Modeling and Analysis for Spectrum Handoffs in Cognitive Radio Networks

Li-Chun Wang, *Fellow, IEEE,* Chung-Wei Wang, *Student Member, IEEE,* and Chung-Ju Chang, *Fellow, IEEE*

Abstract

In this paper, we present an analytical framework to evaluate the latency performance of connection-based spectrum handoffs in cognitive radio (CR) networks. During the transmission period of a secondary connection, multiple interruptions from the primary users result in multiple spectrum handoffs and the need of predetermining a set of target channels for spectrum handoffs. To quantify the effects of channel obsolete issue on the target channel predetermination, we should consider the three key design features: (1) general service time distribution of the primary and secondary connections; (2) different operating channels in multiple handoffs; and (3) queueing delay due to channel contention from multiple interrupted secondary connections. To this end, we propose the preemptive resume priority (PRP) M/G/1 queueing network model to characterize the spectrum usage behaviors with all the three design features. This model aims to analyze the extended data delivery time of the secondary connections with proactively designed target channel sequences under various traffic arrival rates and service time distributions. These analytical results are applied to evaluate the latency performance of the connection-based spectrum handoff based on the target channel sequences mentioned in the IEEE 802.22 wireless regional area networks standard. Then, to reduce the extended data delivery time, a traffic-adaptive spectrum handoff is proposed, which changes the target channel sequence of spectrum handoffs based on traffic conditions. Compared to the existing target channel selection methods, this traffic-adaptive target channel selection approach can reduce the extended data transmission time by 35%, especially for the heavy traffic loads of the primary users.

Index Terms

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

This work was presented in part at the IEEE International Conference on Communications (ICC), 2009. This work was supported in part by the MoE ATU Plan, by the National Science Council (NSC) grants 96-2628-E-009-004-MY3, 97-2221-E-009-099-MY3, and 97-2917-I-009-109.
Li-Chun Wang, Chung-Wei Wang, and Chung-Ju Chang are with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu 300, Taiwan. (E-mail: lichun@cc.nctu.edu.tw, hyper.cm91g@nctu.edu.tw, and cjchang@mail.nctu.edu.tw)

Digital Object Indentifier 10.1109/TMC.2011.155 1536-1233/11/$26.00 © 2011 IEEE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

1 Introduction

Cognitive radio (CR) can significantly improve spectrum efficiency by allowing the secondary users to temporarily access the primary user's under-utilized licensed spectrum [1]–[4]. Spectrum mobility issues arise when the primary user appears at the channels being occupied by the secondary users. The secondary users need to return the occupied channel because the primary users have the preemptive priority to access channels. Spectrum handoff techniques can help the interrupted secondary user vacate the occupied licensed channel and find a suitable target channel to resume its unfinished data transmission [5], [6].

One fundamental issue for spectrum handoff modeling in CR networks is the multiple interruptions from the primary users during each secondary user's connection [7]. The issue of multiple interruptions results in the requirement of designing the target channel pool for a series of spectrum handoffs in a secondary connection. In this paper, we define the *connection-based modeling techniques* for spectrum handoff as the schemes that incorporate the effects of multiple interruptions from the primary users in an event-driven manner, and the *slot-based modeling techniques* mean that the interruptions to the secondary user are modeled in a time-driven manner. That is, the connection-based method models the spectrum handoff only when the primary user appears, while for the slot-based methods the spectrum handoff is considered at each time slot.

Spectrum handoff mechanisms can be generally categorized into two kinds according to the decision timing of selecting target channels [8]. The first kind is called the proactive-decision spectrum handoff[1], which decides the target channels for future spectrum handoffs based on the long-term traffic statistics before data connection is established [22]–[24]. The second kind is called the reactive-decision spectrum handoff scheme [25]. For this scheme, the target channel is searched in an on-demand manner [26], [27]. After a spectrum handoff is requested, spectrum sensing is performed to help the secondary users find idle channels to resume their unfinished data transmission. Both spectrum handoff schemes have their own advantages and disadvantages. A quantitative comparison of the two spectrum handoff schemes was provided in [8].

In this paper, we focus on the modeling technique and performance analysis for the proactive-decision spectrum handoff scheme, while leave the related studies on the reactive-decision spectrum handoff in [25]. Compared to the reactive spectrum handoff scheme, the proactive

1. In this paper, we assume that spectrum handoff request is initiated only when the primary user appears as discussed in the IEEE 802.22 wireless regional area networks (WRAN) standard. In this scheme, the proactive spectrum handoff represents the spectrum handoff scheme with the proactively designed target channel sequences. It is different from the proactive spectrum handoff in [9]–[21] that assumes spectrum handoff can be performed before the appearance of the primary users.

spectrum handoff is easier to achieve a consensus on the target channels between the transmitter and its intended receiver because both the transmitter and receiver can know their target channel sequence for future spectrum handoffs before data transmission. Furthermore, the change switching delay of the proactive spectrum handoff is shorter than that of the reactive spectrum handoff because scanning wide spectrum to determine the target channel is not necessary at the moment of link transition. Nevertheless, the proactive spectrum handoff scheme shall resolve the obsolescent channel issue because the predetermined target channel may not be available any more when a spectrum handoff is requested.

To characterize the channel obsolescence effects and the spectrum usage behaviors with a series of interruptions in the secondary connections, we suggest a new performance metric - the extended data delivery time of the secondary connections. It is defined as the duration from the instant of starting transmitting data until the instant of finishing the whole connection, during which multiple interruptions from the primary users may occur. In the context of the connection-based spectrum handoffs, how to analyze the extended data delivery time is challenging because three key design features must be taken into account: (1) general service time distribution, where the probability density functions (pdfs) of service time of the primary and secondary connections can be any distributions; (2) different operating channels before and after spectrum handoff; and (3) queueing delay due to channel contention from multiple secondary connections. To the best of our knowledge, an analytical model for characterizing all these three features for multiple handoffs has rarely been seen in the literature.

In this paper, we propose a preemptive resume priority (PRP) M/G/1 queueing network model to characterize the spectrum usage behaviors of the connection-based multiple-channel spectrum handoffs. Based on the proposed model, we derive the closed-form expression for the extended data delivery time of different proactively designed target channel sequences under various traffic arrival rates and service time distributions. We apply the developed analytical method to analyze the latency performance of spectrum handoffs based on the target channel sequences specified in the IEEE 802.22 wireless regional area networks (WRAN) standard. We also suggest a traffic-adaptive target channel selection principle for spectrum handoffs under different traffic conditions.

The rest of this paper is organized as follows. Section 2 reviews the current spectrum usage models for the proactive spectrum handoff schemes in the literature. An illustrative example for multiple handoff issue is given in Section 3. Next, we present the PRP M/G/1 queueing network model, which can characterize the spectrum usage behaviors with multiple handoffs, in Section 4. Based on this model, Section 5 evaluates the extended data delivery time of the secondary connections with various target channel sequences. Then, Section 6 investigates the

TABLE 1

Comparison of Various Channel Usage Models.

Model Name	Type of Modeling Technique	General Service Time	Various Operating Channels	Multiple Secondary Connections
Two-State Markov Chain [9]–[12]	Slot-based	×	○	×
Arbitrary ON/OFF Random Process [13]–[21]	Slot-based	○	○	×
Bernoulli Random Process [28]	Connection-based	×	○	×
Multi-dimensional Markov Chain [29]	Connection-based	×	×	○
M/G/1 Queueing Model [30]–[38]	Connection-based	○	×	○
Proposed M/G/1 Queueing Network Model	Connection-based	○	○	○

latency performance of the spectrum handoffs resulting from the two typical target channel sequences mentioned in the IEEE 802.22 WRAN standard. Analytical and simulation results are given in Section 7. Finally, we give our concluding remarks in Section 8.

2 Related Work

In order to characterize the multiple handoff behaviors in CR networks, we should consider the three key design features, consisting of (1) general service time distribution; (2) various operating channels; and (3) queueing delay due to channel contention from multiple secondary connections. Based on these three features, Table 1 classifies the existing modeling techniques for the proactive spectrum handoff. In the table, the signs “ ○ ” and “ × ” indicate that the proposed model “does” and “does not” consider the corresponding feature, respectively. In the literature, the modeling techniques for spectrum handoff behaviors can be categorized into the following five types: (1) the two-state Markov chain; (2) the arbitrary ON/OFF random process; (3) the Bernoulli random process; (4) the birth-death process with multi-dimensional Markov chain; and (5) the PRP 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.

