1}\\ \displaystyle(1-\rho^{(s_i)})\sum_{\forall\mathbb{A}\subseteq\Omega/\{s_{i-1},s_i\}}\left[\frac{1}{1+|\mathbb{A}|}\prod_{\forall\upsilon\in\mathbb{A}}(1-\rho^{(\upsilon)})\prod_{\forall\upsilon'\notin\mathbb{A},\forall\upsilon'\notin\{s_{i-1},s_i\}}\rho^{(\upsilon')}\right] & , \quad s_i\ne s_{i-1}\end{cases} . \tag{8}$$

can be expressed as follows:

$$\mathbf{Pr}[\mathbf{S}^{(\eta)}=\boldsymbol{s}_n]=(1-p_{n,\eta}^{(s_n)})\prod_{i=1}^{n}\mathbf{Pr}[S_{i,\eta}=s_i|S_{i-1,\eta}=s_{i-1}] . \tag{12}$$

By substituting (7) and (8) into (12), the value of $\mathbf{Pr}[\mathbf{S}^{(\eta)}=\boldsymbol{s}_N]$ can be obtained. ■

B. An Example for the Exponentially Distributed Service Time

Now, we show how to derive channel utilization factor in the following special case. We assume that all secondary connections have the same service time distribution. Hence, we have $f_s^{(k)}(x)=f_s(x)$ and $\mathbf{E}[X_s^{(k)}]=\mathbf{E}[X_s]$, where $1\le k\le M$. Furthermore, because this paper focuses on the latency-sensitive traffic for the secondary users, it is reasonable to assume that the service time X_s is exponentially distributed (page 135 in [51]). Hence, we have $f_s(x)=\mu_s e^{-\mu_s x}$, where $\mu_s=\frac{1}{\mathbf{E}[X_s]}$. Based on the above traffic parameters, we derive $\rho^{(k)}$ as follows. Firstly, referring to [5], [16] and Proposition 5, we can have

$$\mathbf{E}[\Phi_{i,\eta}^{(k)}]=\frac{1}{\lambda_p^{(k)}+\mu_s} . \tag{13}$$

Secondly, from (7), it follows that

$$p_{i,\eta}^{(k)}=\lambda_p^{(k)}\mathbf{E}[\Phi_{i,\eta}^{(k)}]=\frac{\lambda_p^{(k)}}{\lambda_p^{(k)}+\mu_s} . \tag{14}$$

Then, according to the Proposition 4, we can derive $\omega_{i,\eta}^{(k)}$ as follows:

$$\omega_{i,\eta}^{(k)}=\sum_{k'=1}^{M}\omega_{i-1,\eta}^{(k')}\cdot\frac{\lambda_p^{(k')}}{\lambda_p^{(k')}+\mu_s}\cdot\mathbf{Pr}[S_{i,\eta}=k|S_{i-1,\eta}=k'] . \tag{15}$$

From (15), we can find that $\omega_{i,\eta}^{(k)}$ is a function of $\rho^{(k')}$ because $\mathbf{Pr}[S_{i,\eta}=k|S_{i-1,\eta}=k']$ is a function of $\rho^{(k')}$ in (8). Furthermore, according to (6), we can find that $\rho^{(k')}$ is a function of $\omega_{i,\eta}^{(k')}$. Hence, we can determine $\omega_{i,\eta}^{(k)}$ and $\rho^{(k)}$ by solving (6) and (15) iteratively.

VI. Analysis of Extended Data Delivery Time

In this section, we show how to evaluate the extended data delivery time, which is an important performance measure for the latency-sensitive traffic of the secondary connections. Without loss of generality, we consider the secondary connection whose default channel is channel η in the following discussions. Its extended data delivery time consists of the original service time $X_s^{(\eta)}$ and the cumulative delay (denoted by $\mathbf{E}[D^{(\eta)}]$) resulting from multiple handoffs. Let $\mathbf{E}[D|\mathbf{S}^{(\eta)}=\boldsymbol{s}_n]$ be its cumulative handoff delay when its target channel is $\boldsymbol{s}_n$. Then, its average extended data delivery time (denoted by $\mathbf{E}[T^{(\eta)}]$) can be expressed as

$$\begin{aligned}\mathbf{E}[T^{(\eta)}] &= \mathbf{E}[X_s^{(\eta)}]+\mathbf{E}[D^{(\eta)}]\\ &= \mathbf{E}[X_s^{(\eta)}]\\ &\quad+\sum_{n=1}^{L}\sum_{\forall\boldsymbol{s}_n\in\Omega^n}\mathbf{Pr}\{\mathbf{S}^{(\eta)}=\boldsymbol{s}_n\}\mathbf{E}[D|\mathbf{S}^{(\eta)}=\boldsymbol{s}_n] .\end{aligned} \tag{16}$$

In order to calculate $\mathbf{E}[T^{(\eta)}]$, we will show how to evaluate $\mathbf{Pr}\{\mathbf{S}^{(\eta)}=\boldsymbol{s}_n\}$ and $\mathbf{E}[D|\mathbf{S}^{(\eta)}=\boldsymbol{s}_n]$ in the following.

A. Derivations of $\mathbf{Pr}\{\mathbf{S}^{(\eta)}=\boldsymbol{s}_n\}$ and $\mathbf{E}[D|\mathbf{S}^{(\eta)}=\boldsymbol{s}_n]$

In order to evaluate $\mathbf{Pr}\{\mathbf{S}^{(\eta)}=\boldsymbol{s}_n\}$ and $\mathbf{E}[D|\mathbf{S}^{(\eta)}=\boldsymbol{s}_n]$, a state diagram is developed by characterizing the evolutions of

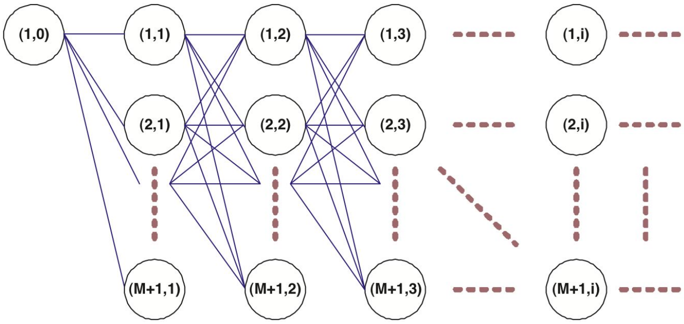


Fig. 4. State diagram of target channel sequence for a secondary connection, where default channel $\eta = 1$.

the target channel sequence and the corresponding cumulative handoff delay for the secondary connection. The proposed state diagram is a two-dimensional chain, where state (k, i) represents that channel k is selected for the target channel at the i^{th} interruption. Because the considered secondary connection's default channel is η, the initial state of the proposed state diagram is $(\eta, 0)$. Furthermore, state $(M+1, i)$ represents that the secondary user can finish its transmission after the $(i-1)^{th}$ interruption, and thus state $(M+1, i)$ is the ending of state transition. Note that the state transition occurs only at two adjacent states. Specifically, a transition link from (k, i) to (k', i') exists if $i' = i+1$, and vice versa. An example of state diagram is shown in Fig. 4, where $\eta = 1$.

In Fig. 4, the state transition path can be regarded as a target channel sequence. For example, for a target channel sequence $\boldsymbol{s}_n \triangleq (s_1, s_2, s_3, \cdots, s_n)$, the corresponding state transition path is $(\eta, 0) \to (s_1, 1) \to (s_2, 2) \to (s_3, 3) \to \cdots \to (s_n, n) \to (M+1, n+1)$. Hence, calculating the average cumulative handoff delay over all possible target channel sequences can be regarded as calculating the cumulative transition cost over all possible state transition paths. In the following, we show how to design the state transition probability and cost in the developed state diagram.

1) State Transition Probability: When an interruption event occurs, the interrupted secondary connection must search its target channel for spectrum handoff through spectrum sensing. The probability that each channel is selected to be the target channel is related to the statistics of channel occupancy. Let $P[(k', i)|(k, i-1)]$ be the transition probability from states $(k, i-1)$ to (k', i). At channel k, maybe the considered secondary connection does not encounter interruption again and can finish its transmission at channel k. In this case, the transition from states $(k, i-1)$ to $(M+1, i)$ will occur. On the other hand, the transition from states $(k, i-1)$ to (k', i) where $1 \le k' \le M$ will occur when the considered secondary connection is interrupted again at channel k. Thus, $P[(k', i)|(k, i-1)]$ can be expressed as follows:

$$P[(k', i)|(k, i-1)] = \begin{cases} 1 - p_{i-1,\eta}^{(k)} & , k' = M+1 \\ p_{i-1,\eta}^{(k)} \cdot \mathbf{Pr}[S_{i,\eta} = k' | S_{i-1,\eta} = k] & , k' \neq M+1 \end{cases} . \tag{17}$$

2) State Transition Cost: The cost of state transition is defined as the handoff delay of the interrupted secondary connection. The handoff delay from channels k to k' depends on the state of channel occupancy. Recall that δ_s and δ_c are the total processing time for executing spectrum handoff procedure when the secondary users stay on the current channel and change to another channel, respectively. If one idle channel exists after spectrum sensing, the interrupted secondary connection will change to this idle channel. Hence, the handoff delay in this case is δ_c. Furthermore, if all channels are busy, the interrupted secondary connection will stay on its current operating channel (i.e., $k = k'$). Hence, the expected handoff delay is the sum of δ_s and the duration from the time instant that channel k is used by the primary connections until the time instant that channel k becomes idle. This duration is called the *busy period* resulting from the transmissions of multiple primary connections at channel k and denoted by $Y_p^{(k)}$. Let $C[k'|k]$ be the transition cost from states $(k, i-1)$ to (k', i) in the state diagram. Then, we can have

$$C[k'|k] = \begin{cases} 0 & , \quad k' = M+1 \\ \delta_s + \mathbf{E}[Y_p^{(k)}] & , \quad k' = k \\ \delta_c & , \quad \text{others} \end{cases} . \tag{18}$$

Note that (18) is unrelated to the number of interruptions i. Furthermore, referring to [16], we can have

$$\mathbf{E}[Y_p^{(k)}] = \frac{\mathbf{E}[X_p^{(k)}]}{1 - \lambda_p^{(k)} \mathbf{E}[X_p^{(k)}]} . \tag{19}$$

Next, from the developed state diagram, $\mathbf{Pr}\{\mathbf{S}^{(\eta)} = \boldsymbol{s}_n\}$ and $\mathbf{E}[D|\mathbf{S}^{(\eta)} = \boldsymbol{s}_n]$ can be expressed as follows:

$$\begin{aligned} & \mathbf{Pr}\{\mathbf{S}^{(\eta)} = \boldsymbol{s}_n\} \\ = \ & P[(M+1, n+1)|(s_n, n)] \prod_{i=0}^{n-1} P[(s_{i+1}, i+1)|(s_i, i)] \ , \end{aligned} \tag{20}$$

and

$$\mathbf{E}[D|\mathbf{S}^{(\eta)} = \boldsymbol{s}_n] = \sum_{i=0}^{n} C[s_{i+1}|s_i] \ . \tag{21}$$

Note that (20) is equivalent to (12). Finally, substituting (20) and (21) into (16), we can obtain the closed-form expression for the extended data delivery time $\mathbf{E}[T^{(\eta)}]$ for any service time distribution $f_s^{(\eta)}(x)$.

B. An Example for the Exponentially Distributed Service Time

Now, we investigate how to derive the cumulative handoff delay in (16) when the secondary connection's service time is exponentially distributed as adopted in Section V-B. Intuitively, we can evaluate the value of $\mathbf{E}[D^{(\eta)}]$ by examining all possible transition paths in the state diagram, which has time complexity of $O(M^L)$. Fortunately, the derivations of the cumulative handoff delay can be simplified due to the memoryless property of the exponential distribution.