- **Discrete-time Two-state Markov chain**: In [9]–[12], the evolutions of the channel usage of the primary networks at each channel were modeled as Gilbert-Elliot channel, i.e., a discrete-time Markov chain which has two occupancy states: busy (ON) and idle (OFF) states. The idle state can be regarded as a potential spectrum opportunity for the secondary users. Note that the Markov chain model is suitable for the exponentially distributed service time, and how to extend it to the case with general service time distribution is not clear. In this model, the target channel selection problem in every time slot is modeled as a Markov decision process. According to the channel occupancy state at the current time slot, a decision maker (secondary user) can preselect the best action (target channel) to optimize its immediate reward (such as expected per-slot throughput [9]–[11], or expected waiting time [12]) at the next time slot. Note that this model belongs to the slot-based modeling technique because the secondary user shall decide its target channel at each time slot. Even though the primary users do not appear at the current operating channel, the secondary user still needs to change its operating channel, resulting in frequent spectrum handoffs.

- **Arbitrary ON/OFF random process**: Unlike the authors in [9]–[12] which assumed that the channel usage behaviors of the primary networks have the Markov property, the authors in [13]–[21] used the continuous-time ON/OFF random process with arbitrary distributed ON/OFF period to characterize the channel usage behaviors of the primary networks at each channel. It was assumed that the secondary user can estimate the distributions of the ON period and the OFF period based on long-term observations. In each time slot, the secondary user must calculate the expected reward such as the average remaining idle periods of primary users [13]–[19] or the average throughout of secondary users [20], [21]. Then, the secondary users will immediately switch to the channel with the largest reward. This model also belongs to the slot-based modeling technique because the target channel is decided in each time slot.
- **Bernoulli random process**: The authors in [28] examined the effects of multiple interruptions from the primary users on the connection maintenance probability in a connection-based environment, where the spectrum usage behaviors of the primary networks on each channel were characterized by a Bernoulli random process. Here, the connection maintenance probability is the probability that a secondary connection can finish its transmission within a predetermined number of handoff trials. Because both the busy and idle periods of the considered primary networks follow the geometrical distributions, it is more difficult to extend this modeling technique to the cases with other general service time distribution.
- **Multi-dimensional Markov chain**: In [29], the spectrum usage behaviors of both the primary and secondary networks were modeled by the multi-dimensional Markov chain. The actions of each primary and secondary user are indicated in the states of the Markov chain. Here, the action of each user can be ”idleness”, ”waiting at queue”, or ”communication”. It was assumed that the secondary user must stay on its current operating channel after the primary user’s interruption. This analytical model is suitable for the single channel network, and the issue of different operating channels in multiple handoffs has not been addressed.

- **M/G/1 queueing model**: Some researchers used the preemptive resume priority (PRP) M/G/1 queueing model to characterize the spectrum usage behaviors in a single-channel CR network. The effects of multi-user sharing and multiple interruptions on the extended data delivery time of the secondary users were studied in [30]–[38]. Note that the authors in [30]–[38] also assumed that the secondary users must stay on the current operating channel to resume their unfinished transmissions when they are interrupted.

To summarize, the first three analytical models, two-state Markov chain, arbitrary ON/OFF random process, and bernoulli random process, did not incorporate the effects of the traffic loads of the secondary users on the statistics of channel occupancy. How to extend these models to consider the queueing delay due to channel contention from multiple secondary connections is unclear. The last two models, multi-dimensional Markov chain and M/G/1 queueing model, can characterize the effects of spectrum sharing between multiple secondary users. However, these two models assumed that the interrupted secondary user must stay on the current operating channel, and have not dealt with the handoff interaction issue among different channels.

In this paper, we propose a PRP M/G/1 queueing *network* model to take into account of all the effects of the general service time distributions of the primary and the secondary connections, various operating channels, and the queueing behaviors of multiple secondary connections. In the next sections, we will discuss the analytical framework of proactive spectrum handoff based on the PRP M/G/1 queueing *network* model.

3 System Model

3.1 Assumptions

In this paper, we consider a CR network with M independent channels, where each channel has its own high-priority and low-priority queues as discussed in [38]. The traffic loads of the primary and secondary users respectively enter the high-priority and low-priority queues before transmitting data. Then, according to the time that traffic arrival at queues, the *primary connections* and the *secondary connections* are established for the primary and the secondary users[2], respectively. Here, we assume that the connections with the same priority follow the first-come-first-served (FCFS) scheduling ploicy in a centralized manner regardless of uplink or downlink[3].

2. Note that we assume the primary and secondary users always have packets to send during the connections, and the considered two queues have an infinite length for simplification.

3. Note that this model can be also applied to the decentralized CR network architectures. In this case, the channel contention time and retransmission in the medium access control (MAC) layer should be taken into account when calculating the latency performance of the secondary connections [8].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Assume that the considered CR network is a time-slotted system as [9], [24], [39]–[41]. In order to detect the presence of primary connections, each secondary user must perform spectrum sensing at the beginning of each time slot. If the current operating channel is idle, the secondary user can transmit or receive data in the remaining duration of this time slot. Otherwise, the secondary user must perform spectrum handoff procedures to resume its unfinished transmission at the preselected target channel. This kind of listen-before-talk channel access scheme is implemented in many wireless techniques, such as the quiet period of the IEEE 802.22 standard [42] and the clear channel assessment (CCA) of the IEEE 802.11 standard [43].

3.2 Illustrative Example of Spectrum Handoffs with Multiple Interruptions

A secondary connection may encounter multiple interruption requests during its transmission period. Because spectrum handoff procedures must be performed whenever an interruption occurs, a set of target channels will be sequentially selected, called the *target channel sequence* in this paper. Fig. 1 shows an example that three spectrum handoff requests occur during the transmission period of the secondary connection SC_A. In this example, SC_A's initial (default) channel is Ch1 and its *target channel sequence* for spectrum handoffs is (Ch2, Ch2, Ch3, $\cdots$). The extended data delivery time of SC_A is denoted by T. Furthermore, D_i is the handoff delay of the i^{th} interruption. Here, the handoff delay is the duration from the instant when the transmission is interrupted until the instant when the unfinished transmission is resumed. We assume that the transmitter of SC_A plans to establish a connection flow consisting of 28 slots to the corresponding receiver. Then, the transmission process with multiple handoffs is described as follows:

1) In the beginning, SC_A is established at its default channel Ch1. When an interruption event occurs, SC_A decides its target channel according to the predetermined target channel sequence.
2) At the first interruption, SC_A changes its operating channel to the idle channel Ch2 from Ch1 because the first predetermined target channel is Ch2. In this case, the handoff delay D_1 is the channel switching time (denoted by t_s).
3) At the second interruption, SC_A stays on its current operating channel Ch2 because the second target channel is Ch2. SC_A cannot be resumed until all the high-priority primary connections finish their transmissions at Ch2. In this case, the handoff delay D_2 is the duration from the time instant that Ch2 is used by the primary 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 changes its operating channel to Ch3 because the third target

channel is Ch3. In this example, because Ch3 is busy, SC$_A$ must wait in the low-priority queue until all the data in the present high-priority and low-priority queues of Ch3 are served[4]. Hence, the handoff delay D_3 is the sum of this waiting time and the channel switching time t_s.

5) Finally, SC$_A$ is completed on Ch3.

When a secondary connection changes its operating channel from channel k to k' where $k' \neq k$, the expected handoff delay is the sum of the channel switching time t_s and the average waiting time of channel k' (denoted by $\mathbf{E}[W_s^{(k')}]$) for the secondary connections. Note that this waiting time $W_s^{(k')}$ is the duration from the time instant that a secondary connection enters the low-priority queue of channel k' until it gets a chance to transmit at channel k'. After the secondary connection's operating channel is changed to channel k', one of two situations will occur. If channel k' is idle as the first interruption in Fig. 1, the expected handoff delay is t_s since $\mathbf{E}[W_s^{(k')}|\text{channel } k' \text{ is idle}] = 0$. On the other hand, the expected handoff delay is $t_s + \mathbf{E}[W_s^{(k')}|\text{channel } k' \text{ is busy}]$ if channel k' is busy as the third interruption in Fig. 1.

4 Analytical Framework

4.1 The PRP M/G/1 Queueing Network Model

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

- Each server (channel) can accept two types of customers (connections): the high priority connections from the primary users and the low-priority connections from the secondary users.
- The primary users have the preemptive priority to use the channel and can interrupt the transmission of the secondary users. The interrupted secondary user can resume the unfinished transmission instead of retransmitting the whole connection [22]. Note that the target channel of an interrupted secondary connection can be different from its current operating channel. This is a key difference from the spectrum usage models of [30]–[38], which is also based on the PRP M/G/1 queueing theory .