For the secondary connection whose default channel is channel k, its cumulative handoff delay $\mathbf{E}[D^{(k)}]$ in (16) is derived as follows. Because the considered secondary connection's service time distribution is the exponential distribution, its remaining service time after an interruption event occurs also follows the identical exponential distribution. Hence, for the secondary connections at state (k, i) and (k', i'), they

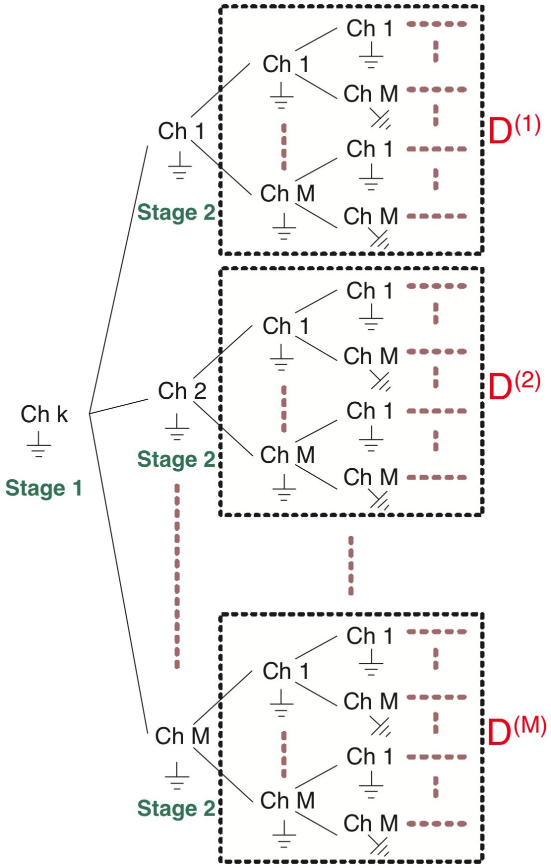


Fig. 5. Tree-structured representations of the proposed state diagram where the grounding symbols represent the ending of state transition. Note that this figure considers the secondary connections whose default channels are Chk.

will encounter the same cumulative handoff delay and interrupted probability in their remaining transmissions if $k = k'$, $k \neq M+1$, and $k' \neq M+1$. From the aforementioned discussions, we can re-plot the state diagram expression (in Fig. 4) as a tree-structured representation for the target channel selection. The tree-structured representation is shown in Fig. 5, where Chk represents that channel k is selected for the target channel and the "grounding symbols" represent the endings of state transition. Note that at the second stage of Fig. 5, the average cumulative handoff delay of the type-(1) secondary connection is also equal to $\mathbf{E}[D^{(k)}]$ when this type-(1) secondary connection's current operating channel is channel k. Furthermore, because the state transition probability is independent of the number of interruptions for the secondary connections due to memoryless property, we can have $P[(k', i+1)|(k, i)] = P[k'|k]$ for each $i \geq 0$. Hence, it follows that

$$\mathbf{E}[D^{(k)}] = P[M+1|k] \cdot C[M+1|k] + \sum_{k'=1}^{M} P[k'|k] \cdot (C[k'|k] + \mathbf{E}[D^{(k')}]) \quad (22)$$

where $1 \leq k \leq M$. Finally, substituting (17) and (18) into (22), we can obtain M independent equations. Hence, the closed-form expressions for the cumulative handoff delay $\mathbf{E}[D^{(k)}]$ can be derived by solving these simultaneous equations iteratively.

VII. Numerical Results

We show numerical results to reveal the importance of the three key design features for modeling spectrum handoffs as discussed in Section II, which consist of (1) various arrival rates of the secondary connections; (2) heterogeneous arrival rates of the primary connections; and (3) the handoff processing time.

A. Simulation Setting

In order to validate the proposed analytical model, we perform simulations based on the Monte-carlo method by MATLAB software. The continuous-time cognitive radio system, where the inter-arrival time and service time can be the duration of non-integer time slots, is simulated. We consider a two-channel CR system with Poisson arrival processes of rates $\lambda_p^{(k)}$ and $\lambda_s^{(k)}$ for the high-priority primary connections and the low-priority secondary connections, respectively. The high-priority connections can interrupt the transmissions of the low-priority connections. Furthermore, the connections with the same priority follow the first-come-first-served (FCFS) scheduling discipline.We adopt time slot duration of 10 msec in our simulations as discussed in the IEEE 802.22 standard [52].

B. Effects of Various Arrival Rates for the Secondary Connections

Firstly, we investigate the effects of various arrival rates of the secondary connections on the channel utilization and the extended data delivery time of the secondary connections. We consider a two-channel CR network, where $\lambda_p^{(1)} = \lambda_p^{(2)} \triangleq \lambda_p$, $(\lambda_s^{(1)}, \lambda_s^{(2)}) = (0.01, 0.02)$ (arrivals/slot), $(\mathbf{E}[X_p^{(1)}], \mathbf{E}[X_p^{(2)}]) = (20, 20)$ (slots/arrival), and $(\mathbf{E}[X_s^{(1)}], \mathbf{E}[X_s^{(2)}]) = (10, 10)$ (slots/arrival). We only consider the case that $0 \leq \lambda_p \leq 0.04$ (arrivals/slot) in the following numerical results. When $\lambda_p \geq 0.05$ (arrivals/slot), the overall normalized traffic workloads in the considered CR network will be saturated because $\lambda_p \mathbf{E}[X_p^{(1)}] + \lambda_p \mathbf{E}[X_p^{(2)}] + \lambda_s^{(1)} \mathbf{E}[X_s^{(1)}] + \lambda_s^{(1)} \mathbf{E}[X_s^{(1)}] > 2$.

Figure 6 shows the effects of the arrival rates of the primary connections (λ_p) on channel utilizations of channels 1 and 2. As λ_p increases, the channel utilizations of the two channels increase. Because channels 1 and 2 have the same arrival rate of the primary connections, the difference between channel utilizations of the both channels will keep a constant, which equals to 0.1.

Figure 7 shows the effects of the arrival rates of the primary connections (λ_p) on the extended data delivery time of the secondary connections whose initial default channels are channels 1 and 2. We have three important observations. First, because channels 1 and 2 have the same arrival rate of the primary connections, the secondary connections at the two channels will encounter same interrupted probability according to (14). Hence, the secondary connections whose default channels are channels 1 and 2 have similar extended data delivery time even though channels 1 and 2 have different channel utilization. Next, it is shown that the extended data delivery time of the secondary connections increases as λ_p increases because a

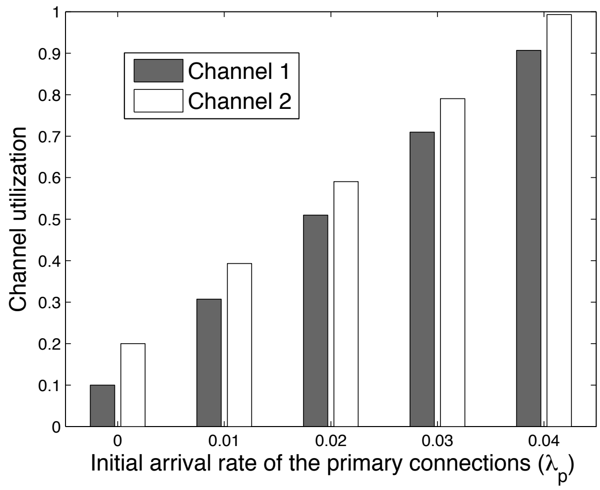


Fig. 6. Effects of the arrival rate of the primary connections (λ_p) on the channel utilizations at channels 1 and 2, where $\delta_s = 1$ and $\delta_c = 2$.

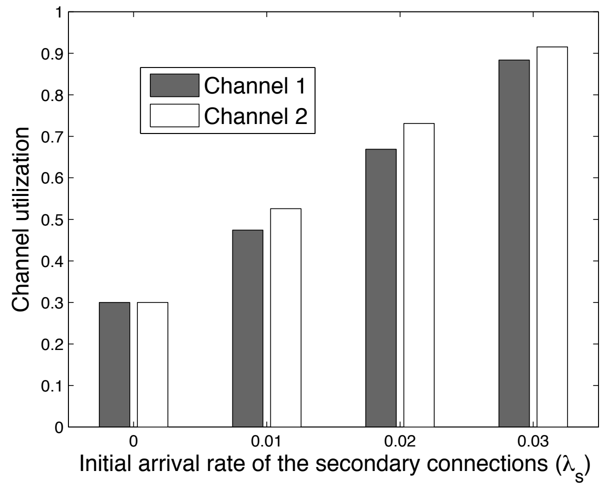


Fig. 8. Effects of the initial arrival rate of the secondary connections (λ_s) on the channel utilizations at channels 1 and 2.

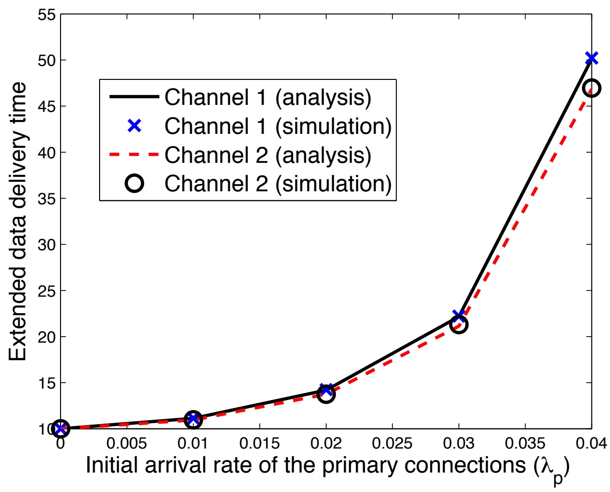


Fig. 7. Effects of the arrival rate of the primary connections (λ_p) on the average extended data delivery time of the secondary connections whose default channels are channels 1 and 2.

larger λ_p leads to higher channel busy probabilities and longer average handoff delay. More importantly, we find that the simulation results match the analytical results quite well. It can validate the slot-based assumption used in our analysis.

C. Effects of Heterogenous Arrival Rates for the Primary Users' Connections

Secondly, we demonstrate the effects of heterogeneous arrival rates of the primary connections on the channel utilization and the extended data delivery time of the secondary connections. We also consider a two-channel network, where $\lambda_s^{(1)} = \lambda_s^{(2)} \triangleq \lambda_s$, $(\lambda_p^{(1)}, \lambda_p^{(2)}) = (0.03, 0.01)$ (arrivals/slot), $(\mathbf{E}[X_s^{(1)}], \mathbf{E}[X_s^{(2)}]) = (20, 20)$ (slots/arrival), and $(\mathbf{E}[X_p^{(1)}], \mathbf{E}[X_p^{(2)}]) = (10, 30)$ (slots/arrival). Note that the two channels have the same channel utilizations resulting from the primary connections. Specifically, we have $\rho_p^{(1)} = \rho_p^{(2)} = 0.3$. We only consider the case that $0 \leq \lambda_s \leq 0.03$ (arrivals/slot) in the following numerical results. When $\lambda_s \geq 0.04$ (arrivals/slot), the overall normalized traffic workloads in the considered CR network will be saturated because $\lambda_p^{(1)}\mathbf{E}[X_p^{(1)}] + \lambda_p^{(1)}\mathbf{E}[X_p^{(2)}] + \lambda_s\mathbf{E}[X_s^{(1)}] + \lambda_s\mathbf{E}[X_s^{(1)}] > 2$.