4. Here, the 1-persistent waiting policy is adopted. That is, the interrupted secondary user must stay on the selected target channel even though the selected channel is busy and then transmit unfinished data when channel becomes idle. Another possible approach is to reselect a new channel at the next time slot when a busy channel is selected in the current time slot. However, this approach is more impractical because it will lead to many channel-switching behaviors during a secondary connection.

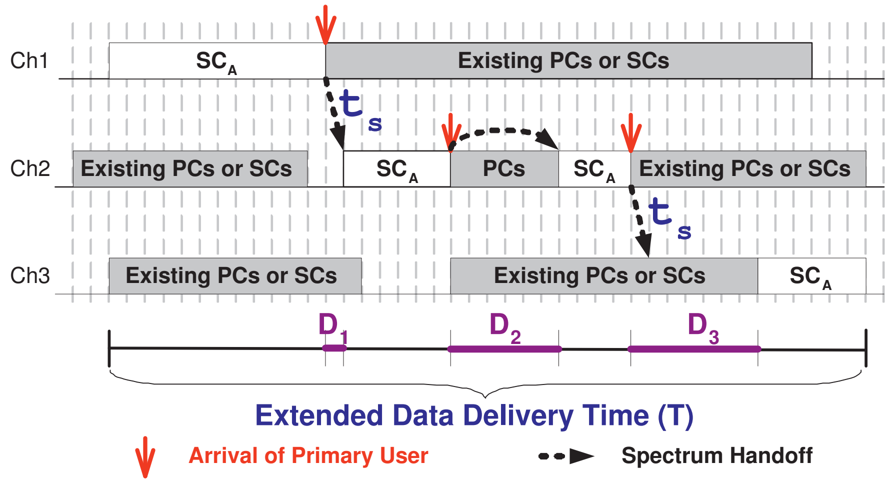


Fig. 1. An example of transmission process for the secondary connection SC$_A$, where t_s is the channel switching time, T is the extended data delivery time of SC$_A$, and D_i is the handoff delay of the i^{th} interruption. The gray areas indicate that the channels are occupied by the existing primary connections (PCs) or the other secondary users' connections (SCs). Because SC$_A$ is interrupted three times in total, the overall data connection is divided into four segments.

- During the transmission period, a secondary connection may encounter multiple interruptions from the primary users.

To ease analysis, we further make the following assumptions:

- A default channel is preassigned to each secondary user through spectrum decision algorithms in order to balance the overall traffic loads of the secondary users to all the channels [44]. When a secondary transmitter has data, it can transmit handshaking signal at the default channel of the intended receiver to establish a secondary connection [45], [46]. If the corresponding receiver's default channel is busy, the secondary transmitter must wait at this channel until it becomes available [9].
- Each primary connection is assigned with a default or licensed channel.
- Each secondary user can detect the presence of the primary user. In fact, this model can be also extended to consider the effects of false alarm and missed detection [47].
- Any time only one user can transmit data at one channel.

4.2 Example

Figure 2 shows an example of the PRP M/G/1 queueing network model with three channels, in which 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. When a primary connection appears at the channel being occupied by the secondary connection, the interruption

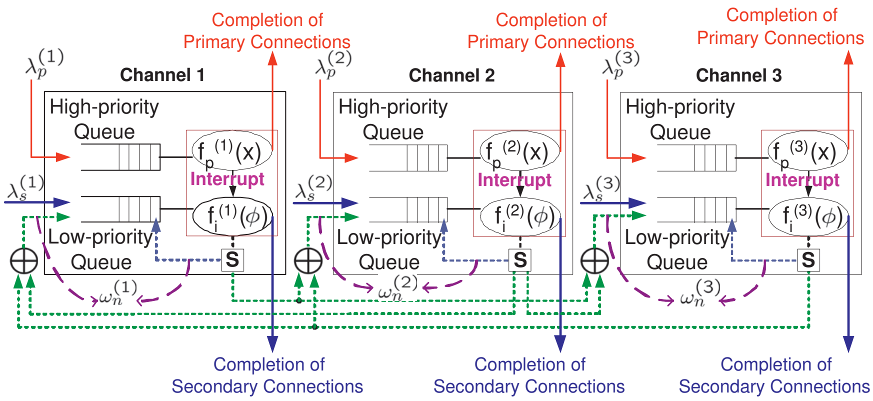


Fig. 2. The PRP M/G/1 queueing network model with three channels where $\lambda_p^{(k)}$, $\lambda_s^{(k)}$, and $\omega_n^{(k)}$ are the arrival rates of the primary connections, the secondary connections, and the type-n secondary connections ($n \geq 1$) at channel k. Furthermore, $f_p^{(k)}(x)$ and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

event occurs. The interrupted secondary connection decides its target channel for spectrum handoff according to the target channel predetermination algorithm which is implemented in the channel selection point $\boxed{\text{S}}$. In our queueing network model, the interrupted secondary connection can either stay on its current channel or change to another channel through different feedback paths. If a secondary connection chooses to stay on its current operating channel, its remaining data will be connected to the head of the low-priority queue of its current operating channel. On the other hand, if the decision is to change its operating channel, the remaining data of the interrupted secondary connection will be connected to the tail of the low-priority queue of the selected channel after channel switching time t_s. In order to characterize the handoff delay from channel switching time t_s, $\boxed{\text{S}}$ must be regarded as a server with constant service time t_s. Note that $\oplus$ in the figure represents that the traffic of the interrupted secondary connections 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 in multi-channel spectrum handoffs.

4.3 Traffic Parameters

The proposed PRP M/G/1 queueing network model requires the following traffic parameters. Assume that the arrival processes of the primary and the secondary connections on the queues of each channel are Poisson. Let $\lambda_p^{(\eta)}$ (arrivals/slot) be the traffic arrival rate of the primary connections at channel η, and $\lambda_s^{(\eta)}$ (arrivals/slot) be the secondary connection's initial arrival rate at channel η. Furthermore, let $X_p^{(\eta)}$ (slots/arrival) and $X_s^{(\eta)}$ (slots/arrival) be the corresponding service time of the primary and the secondary connections, respectively, as well as $f_p^{(\eta)}(x)$ and $f_s^{(\eta)}(x)$ be the pdfs of $X_p^{(\eta)}$ and $X_s^{(\eta)}$, respectively. If the four traffic parameters $\lambda_p^{(\eta)}$, $\lambda_s^{(\eta)}$, $f_p^{(\eta)}(x)$, and $f_s^{(\eta)}(x)$ can be obtained by certain traffic pattern prediction methods [48], many performance measures for the multi-channel spectrum handoffs with multiple interruptions can be derived.

Now, we define the type-i secondary connection as the secondary connection that has experienced i interruptions, where $i \geq 0$. For the type-i secondary connections, two more important system parameters $\omega_i^{(k)}$ and $\Phi_i^{(k)}$ are defined as follows:

- $\omega_i^{(k)}$ is the arrival rate of the type-i secondary connections at channel k[5] How to derive $\omega_i^{(k)}$ from the four traffic parameters is discussed in Appendix A.
- $\Phi_i^{(k)}$ is the effective service time of the type-i secondary connections at channel k. That is, $\Phi_i^{(k)}$ is the transmission duration of a secondary connection between the i^{th} and the $(i+1)^{th}$ interruptions at channel k. Furthermore, let $f_i^{(k)}(\phi)$ be the pdf of $\Phi_i^{(k)}$. In Appendix B, we will discuss how to derive $f_i^{(k)}(\phi)$ from the four traffic parameters.

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. Moreover, the busy probability of channel k is denoted by $\rho^{(k)}$. Let n_{max} be the maximum allowable number of interruptions for the secondary connections. That is, a secondary connection will be dropped when it encounters the $(n_{max}+1)^{th}$ interruption[6]. Moreover, in an M-channel network, the following constraint shall be satisfied for $1 \leq k \leq M$:

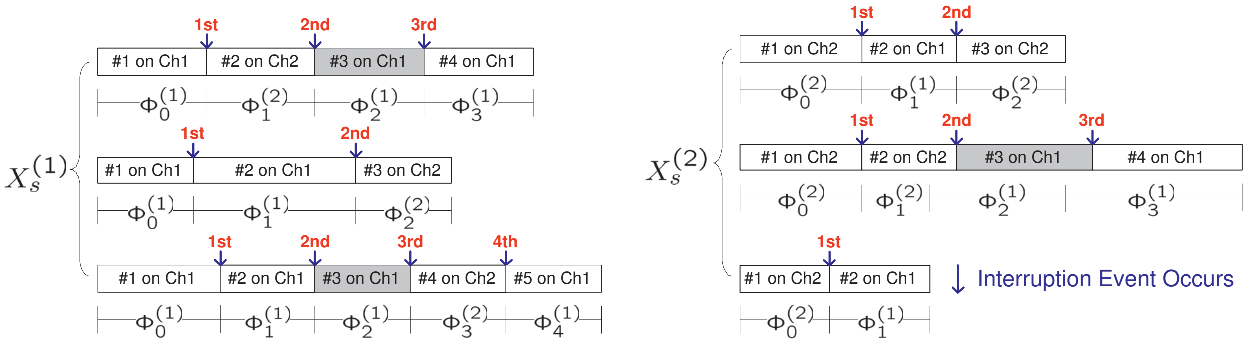
$$\rho^{(k)} = \frac{\tau_p}{\tau} + \sum_{i=0}^{n_{max}} \frac{\tau_i}{\tau} = \rho_p^{(k)} + \sum_{i=0}^{n_{max}} \rho_i^{(k)} < 1 \ , \tag{1}$$

where τ_p and τ_i are the time occupied by primary and type-i secondary connections, and τ is the total observed time. Note that $\rho_p^{(k)} = \lambda_p^{(k)} \mathbf{E}[X_p^{(k)}] < 1$ and $\rho_i^{(k)} = \omega_i^{(k)} \mathbf{E}[\Phi_i^{(k)}] < 1$ as well as $\rho^{(k)}$ can be also interpreted as the utilization factor of channel k for each k .

Figure 3 illustrates the physical meaning of random variable $\Phi_i^{(k)}$. Consider a two-channel network with the service time of the secondary connections $X_s^{(1)}$ and $X_s^{(2)}$ at channels 1 and 2, respectively. In channel 1, random variable $X_s^{(1)}$ are generated three times in Fig. 3(a). Similarly, Fig. 3(b) shows the three realizations of $X_s^{(2)}$ for channel 2. Each secondary connection is

5. Note that when a new secondary connection arrives at channel k, it will become the type-0 secondary connection at channel k because this secondary connection has experiences 0 interruptions. Hence, we have $\omega_0^{(\eta)} = \lambda_s^{(\eta)}$.

6. Intuitively, a larger value of n_{max} results in the higher complexity to determine the optimal target channel. However, a smaller value of n_{max} will reduce the quality of service (QoS) performance of the secondary users because a secondary will be dropped more frequently. Hence, determining the optimal n_{max} is a system-dependent issue.



(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, $\Phi_2^{(1)}$ is one of the third segments (gray areas) of the first and the third secondary connections in (a) as well as the second secondary connection in (b). Note that the third secondary connection in (b) does not have the third segment because it is interrupted only once.

divided into many segments due to multiple primary users' interruptions. For example, the first secondary connection in Fig. 3(a) is divided into four segments because it encounters three interruptions in total. The first, second, third, and fourth segments are transmitted at channels 1, 2, 1, and 1, respectively. Thus, this secondary connection's default channel is Ch1 and its target channel sequence is (Ch2, Ch1, Ch1). In Fig. 3, random variable $\Phi_2^{(1)}$, one of the gray regions, represents the transmission duration of a secondary connection between the 2^{nd} and the 3^{rd} interruptions at channel 1. That is, $\Phi_2^{(1)}$ is the third segment of the first secondary connections or the third segment of the third secondary connections in Fig. 3(a), or the third segment of the second secondary connection in Fig. 3(b).

In this paper, each secondary connection is divided into many segments due to multiple interruptions as shown in Fig. 3. Note that the operating channel of each segment can be any channel. Because the effective service time ($\Phi_i^{(k)}$) of each segment is dependent on the traffic statistics of the primary and other secondary users of the selected target channels, it is quite complex to find the probability density function of the effective service time of each segment. Fortunately, based on the proposed analytical framework, we provide a systematic approach to study the effects of various system parameters on the effective service time and then can derive the closed-form expression for the average effective service time of each segment.

TABLE 2

Definitions of notations.

$\lambda_p^{(\eta)}$	traffic arrival rate of the primary connections at channel η
$\lambda_s^{(\eta)}$	initial traffic arrival rate of the secondary connections at default channel η
$X_p^{(\eta)}$	service time of the primary connections at channel η
$X_s^{(\eta)}$	service time of the secondary connections at default channel η
t_s	channel switching time
D_i	handoff delay for the i^{th} interruption
n_{max}	maximum number of interruptions for the secondary connections
$s_{i,\eta}$	the target channel at the i^{th} interruption
$p_i^{(s_{i,\eta})}$	interrupted probability when the secondary connection has experienced i interruptions

5 Analysis of Extended Data Delivery Time

Based on the proposed PRP M/G/1 queuing network model, we can evaluate many performance metrics of the secondary connections with various target channel sequences. In this paper, we focus on the analysis of the extended data delivery time, which is an important performance measure for the latency-sensitive traffic of the secondary connections. Some important notations used in this paper had been summarized in Table 2.

A secondary connection may encounter many interruptions during its transmission period. Without loss of generality, we consider a secondary connection whose default channel is channel η in the following discussions. Let N be the total number of interruptions of this secondary connection. Then, the average extended data delivery time of this secondary connection can be expressed as

$$\mathbf{E}[T] = \sum_{n=1}^{n_{max}} \mathbf{E}[T|N=n]\mathbf{Pr}(N=n) \ . \tag{2}$$

Note that we can evaluate the extended data delivery time resulting from various target channel sequences from (2). Then, by comparing the extended data delivery time resulting from all possible target channel sequences, the optimal target channel sequence can be determined to minimize the extended data delivery time.

First, we show how to derive the value of $\mathbf{E}[T|N=n]$ of (2). The considered secondary connection can be divided into many segments due to multiple interruptions as discussed in Fig. 1. Hence, the extended data delivery time of this secondary connection consists of the original service time and the cumulative delay resulting from multiple handoffs. Let D_i be the handoff delay of the considered secondary connection for the i^{th} interruption. When $N=n$, we have $D_i=0$ if $i \geq n+1$. Then, the conditional expectation of the extended data delivery time of the considered secondary connection given the event $N = n$ can be derived as

$$\mathbf{E}[T|N = n] = \mathbf{E}[X_s^{(\eta)}] + \sum_{i=1}^{n} \mathbf{E}[D_i] \ . \quad (3)$$

Next, we investigate how to derive the value of $\mathbf{Pr}(N = n)$ of (2). For the considered secondary connection, denote $s_{0,\eta}$ and $s_{i,\eta}$ as its default channel and its target channel at the i^{th} interruption, respectively. Thus, we have $s_{0,\eta} = \eta$ and this secondary connection's target channel sequence can be expressed as $(s_{1,\eta}, s_{2,\eta}, s_{3,\eta}, \cdots)$. Let $p_i^{(s_{i,\eta})}$ be the probability that the considered secondary connection is interrupted again at channel $s_{i,\eta}$ when it has experienced i interruption. Then, the probability that the considered secondary connection is interrupted exactly n times can be expressed as

$$\mathbf{Pr}(N = n) = (1 - p_n^{(s_{n,\eta})}) \prod_{i=0}^{n-1} p_i^{(s_{i,\eta})} \ . \quad (4)$$

Finally, substituting (3) an (4) into (2) yields

$$\mathbf{E}[T] = \mathbf{E}[X_s^{(\eta)}] + \sum_{n=1}^{n_{max}} \left[\left(\sum_{i=1}^{n} \mathbf{E}[D_i] \right) (1 - p_n^{(s_{n,\eta})}) \prod_{i=0}^{n-1} p_i^{(s_{i,\eta})} \right] , \quad (5)$$

where the values of $\mathbf{E}[D_i]$ and $p_i^{(k)}$ can be obtained from the Propositions 1 and 2, respectively.