The effects of the initial arrival rates of the secondary connections (λ_s) on the channel utilizations of channels 1 and 2 is shown in Fig. 8. When $\lambda_s = 0$, the two channels have the same channel utilizations of 0.3. As λ_s increases, the channel utilizations of the two channels increase. However, the increments of the two channels are different even though the two channels have the same busy probability resulting from the primary connections. Firstly, compared to the secondary connections at channel 1, the secondary connections at channel 2 will encounter lower interrupted probability because channel 2 has smaller arrival rate of the primary connections. Thus, the time that the secondary connections can use channel 2 is longer than the time that the secondary connections can use channel 1. Hence, the increment of channel utilization at channel 2 is larger than that at channel 1 when λ_s is increased from 0.01 to 0.02. Furthermore, when interruption event occurs, the secondary connection at channel 1 will stay on the current operating channel with higher probability compared to the secondary connection at channel 2 because channel 2 has higher busy probability. Hence, the increment of channel utilization at channel 1 is larger than that at channel 1 when λ_s is increased from 0.02 to 0.03.

Figure 9 shows the effects of the initial arrival rates of the secondary connections (λ_s) on the extended data delivery time of the secondary connections whose initial channels are channels 1 and 2. We can find that the extended data delivery time of the secondary connections increases as λ_s increases because a larger λ_s will lead to a higher channel busy probability as shown in Fig. 8. Furthermore, the secondary connections whose initial channel is channel 2 has shorter extended data delivery time compared to the secondary connections whose initial channel is channel 1. This is because the secondary connections whose initial channel is channel 2 can have lower interrupted probability and the smaller number of interruptions during their transmission period.

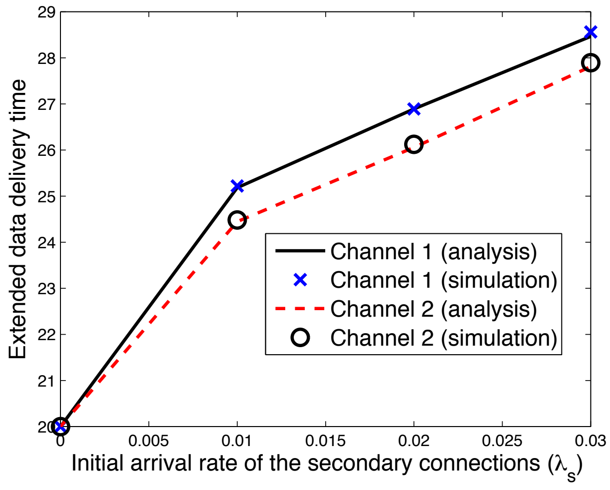


Fig. 9. Effects of the initial arrival rate of the secondary connections (λ_s) on the average extended data delivery time of the secondary connections whose default channels are channels 1 and 2, where $\delta_s = 1$ and $\delta_c = 2$.

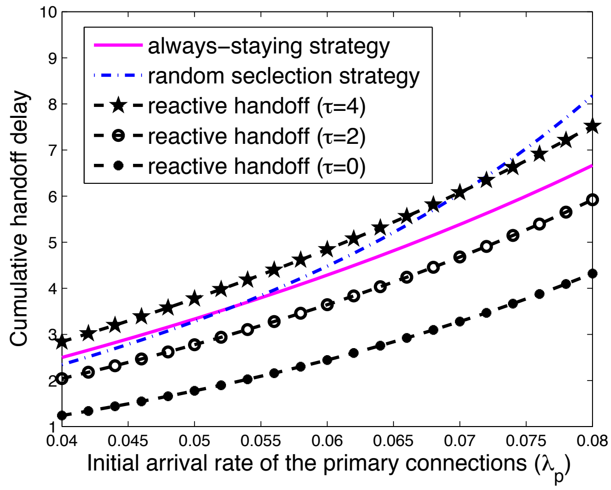


Fig. 10. Comparison of average cumulative handoff delay for different target channel selection schemes.

D. Effects of Handoff Processing Time

Finally, we discuss the effects of handoff processing time. A two-channel CR network is considered with the following parameters: $t_h = 0$ (slot), $t_s = 1$ (slot), $\lambda_p^{(1)} = \lambda_p^{(2)} \triangleq \lambda_p$, $\lambda_s^{(1)} = \lambda_s^{(2)} = 0.02$ (arrivals/slot), $\mathbf{E}[X_p^{(1)}] = \mathbf{E}[X_p^{(2)}] = 5$ (slots/arrival) and $\mathbf{E}[X_s^{(1)}] = \mathbf{E}[X_s^{(2)}] = 10$ (slots/arrival). Then, based on the proposed analytical model, we can evaluate the average cumulative handoff delay and then design the admission control rule for the secondary users as shown in Figs. 10 and 11.

Figure 10 compares the cumulative handoff delay $\mathbf{E}[D]$ of the following three target channel selection schemes: (1) the always-staying strategy; (2) the random selection strategy; and (3) the reactive-decision selection strategy. For the always-staying approach, the interrupted secondary user always stays on its default channel to resume its unfinished data transmission. The approach is one kind of proactive-decision spectrum handoff [16] because the target channels are predetermined and its concept is similar to the non-hopping mode of IEEE 802.22 [53]. In the random selection approach, the interrupted user randomly selects a target channel from all channels. From this figure, we have the following three important observations. First, we find that the cumulative handoff delay resulting from the random selection method is longer than that resulting from the always-staying strategy when $\lambda_p \geq 0.052$. For a larger value of λ_p, the interrupted secondary users with the random selection method must spend much more time to wait when it changes its operating channel because the selected target channel is likely busy. Thus, the handoff delay of the random selection method becomes longer in this case. Next, it is shown that the reactive-decision spectrum handoff can result in the shortest cumulative handoff delay when $\tau = 0$ (slot) because it can reliably find idle channels by performing spectrum sensing. In this case, the cumulative handoff delay can be shortened around 40% compared to the other approaches under various arrival rates of the primary connections. However, when the spectrum sensing time becomes longer (e.g., $\tau = 4$ slots), the reactive-decision spectrum handoff is worse than the random selection method in terms of cumulative handoff delay when $\lambda_p \leq 0.071$. This is because handoff delay increases as τ increases. Finally, we find that the sensing technology can effectively shorten the average cumulative handoff delay only when $\tau \leq 2$ compared to the always-staying strategy.