Proposition 1.

$$\boldsymbol{E}[D_i] = \begin{cases} \boldsymbol{E}[Y_p^{(s_{i,\eta})}] & , \quad s_{i-1,\eta} = s_{i,\eta} \\ \boldsymbol{E}[W_s^{(s_{i,\eta})}] + t_s & , \quad s_{i-1,\eta} \neq s_{i,\eta} \end{cases} , \quad (6)$$

where

$$\boldsymbol{E}[Y_p^{(k)}] = \frac{\boldsymbol{E}[X_p^{(k)}]}{1 - \rho_p^{(k)}} = \frac{\boldsymbol{E}[X_p^{(k)}]}{1 - \lambda_p^{(k)} \boldsymbol{E}[X_p^{(k)}]} , \quad (7)$$

and

$$\boldsymbol{E}[W_s^{(k)}] = \frac{\lambda_p^{(k)} \boldsymbol{E}[(X_p^{(k)})^2] + \sum_{i=0}^{n_{max}} \omega_i^{(k)} \boldsymbol{E}[(\Phi_i^{(k)})^2] + \frac{(\lambda_p^{(k)})^2 \boldsymbol{E}[(X_p^{(k)})^2]}{1 - \lambda_p^{(k)} \boldsymbol{E}[X_p^{(k)}]} \boldsymbol{E}[X_p^{(k)}]}{2(1 - \lambda_p^{(k)} \boldsymbol{E}[X_p^{(k)}] - \sum_{i=0}^{n_{max}} \omega_i^{(k)} \boldsymbol{E}[\Phi_i^{(k)}])} \ . \quad (8)$$

Proof: The handoff delay $\mathbf{E}[D_i]$ depends on which channel is selected for the target channel at the i^{th} interruption. For the secondary connection with $(i-1)$ interruptions, its current operating channel is $s_{i-1,\eta}$. When it is interrupted again, its new operating channel is $s_{i,\eta}$. When $s_{i-1,\eta} = s_{i,\eta}$, it means that the considered secondary connection will stay on the current channel. When $s_{i-1,\eta} \neq s_{i,\eta}$, it represents that the considered secondary connection will change its operating channel to another channel. Both cases are discussed as follows.

(1) Staying case: When the considered secondary connection stays on its current operating channel $s_{i,\eta} = k$, it cannot be resumed until all the high-priority primary connections of channel k finish their transmissions. Hence, the handoff delay is the busy period resulting from multiple primary connections of channel k (denoted by $Y_p^{(k)}$) as discussed in Section 3.2. That is, we can have $\mathbf{E}[D_i] = \mathbf{E}[Y_p^{(k)}]$.

The value of $\mathbf{E}[Y_p^{(k)}]$ can be derived as follows. Denote I_p as the idle period resulting from the primary connections. This idle period is the duration from the termination of the busy period to the arrival of the next primary connection. Because of the memoryless property, the idle period follows the exponential distribution with rate $\lambda_p^{(k)}$ [49]. Hence, we have

$$\mathbf{E}[I_p^{(k)}] = \frac{1}{\lambda_p^{(k)}} \ . \tag{9}$$

Next, according to the definition of the utilization factor at channel k, we have

$$\rho_p^{(k)} = \lambda_p^{(k)}\mathbf{E}[X_p^{(k)}] \ . \tag{10}$$

Because $\rho_p^{(k)}$ is also the busy probability resulting from the primary connections of channel k, we have

$$\rho_p^{(k)} = \frac{\mathbf{E}[Y_p^{(k)}]}{\mathbf{E}[Y_p^{(k)}] + \mathbf{E}[I_p^{(k)}]} \ . \tag{11}$$

Then, substituting (9) and (10) into (11), we can obtain (7).

(2) Changing case: In this case, the considered secondary connection will change to channel $s_{i,\eta} = k'$. After switching channel from channel k to k', it must wait in the low-priority queue of channel k' until all the traffic in the high-priority and the present low-priority queues of channel k' are served as discussed in Section 3.2. Denote $W_s^{(k')}$ as this waiting time for the secondary connections at channel k'[7]. Hence, we have $\mathbf{E}[D_i] = \mathbf{E}[W_s^{(k')}] + t_s$.

The value of $\mathbf{E}[W_s^{(k')}]$ can be derived as follows. Let $\mathbf{E}[Q_p^{(k')}]$ be the average number of the primary connections which are waiting in the high-priority queue of channel k' and $\mathbf{E}[Q_i^{(k')}]$ be the average number of the type-i secondary connections which are waiting in the low-priority queue of channel k'. Note that the type-i and type-j secondary connections have the same priority to access channel for any i and j. Because the newly arriving secondary connections cannot be established until all the secondary connections in the low-priority queue and the primary connections in the high-priority queue have been served, the average waiting time of channel k' is expressed as

$$\mathbf{E}[W_s^{(k')}] = \mathbf{E}[R_s^{(k')}] + \mathbf{E}[Q_p^{(k')}]\mathbf{E}[X_p^{(k')}] + \sum_{i=0}^{n_{max}} \mathbf{E}[Q_i^{(k')}]\mathbf{E}[\Phi_i^{(k')}] + \lambda_p^{(k')}\mathbf{E}[W_s^{(k')}]\mathbf{E}[X_p^{(k')}] \ , \tag{12}$$

7. A secondary connection needs to change its operating channel only when a primary connection appears. Because the arrivals of the primary connections follow Poisson distribution, the arrivals of the interrupted secondary connections at channel k' also follow Poisson distribution. Applying the property of Poisson arrivals see time average (PASTA) to the arrivals of the interrupted secondary connections at channel k' [50], all of them must spend time duration $\mathbf{E}[W_s^{(k')}]$ on average to wait for an idle channel k'. This waiting time is uncorrelated to the number of interruptions.

where $\mathbf{E}[R_s^{(k')}]$ is the average residual effective service time of channel k'. That is, $\mathbf{E}[R_s^{(k')}]$ is the remaining time to complete the service of the connection being served at channel k'. This connection being served can be the primary connection or the type-i secondary connection. Furthermore, $\mathbf{E}[Q_p^{(k')}]\mathbf{E}[X_p^{(k')}]$ and $\sum_{i=0}^{n_{max}} \mathbf{E}[Q_i^{(k')}]\mathbf{E}[\Phi_i^{(k')}]$ in (12) are the cumulative workload resulting from the primary connections and the secondary connections in the present queues of channel k', respectively. Moreover, the fourth term ($\lambda_p^{(k')}\mathbf{E}[W_s^{(k')}]\mathbf{E}[X_p^{(k')}]$) in (12) is the cumulative workload resulting from the arrivals of the primary connections during $W_s^{(k')}$.

In (12), the closed-form expression for $\mathbf{E}[\Phi_i^{(k')}]$ is derived in Appendix B. Next, we will derive $\mathbf{E}[R_s^{(k')}]$, $\mathbf{E}[Q_p^{(k')}]$, and $\mathbf{E}[Q_i^{(k')}]$. Firstly, according to the definition of residual time in [51], we have

$$\mathbf{E}[R_s^{(k')}] = \frac{1}{2}\lambda_p^{(k')}\mathbf{E}[(X_p^{(k')})^2] + \frac{1}{2}\sum_{i=0}^{n_{max}} \omega_i^{(k')}\mathbf{E}[(\Phi_i^{(k')})^2] \ , \tag{13}$$

where $\omega_i^{(k')}$ is derived in Appendix A. Secondly, according to Little's formula, it follows that

$$\mathbf{E}[Q_p^{(k')}] = \lambda_p^{(k')}\mathbf{E}[W_p^{(k')}] \ , \tag{14}$$

where $\mathbf{E}[W_p^{(k')}]$ is the average waiting time of the primary connections at channel k'. It is the duration from the time instant that a primary connection enters the high-priority queue of channel k' until it gets a chance to transmit at channel k'. Hence, it follows that

$$\mathbf{E}[W_p^{(k')}] = \mathbf{E}[R_p^{(k')}] + \mathbf{E}[Q_p^{(k')}]\mathbf{E}[X_p^{(k')}] \ , \tag{15}$$

where $\mathbf{E}[R_p^{(k')}]$ is the average residual service time resulting from only the primary connections of channel k' and $\mathbf{E}[Q_p^{(k')}]\mathbf{E}[X_p^{(k')}]$ is the total workload of the primary connections in the present high-priority queue of channel k'. According to [51], we have $\mathbf{E}[R_p^{(k')}] = \frac{1}{2}\lambda_p^{(k')}\mathbf{E}[(X_p^{(k')})^2]$. Then, solving (14) and (15) simultaneously yields

$$\mathbf{E}[W_p^{(k')}] = \frac{\mathbf{E}[R_p^{(k')}]}{1-\rho_p^{(k')}} = \frac{\lambda_p^{(k')}\mathbf{E}[(X_p^{(k')})^2]}{2(1-\lambda_p^{(k')}\mathbf{E}[X_p^{(k')}])} \ , \tag{16}$$

and

$$\mathbf{E}[Q_p^{(k')}] = \frac{\lambda_p^{(k')}\mathbf{E}[R_p^{(k')}]}{1-\rho_p^{(k')}} = \frac{(\lambda_p^{(k')})^2\mathbf{E}[(X_p^{(k')})^2]}{2(1-\lambda_p^{(k')}\mathbf{E}[X_p^{(k')}])} \ . \tag{17}$$

Next, according to Little's formula, we can obtain

$$\mathbf{E}[Q_i^{(k')}] = \omega_i^{(k')}\mathbf{E}[W_s^{(k')}] \ . \tag{18}$$

Finally, substituting (13), (17), and (18) into (12), we can obtain (8).

□

Proposition 2.

$$p_i^{(k)} = \begin{cases} \lambda_p^{(k)}\mathbf{E}[\Phi_i^{(k)}] & , \quad k = s_{i,\eta} \\ 0 & , \quad k \neq s_{i,\eta} \end{cases} \ . \tag{19}$$

Proof: The value of $p_i^{(k)}$ can be evaluated as follows. Because the considered secondary connection will operate at channel $s_{i,\eta}$ after i^{th} interruption, we have $p_i^{(k)} = 0$ when $k \neq s_{i,\eta}$. Furthermore, for the case that $k = s_{i,\eta}$, we consider the time interval $[0, t]$ at channel k. Total $\lambda_p^{(k)} t$ primary connections and $\omega_i^{(k)} t$ type-i secondary connections arrive at channel k during this interval. Hence, there are total $\omega_i^{(k)} t p_i^{(k)}$ type-i secondary connections will be interrupted on average during this interval. Furthermore, applying the property of Poisson arrivals see time average (PASTA) to the arrivals of the primary connections [50], we can obtain the probability that a primary connection finds channel k being occupied by the type-i secondary connections is $\rho_i^{(k)}$. Thus, during this interval, total $\lambda_p^{(k)} t \rho_i^{(k)}$ primary connections can see a busy channel being occupied by the type-i secondary connections. For each primary connection, it can interrupt only one secondary connection when it arrives at a busy channel being occupied by the secondary connection because only one secondary user can transmit at any instant of time. Thus, the total number of the interrupted secondary connections at channel k is also $\lambda_p^{(k)} t \rho_i^{(k)}$. Hence, we have $\omega_i^{(k)} t p_i^{(k)} = \lambda_p^{(k)} t \rho_i^{(k)}$. That is,

$$\rho_i^{(k)} = \frac{\omega_i^{(k)}}{\lambda_p^{(k)}} p_i^{(k)} \quad . \tag{20}$$

Next, we consider a type-i secondary connection at channel k. Before the $(i+1)^{th}$ interruption event occurs, its effective service time is $\mathbf{E}[\Phi_i^{(k)}]$. Thus, from queueing theory, we can have

$$\rho_i^{(k)} = \omega_i^{(k)} \mathbf{E}[\Phi_i^{(k)}] \quad . \tag{21}$$

Comparing (20) and (21), we can obtain (19). □

6 APPLICATIONS TO PERFORMANCE ANALYSIS BASED ON TWO TYPICAL TARGET CHANNEL SEQUENCES

To demonstrate the usefulness of the developed analytical method, we apply these analytical results in Section 5 to two typical target channel sequences used in the IEEE 802.22 WRAN standard[8]. Specifically, we consider the *always-staying* and *always-changing* spectrum handoff sequences, which are respectively introduced in the non-hopping mode and the phase-shifting hopping mode of the IEEE 802.22 standard [52]. From the analytical results, an adaptive target channel selection approach can be provided.

8. IEEE 802.22 is the first worldwide wireless standard based on cognitive radios (CR). Because IEEE 802.22 is a well-known standard in the research area of CR, we take it as an example to demonstrate how to use the proposed analytical model.

6.1 Derivation of Extended Data Delivery Time

For the *always-staying* sequence, a secondary connection always stays on its default channel η when it is interrupted. That is, its target channel sequence can be expressed as (Chη, Chη, Chη, $\cdots$) and thus $s_{i,\eta} = \eta$ for each i. Hence, we can have $\mathbf{E}[D_i] = \mathbf{E}[Y_p^{(\eta)}]$ for each i in (5). Then, the average extended data delivery time of the secondary connections for the always-staying sequence can be expressed as follows:

$$\mathbf{E}[T_{stay}] = \mathbf{E}[X_s^{(\eta)}] + \sum_{n=1}^{n_{max}} \left(\sum_{i=1}^{n} \mathbf{E}[Y_p^{(\eta)}] \right) (1 - p_i^{(\eta)}) \prod_{i=0}^{n-1} p_i^{(\eta)} \ . \tag{22}$$

Next, we consider the *always-changing* sequence. In this case, the secondary connection sequentially changes its operating channel to the next neighboring channel. Without loss of generality, its corresponding target channel sequence can be expressed as (Ch$(\eta+1)$, Ch$(\eta+2)$, $\cdots$, ChM, Ch1, Ch2, $\cdots$, Chη, Ch$(\eta+1)$, $\cdots$), where channel η is the default channel of the secondary connection. That is, at the i^{th} interruption, the target channel of the interrupted secondary connection is channel $s_{i,\eta} \equiv \mathcal{MOD}(i+\eta, M)$ where $\mathcal{MOD}(a,b)$ is the Modulus function and it returns the remainder resulting when a is divided by b. Hence, we have $\mathbf{E}[D_i] = \mathbf{E}[W_s^{(s_{i,\eta})}] + t_s$ for each i in (5). Thus, the average extended data delivery time of the secondary connections for the always-changing sequence can be expressed as follows:

$$\mathbf{E}[T_{change}] = \mathbf{E}[X_s^{(\eta)}] + \sum_{n=1}^{n_{max}} \left[\left(\sum_{i=1}^{n} (\mathbf{E}[W_s^{(s_{i,\eta})}] + t_s) \right) (1 - p_i^{(s_{n,\eta})}) \prod_{i=0}^{n-1} p_i^{(s_{i,\eta})} \right] \ . \tag{23}$$

Based on the analytical results, the secondary connection can adaptively adopt the better target channel sequence to reduce its extended data delivery time. Thus, the average extended data delivery time with this adaptive channel selection principle (denoted by $\mathbf{E}[T^*]$) can be expressed as follows:

$$\mathbf{E}[T^*] = \min \left(\mathbf{E}[T_{stay}], \mathbf{E}[T_{change}] \right) \ . \tag{24}$$

6.2 An Example for Homogeneous Traffic Loads

Now, we give an example to explain how to apply our analytical results to find the better target channel sequence when traffic parameters are given. We consider a special case that the primary and the secondary connections have the same traffic parameters in a three-channel system (i.e., $\lambda_p^{(1)} = \lambda_p^{(2)} = \lambda_p^{(3)} \equiv \lambda_p$, $\lambda_s^{(1)} = \lambda_s^{(2)} \equiv \lambda_s$, and $\mathbf{E}[X_p^{(1)}] = \mathbf{E}[X_p^{(2)}] = \mathbf{E}[X_p^{(3)}] \equiv \mathbf{E}[X_p]$). Because the three channels are identical, three channels have the same performance metrics. Thus, the superscript (k) can be dropped to ease the notations. Furthermore, we assume that the service time of the secondary connections follows the same exponential distribution, i.e., $f_s^{(1)}(x) = f_s^{(2)}(x) = f_s^{(3)}(x) \equiv f_s(x) = \mu_s e^{-\mu_s x}$. Hence, we have $\mathbf{E}[X_s^{(1)}] = \mathbf{E}[X_s^{(2)}] = \mathbf{E}[X_s^{(3)}] \equiv \mathbf{E}[X_s] = \frac{1}{\mu_s}$.

6.2.1 Derivation of $p_i^{(\eta)}$ and $\mathbf{E}[Y_p^{(\eta)}]$ in (22)

First, according to Appendix B, we can derive $\mathbf{E}[\Phi_i^{(\eta)}]$ as follows:

$$\mathbf{E}[\Phi_i^{(\eta)}] = \mathbf{E}[\Phi_i] = \frac{1}{\lambda_p + \mu_s} \quad . \tag{25}$$

Then, the value of $p_i^{(\eta)}$ can be derived from (19) as follows:

$$p_i^{(\eta)} = \lambda_p^{(\eta)} \mathbf{E}[\Phi_i^{(\eta)}] = \frac{\lambda_p}{\lambda_p + \mu_s} \equiv p_i \quad . \tag{26}$$

Next, referring to (7), it follows that

$$\mathbf{E}[Y_p^{(\eta)}] = \mathbf{E}[Y_p] = \frac{\mathbf{E}[X_p]}{1 - \lambda_p \mathbf{E}[X_p]} \quad . \tag{27}$$

Finally, substituting (26) and (27) into (22), we can obtain the closed-form expression for the extended data delivery time with the always-staying target channel sequence.