The analytical results developed in this paper can be used to design the admission control rule for the arriving secondary users subject to their latency requirement. Fig. 11 shows the admissible region for the normalized traffic workloads (or channel utilities) (ρ_p, ρ_s) for the Voice over IP (VoIP) services when $\tau = 0$ (slot), where $\rho_p = \lambda_p\mathbf{E}[X_p]$ and $\rho_s = \lambda_s\mathbf{E}[X_s]$. The maximum allowable average cumulative delay resulting from multiple handoffs is 20 ms for the VoIP traffic [54]. According to this figure, the admission control policy can be designed. When $\rho_p < 0.1667$, a CR network can accept all arrival requests from the secondary users until the CR network is saturated, i.e., $\rho_p + \rho_s \simeq 1$. Furthermore, when $0.1667 < \rho_p < 0.3397$, a part of traffic workloads of the secondary users must be rejected in order to satisfy the delay constraint of the secondary users. For example, when $\rho_p = 0.25$, a CR network can support at most 0.279 workload for the secondary users. That is, a CR network can accept the traffic of the secondary connections with rate $\lambda_s = 0.0279$ (arrivals/slot) at most when $\lambda_p = 0.05$ (arrivals/slot). In order to design the most allowable λ_s to achieve this arrival rate upper bound for the secondary connections, many arrival-rate control methods can be considered, such as the p-persistent carrier sense multiple access (CSMA) protocol in [55] and the call admission control mechanisms in [30], [56], [57]. Finally, when $\rho_p > 0.3397$, no secondary user can be accepted. Note that the size of the admissible region decreases as τ increases.

VIII. Conclusions

In this paper, we have investigated the effects of reactive-decision spectrum handoff on the channel utilization and the extended data delivery time of the secondary connections by considering the three key design features for spectrum

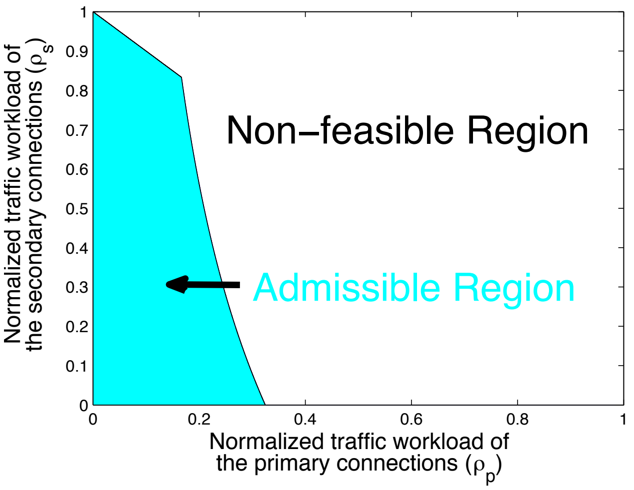


Fig. 11. Admissible region (λ_p, λ_s), where the average extended data delivery time constraint can be satisfied when $\tau = 0$.

handoffs, consisting of (1) heterogeneous arrival rates of the primary users; (2) various arrival rates of the secondary users; (3) handoff processing time. Firstly, we proposed a PRP M/G/1 queueing network model to characterize the spectrum usage behaviors between the primary and the secondary connections with multiple handoffs. Next, we developed a state diagram to characterize the effect of multiple handoff delay on the extended data delivery time of the secondary connections. Based on the proposed unifying model, an insightful study to quantify the effect of the three design features on the channel utilizations and the extended data delivery time under various traffic arrival rates and service time distributions can be provided. More importantly, these analytical results can facilitate the designs of admission control rule for the secondary users and can provide a framework to determine whether the spectrum sensing technology can effectively shorten the data delivery time under various sensing time.

Some interesting research issues can be extended from this paper. Firstly, one can relax the assumption that the interrupted secondary connection must stay on the current operating channel if channels are all busy, and investigate the handoff delay resulting from various target channel selection approaches. For example, when channels are all busy, the interrupted secondary user can perform sensing again after a predetermined period. Another interesting scenario is when the interrupted secondary user may need to retransmit the whole connection rather than resuming the unfinished transmission. In this case, how to use preemptive *repeat* priority queueing network to model CR networks is still an open research issue.

References

[1] J. Mitola and G. Q. Maguire, "Cognitive Radio: Making Software Radios More Personal," *IEEE Personal Commun.*, vol. 6, pp. 13–18, Aug. 1999.

[2] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] R. W. Thomas, L. A. DaSilva, and A. B. MacKenzie, "Cognitive Networks," *IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, Nov. 2005.

[4] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey," *Computer Networks Journal (Elsevier)*, vol. 50, pp. 2127–2159, Sep. 2006.

[5] C.-W. Wang, L.-C. Wang, and F. Adachi, "Performance Gains for Spectrum Utilization in Cognitive Radio Networks with Spectrum Handoff," *International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Sep. 2009.

[6] L.-C. Wang, C.-W. Wang, Y.-C. Lu, and C.-M. Liu, "A Concurrent Transmission MAC Protocol for Enhancing Throughput and Avoiding Spectrum Sensing in Cognitive Radio," *IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2007.

[7] T. A. Weiss and F. K. Jondral, "Spectrum Pooling: An Innovative Dtrategy for the Enhancement of Spectrum Efficiency," *IEEE Radio Commun. Mag.*, vol. 42, no. 3, pp. S8–S14, Mar. 2004.

[8] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A Survey on Spectrum Management in Cognitive Radio Networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr. 2008.

[9] M. D. Silvius, R. Rangnekar, A. B. MacKenzie, and C. W. Bostian, "The Smart Radio Channel Change Protocol: A Primary User Avoidance Technique for Dynamic Spectrum Sharing Cognitive Radios to Facilitate Co-existence in Wireless Communication Networks," *International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, Jun. 2009.

[10] L.-C. Wang, C.-W. Wang, and K.-T. Feng, "A Queueing-Theoretical Framework for QoS-Enhanced Spectrum Management in Cognitive Radio Networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 6, pp. 18–26, Dec. 2011.

[11] S. Srinivasa and S. A. Jafar, "The Throughput Potential of Cognitive Radio: A Theoretical Perspective," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 73–79, May 2007.

[12] Q. Shi, D. Taubenheim, S. Kyperountas, P. Gorday, and N. Correal, "Link Maintenance Protocol for Cognitive Radio System with OFDM PHY," *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Apr. 2007.

[13] H. Su and X. Zhang, "Channel-hopping Based Single Transceiver MAC for Cognitive Radio Networks," *IEEE Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2008.

[14] D. Willkomm, J. Gross, and A. Wolisz, "Reliable Link Maintenance in Cognitive Radio Systems," *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Nov. 2005.

[15] J. Tian and G. Bi, "A New Link Maintenance and Compensation Model for Cognitive UWB Radio Systems," *International Conference on ITS Telecommunications Proceedings*, Jun. 2006.

[16] L.-C. Wang, C.-W. Wang, and C.-J. Chang, "Modeling and Analysis for Spectrum Handoffs in Cognitive Radio Networks," *IEEE Trans. Mobile Computing*, vol. 11. no. 9, pp. 1499–1513, Sept. 2011.

[17] ——, "Optimal Target Channel Sequence Design for Multiple Spectrum Handoffs in Cognitive Radio Networks," *IEEE Trans. Commun.*, Sept. 2012.

[18] L.-C. Wang, C.-W. Wang, and F. Adachi, "Load-Balancing Spectrum Decision for Cognitive Radio Networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 757–769, Apr. 2011.

[19] C.-T. Chou, Sai Shankar N, H. Kim, and K. G. Shin, "What and How Much to Gain from Spectral Agility?" *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 576–588, Apr. 2007.

[20] J. Heo, J. Shin, J. Nam, Y. Lee, J. G. Park, and H.-S. Cho, "Mathematical Analysis of Secondary User Traffic in Cognitive Radio System," *IEEE Vehicular Technology Conference Fall*, Sep. 2008.

[21] L.-C. Wang and A. Chen, "On the Performance of Spectrum Handoff for Link Maintenance in Cognitive Radio," *IEEE International Symposium on Wireless Pervasive Computing (ISWPC)*, May 2008.

[22] Y. Zhang, "Spectrum Handoff in Cognitive Radio Networks: Opportunistic and Negotiated Situations," *IEEE International Conference on Communications (ICC)*, Jun. 2009.

[23] F. Capar, I. Martoyo, T. Weiss, and F. Jondral, "Comparison of Bandwidth Utilization for Controlled and Uncontrolled Channel Assignment in a Spectrum Pooling System," *IEEE Vehicular Technology Conference Spring*, May 2002.

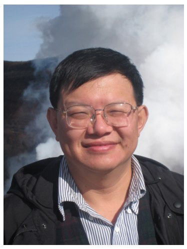
[24] B. Ishibashi, N. Bouabdallah, and R. Boutaba, "QoS Performance Analysis of Cognitive Radio-based Virtual Wireless Networks," *IEEE International Conference on Computer Communications (INFOCOM)*, Apr. 2008.

[25] X. Zhu, L. Shen, and T.-S. P. Yum, "Analysis of Cognitive Radio Spectrum Access with Optimal Channel Reservation," *IEEE Commun. Lett.*, vol. 11, no. 4, pp. 304–306, Apr. 2007.