6.2.2 Derivation of $\mathbf{E}[W_s^{(s_{i,\eta})}]$ and $p_i^{(s_{i,\eta})}$ in (23)

Referring to Appendixes A and B, we can have

$$\omega_i^{(s_{i,\eta})} = \omega_i = \lambda_s \left(\frac{\lambda_p}{\lambda_p + \mu_s}\right)^i \quad , \tag{28}$$

and

$$\mathbf{E}[(\Phi_i^{(s_{i,\eta})})^2] = \mathbf{E}[(\Phi_i)^2] = \frac{2}{(\lambda_p + \mu_s)^2} \quad . \tag{29}$$

Next, substituting (25), (28), and (29) into (8), we can have

$$\mathbf{E}[W_s^{(s_{i,\eta})}] = \mathbf{E}[W_s] = \frac{\lambda_p \mathbf{E}[(X_p)^2] + \frac{2\lambda_s \mathbf{E}[X_s]}{(\lambda_p + \mu_s)} + \frac{(\lambda_p)^2 \mathbf{E}[(X_p)^2]}{1 - \lambda_p \mathbf{E}[X_p]} \mathbf{E}[X_p]}{2(1 - \lambda_p \mathbf{E}[X_p] - \lambda_s \mathbf{E}[X_s])} \quad . \tag{30}$$

Then, referring to (19), it follows that

$$p_i^{(s_{i,\eta})} = p_i = \frac{\lambda_p}{\lambda_p + \mu_s} \quad . \tag{31}$$

Finally, substituting (30) and (31) into (23), we can obtain the closed-form expression for the extended data delivery time with the always-changing target channel sequence. Note that this closed-form expression for p_i in this special case had been discussed in [7]. However, [7] cannot extend to the case with the general service time distribution.

In summary, the average extended data delivery time with our adaptive target channel selection approach can be expressed as follows:

$$\mathbf{E}[T^*] = \begin{cases} \mathbf{E}[T_{stay}] & , \quad \mathbf{E}[Y_p] \leq \mathbf{E}[W_s] + t_s \\ \mathbf{E}[T_{change}] & , \quad \mathbf{E}[Y_p] \geq \mathbf{E}[W_s] + t_s \end{cases} \quad . \tag{32}$$

Note that the always-staying and the always-changing sequences have the same extended data delivery time when $\mathbf{E}[Y_p] = \mathbf{E}[W_s] + t_s$.

7 NUMERICAL RESULTS

We show numerical results to reveal the importance of the three key design features for modeling spectrum handoffs as discussed in Section 2, which consist of 1) general service time distribution; 2) various operating channels; and 3) queueing behaviors of multiple secondary connections. Here, we only show the effects of these key design features on $\mathbf{E}[T]$. The effects on other performance metrics (such as $\mathbf{E}[T]/\mathbf{E}[X_s]$) can be derived based on similar manners.

7.1 Simulation Setup

In order to validate the proposed analytical model, we perform simulations in non-slot-based (continuous-time) cognitive radio systems, where the inter-arrival time and service time can be the duration of non-integer time slots. We consider a three-channel CR system with Poisson arrival processes of rates λ_p and λ_s 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, and the connections with the same priority follow the first-come-first-served (FCFS) scheduling discipline[9]. Referring to the IEEE 802.22 standard, we adopt time slot duration of 10 msec in our simulations [54].

7.2 Effects of Various Service Time Distributions for Primary Connections

Firstly, we investigate the effects of various service time distributions for primary connections on the extended data delivery time of the secondary connections. The truncated Pareto distribution and the exponential distribution are considered in our simulations. Referring to [49], these two distributions match the actual data and voice traffic measurements very well, respectively. The truncated Pareto distribution is expressed as follows:

$$f_X(x) = \begin{cases} \alpha \frac{K^\alpha}{x^{\alpha+1}} & , \quad K \le x \le m \\ \frac{K^\alpha}{m^\alpha} & , \quad x = m \end{cases} . \tag{33}$$

According to [55], the traffic shaping parameter $\alpha = 1.1$ and the scale parameter $K = 81.5$, and the truncated upper bound $m = 66666$ bytes in (33). Then, the average connection length is 480 bytes for the primary connections. Similarly, we also assume that the average connection length is 480 bytes when the exponentially distributed primary connections are considered. Moreover, we assume that $\mathbf{E}[X_s^{(1)}] = \mathbf{E}[X_s^{(2)}] = \mathbf{E}[X_s^{(3)}] \equiv \mathbf{E}[X_s] = 10$ (slots/arrival), and $\mathbf{E}[X_p^{(1)}] = \mathbf{E}[X_p^{(2)}] = \mathbf{E}[X_p^{(3)}] \equiv \mathbf{E}[X_p]$. When the data rate of the primary connections is 19.2

9. In fact, the analytical results of mean values obtained in this paper can be applied to other scheduling discipline which is independent of the service time of the primary and secondary connections because the averages of system performance metrics will be invariant to the order of service in this case (see page 113 in [53]).

Kbps, we have $\mathbf{E}[X_p] = \frac{480\times 8 \text{ bits}}{19.2 \text{ Kbps}} \div \frac{10 \text{ msec}}{\text{slot}} = 20$ (slots/arrival) for the Pareto and the exponential distributions. Furthermore, we consider that $\lambda_s = 0.01$ (arrivals/slot). Recall that ρ_p is the channel busy probability resulting from the transmissions of the primary connections. We only consider the case that $0 \leq \rho_p = lambda_p \mathbf{E}[X_p] < 1 - \lambda_s \mathbf{E}[X_s] = 0.9$ in the following numerical results. When $\rho_p + \lambda_s \mathbf{E}[X_s] \geq 1$ (or equivalently $\lambda_p \geq \frac{\lambda_s \mathbf{E}[X_s]}{\mathbf{E}[X_p]} = 0.045$ (arrivals/slot)), the secondary connections will encounter the infinite extended data delivery time on average.

Figure 4 compares the effects of Pareto and exponential service time distributions for primary connections when the always-changing spectrum handoff sequence is adopted. First, we find that the simulation results match the analytical results quite well, which can validate the slot-based assumption used in our analysis. Next, compared to the exponentially distributed service time for primary connections, the Pareto distributed service time results in longer average extended data delivery time in the secondary connections. This phenomenon can be interpreted as follows. Because of the heavy tail property of Pareto distribution, the second moment $\mathbf{E}[(X_p)^2]$ of service time with Pareto distribution is larger than that with exponential distribution. According to (30) and (23), an interrupted secondary connection will encounter longer waiting time and extended data delivery time when the primary connections' service time distribution is Pareto. For example, when $\rho_p = 0.44$ or equivalently $\lambda_p = \frac{\rho_p}{\mathbf{E}[X_p]} = 0.022$ (arrivals/slot), the average extended data delivery time with the Pareto-typed primary connection service time is four times longer than that with the exponential-typed primary connection service time. Because the developed analytical framework can characterize the effects of general service time distribution, it is quite useful.

When the always-staying spectrum handoff sequence is adopted, Fig. 5 shows the average extended data delivery time of the secondary connections. According to (22), the extended data delivery time in this case is related to the average busy period $\mathbf{E}[Y_p]$ for the primary connections. Because the considered Pareto and exponential distributions have the same average service time, these two distributions result in the same average busy period $\mathbf{E}[Y_p]$ for the primary connections according to (27), resulting in the same average extended data delivery time as well.