[26] J. Martinez-Bauset, V. Pla, and D. Pacheco-Paramo, "Comments on

Analysis of Cognitive Radio Spectrum Access with Optimal Channel Reservation," *IEEE Commun. Lett.*, vol. 13, no. 10, p. 739, Oct. 2009.
[27] W. Ahmed, J. Gao, H. A. Suraweera, and M. Faulkner, "Comments on Analysis of Cognitive Radio Spectrum Access with Optimal Channel Reservation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4488–4491, Sep. 2009.
[28] D. Pacheco-Paramo, V. Pla, and J. Martinez-Bauset, "Optimal Admission Control in Cognitive Radio Networks," *International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, Jun. 2009.
[29] W. Ahmed, J. Gao, and M. Faulkner, "Performance Evaluation of a Cognitive Radio Network with Exponential and Truncated Usage Models," *IEEE International Symposium on Wireless Pervasive Computing (ISWPC)*, Oct. 2009.
[30] M. Huang, R. Yu, and Y. Zhang, "Call Admission Control with Soft-QoS Based Spectrum Handoff in Cognitive Radio Networks," *International Conference on Wireless Communications and Mobile Computing (IWCMC)*, Jun. 2009.
[31] I. Suliman, J. Lehtomaki, T. Braysy, and K. Umebayashi, "Analysis of Cognitive Radio Networks with Imperfect Sensing," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2009.
[32] Y. Y. Mihov and B. P. Tsankov, "Call-Level Performance Evaluation and QoS Provisioning in Cognitive Radio Networks," *IEEE Africon*, Sep. 2011.
[33] S. Tang and B. L. Mark, "Performance Analysis of a Wireless Network with Opportunistic Spectrum Sharing," *IEEE Global Communications Conference (GLOBECOM)*, Nov. 2007.
[34] ——, "Modeling and Analysis of Opportunistic Spectrum Sharing with Unreliable Spectrum Sensing," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1934–1943, Apr. 2009.
[35] ——, "Modeling an Opportunistic Spectrum Sharing System with a Correlated Arrival Process," *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2008.
[36] ——, "An Analytical Performance Model of Opportunistic Spectrum Access in a Military Environment," *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2008.
[37] E. W. M. Wong and C. H. Foh, "Analysis of Cognitive Radio Spectrum Access with Finite User Population," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 294–296, May 2009.
[38] M. M. Rashid, M. J. Hossain, E. Hossain, and V. K. Bhargava, "Opportunistic Spectrum Access in Cognitive Radio Networks: A Queueing Analytic Model and Admission Controller Design," *IEEE Global Communications Conference (GLOBECOM)*, Nov. 2007.
[39] ——, "Opportunistic Spectrum Scheduling for Multiuser Cognitive Radio: A Queueing Analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5259–5269, Oct. 2009.
[40] Y. Zhang, "Dynamic Spectrum Access in Cognitive Radio Wireless Networks," *IEEE International Conference on Communications (ICC)*, May 2008.
[41] C. Wu, C. He, L. Jiang, and Y. Chen, "A Novel Spectrum Handoff Scheme With Spectrum Admission Control in Cognitive Radio Networks," *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2011.
[42] Sai Shankar N, "Squeezing the Most Out of Cognitive Radio: A Joint MAC/PHY Perspective," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, Apr. 2007.
[43] Sai Shankar N, C.-T. Chou, K. Challapali, and S. Mangold, "Spectrum Agile Radio: Capacity and QoS Implications of Dynamic Spectrum Assignment," *IEEE Global Communications Conference (GLOBECOM)*, Nov. 2005.
[44] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, April 2007.
[45] Y.-C. Liang, Y. Zeng, E. C. Peh, and A. T. Hoang, "Sensing-Throughput Tradeoff for Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, Apr. 2008.
[46] P. Wang, L. Xiao, S. Zhou, and J. Wang, "Optimization of Detection Time for Channel Efficiency in Cognitive Radio Systems," *IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2007.
[47] W.-Y. Lee and I. F. Akyildiz, "Optimal Spectrum Sensing Framework for Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3845–3857, Oct. 2008.
[48] H.-P. Shiang and M. van der Schaar, "Queuing-Based Dynamic Channel Selection for Heterogeneous Multimedia Applications Over Cognitive Radio Networks," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 896–909, Aug. 2008.
[49] L.-C. Wang, Y.-C. Lu, C.-W. Wang, and D. S.-L. Wei, "Latency Analysis for Dynamic Spectrum Access in Cognitive Radio: Dedicated or Embedded Control Channel?" *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2007.
[50] X. Li and S. A. Zekavat, "Traffic Pattern Prediction and Performance Investigation for Cognitive Radio Systems," *IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2008.
[51] Chee-Hock Ng and Boon-Hee Soong, *Queueing Modelling Fundamentals with Applications in Communication Networks, 2nd.* John Wiley & Sons Inc., 2008.
[52] *Draft Standard for Wireless Regional Area Networks Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE 802.22 Working Group.
[53] W. Hu, D. Willkomm, G. Vlantis, M. Gerla, and A. Wolisz, "Dynamic Frequency Hopping Communities for Efficient IEEE 802.22 Operation," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 80–87, May 2007.
[54] C. R. Stevenson, C. Cordeiro, E. Sofer, and G. Chouinard, "Functional Requirements for The 802.22 WRAN Standard," *IEEE 802.22-05/0007r46*, Sep. 2005.
[55] A. Banaei and C. N. Georghiades, "Throughput Analysis of A Randomized Sensing Scheme in Cell-Based Ad-Hoc Cognitive Networks," *IEEE International Conference on Communications (ICC)*, Jun. 2009.
[56] B. Wang, Z. Ji, K. J. Ray Liu, and T. C. Clancy, "Primary-Prioritized Markov Approach for Dynamic Spectrum Allocation," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1854–1865, Apr. 2009.
[57] C.-W. Wang, L.-C. Wang, and F. Adachi, "Optimal Admission Control in Cognitive Radio Networks with Sensing Errors," *IEICE Tech. Rep.*, vol. 109, no. 440, pp. 491–496, Mar. 2010.

Chung-Wei Wang (S'07) received the B.S. degree in electrical engineering from Tamkang University, Taipei, Taiwan, in 2003, and the Minor M.S. and Ph.D. degrees in applied mathematics and communication engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 2007 and 2010, respectively. From 2009 to 2010, he was also a visiting scholar in Tohoku University, Sendai, Japan. He was awarded the student travel grants from IEEE ICC 2009 and GLOBECOM 2010. He is currently working as a Principal Engineer at MStar Semiconductor, Inc., Taipei, Taiwan, and is responsible for the development of 3GPP technologies. His current research interests include cross-layer optimization, MAC protocols design, and radio resource management in cellular system, wireless sensor networks, ad hoc networks, and cognitive radio networks.



Li-Chun Wang (M'96 – SM'06 – F'11) received the B.S. degree from National Chiao Tung University , Taiwan, R. O. C. in 1986, the M.S. degree from National Taiwan University in 1988, and the Ms. Sci. and Ph. D. degrees from the Georgia Institute of Technology , Atlanta, in 1995, and 1996, respectively, all in electrical engineering. From 1996 to 2000, he was with AT&T Laboratories, where he was a Senior Technical Staff Member in the Wireless Communications Research Department. Since August 2000, he has been an Associate Professor in the Department of Communication Engineering of National Chiao Tung University in Taiwan.

Dr. Wang was a co-recipient (with Gordon L. Stüber and Chin-Tau Lea) of the 1997 IEEE Jack Neubauer Best Paper Award. He has published over 150 journal and international conference papers and is holding three US patents. He was elected to the IEEE fellow grade in 2011 for his contributions in cellular architectures and radio resource management in wireless networks. He served as an Associate Editor for the IEEE Trans. on Wireless Communications from 2001 to 2005, the Guest Editor of Special Issue on "Mobile Computing an Networking" for IEEE Journal on Selected Areas in Communications in 2005 and on "Radio Resource Management and Protocol Engineering in Future IEEE Broadband Networks" for IEEE Wireless Communications Magazine in 2006. He is holding eight US patents.