7.3 Traffic-adaptive Target Channel Selection Principle

Figure 6 compares the extended data delivery time of the always-staying and the always-changing spectrum handoff sequences when the service time of the primary connections is exponentially distributed. Based on (32), the traffic-adaptive channel selection approach can appropriately change to better target channel sequence according to traffic conditions. This figure shows that a cross-point occurs when $\rho_p = 0.44$ or equivalently $\lambda_p = \frac{\rho_p}{\mathbf{E}[X_p]} = 0.022$ (arrivals/slot), where the always-staying and the always-changing sequences result in the same

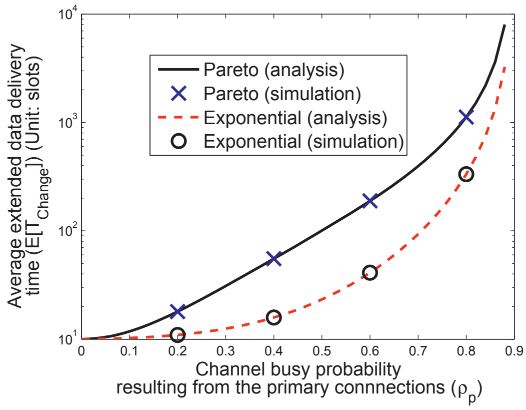


Fig. 4. Effects of Pareto and exponential service time distributions for primary connections on the extended data delivery time (**E**$[T_{change}]$) of the secondary connections when the **always-changing** spectrum handoff sequence is adopted, where $t_s = 1$ (slot), $\lambda_s = 0.01$ (arrivals/slot), **E**$[X_s] = 10$ (slots/arrival), and **E**$[X_p] = 20$ (slots/arrival).

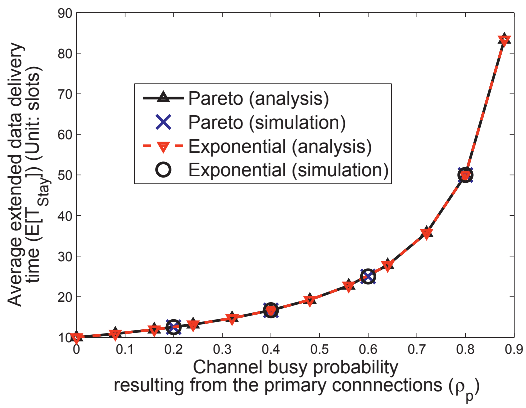


Fig. 5. Effects of Pareto and exponential service time distributions for primary connections on the extended data delivery time (**E**$[T_{stay}]$) of the secondary connections when the **always-staying** spectrum handoff sequence is adopted, where $t_s = 1$ (slot), $\lambda_s = 0.01$ (arrivals/slot), **E**$[X_s] = 10$ (slots/arrival), and **E**$[X_p] = 20$ (slots/arrival).

extended data delivery time. When $\rho_p > 0.44$, the interrupted user prefers the always-staying sequence. This phenomenon can be interpreted as follows. A larger value of ρ_p (or equivalently a larger value of λ_p) will increase the probability that an interrupted secondary user experiences long waiting time when it changes its operating channel. As a result, the average handoff delay for changing operating channel (i.e., $\mathbf{E}[W_s] + t_s$) will be extended. Then, the average extended data delivery time will be also prolonged. In our case, the secondary user prefers staying on the current operating channel when $\rho_p > 0.44$. By contrast, when $\rho_p < 0.44$, the

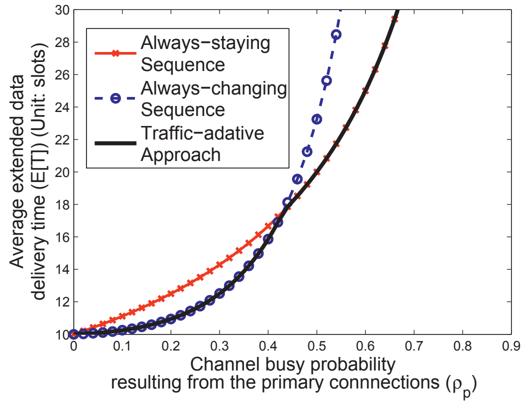


Fig. 6. Comparison of the extended data delivery time for the always-staying and always-changing spectrum handoff sequences as well as the traffic-adaptive channel selection approach, where $t_s = 1$ (slot), $\lambda_s = 0.01$ (arrivals/slot), $\mathbf{E}[X_p] = 20$ (slots/arrival), and $\mathbf{E}[X_s] = 10$ (slots/arrival).

traffic-adaptive channel selection approach can improve latency performance by changing to the always-changing sequence. For example, when $\rho_p = 0.2$, the traffic-adaptive approach can improve the extended data delivery time by 15% compared to the always-staying sequence. Compared to the single-channel spectrum handoff model [30]–[38], the developed analytical framework for multi-channel spectrum handoff is more general because it can incorporate the effects of changing operating channels.

Figure 7 shows the effect of secondary connections' service time $\mathbf{E}[X_s]$ on the cross-point for traffic-adaptive channel selection approach. According to (32), for a lager value of $\mathbf{E}[X_s]$, the interrupted secondary connection prefers staying on the current channel because the average handoff delay for changing its operating channel is longer than that for staying on the current channel. Thus, the cross-point of "always-staying" and "always-changing" sequences moves toward left-hand side as $\mathbf{E}[X_s]$ increases as seen in the figure.

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[10]. Fig. 8 shows the admissible region for the normalized traffic workloads (or channel utilities) (ρ_p, ρ_s)[11] for the Voice over IP (VoIP) services. The maximum allowable average cumulative delay resulting from multiple handoffs is 20 ms for the VoIP traffic [56]. Assume $\mathbf{E}[X_p] = 20$ (slots/arrival) and $\mathbf{E}[X_s] = 10$ (slots/arrival). The admission control policy can be designed according to this figure.

10. The admissible region can be determined by comparing the derived extended data delivery time and the maximum allowable average cumulative delay.

11. $\rho_p = \lambda_p \mathbf{E}[X_p]$ and $\rho_s = \lambda_s \mathbf{E}[X_s]$.

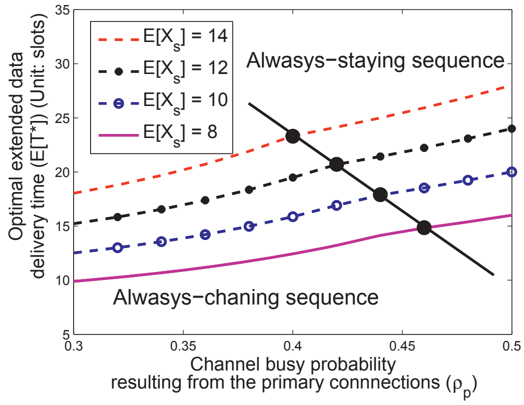


Fig. 7. Effects of secondary connections' service time **E**$[X_s]$ on the cross-point for the traffic-adaptive channel selection approach, where $t_s = 1$ (slot), **E**$[X_p] = 20$ (slots/arrival), and $\lambda_s = 0.01$ (arrivals/slot).

When $\rho_p < 0.166$, 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 $0.166 < \rho_p < 0.312$, a part of traffic workloads of the secondary users must be rejected in order to satisfy the delay constraint for the secondary users. In this case, $0.31 < \rho_p + \rho_s < 0.645$. For example, when $\rho_p = 0.25$, a CR network can support at most 0.214 workload for the secondary users. That is, a CR network can accept at most $\lambda_s = 0.0214$ (arrivals/slot) based on the results shown in the figure when $\lambda_p = 0.0125$ (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 [57] and the call admission control mechanisms in [29], [58], [59]. Finally, when $\rho_p > 0.312$, no secondary user can be accepted.

7.4 Performance Comparison between Different Channel Selection Methods

Now we compare the extended data delivery time of the following three schemes: (1) the slot-based target channel selection scheme; (2) the random-based target channel selection scheme; and (3) the traffic-adaptive target channel selection scheme. We consider a three-channel network with various traffic loads, where $\lambda_p^{(1)} = \lambda_p^{(2)} = \lambda_p^{(3)} \equiv \lambda_p$, $\lambda_s^{(1)} = \lambda_s^{(2)} = \lambda_s^{(3)} \equiv 0.01$ (arrivals/slot), $(\mathbf{E}[X_p^{(1)}], \mathbf{E}[X_p^{(2)}], \mathbf{E}[X_p^{(3)}]) = (5, 15, 25)$ (slots/arrival), and $(\mathbf{E}[X_s^{(1)}], \mathbf{E}[X_s^{(2)}], \mathbf{E}[X_s^{(3)}]) = (15, 15, 15)$ (slots/arrival). For the slot-based scheme, the secondary connections prefer selecting the channel which has the lowest busy probability resulting from the primary connections in each time slot. That is, when handoff procedures are initiated in the beginning of each time slot, all the secondary connections will select channel 1 to be their target channels. Furthermore, the random-based scheme selects one channel out of all the three channels for the target channel. Hence, each

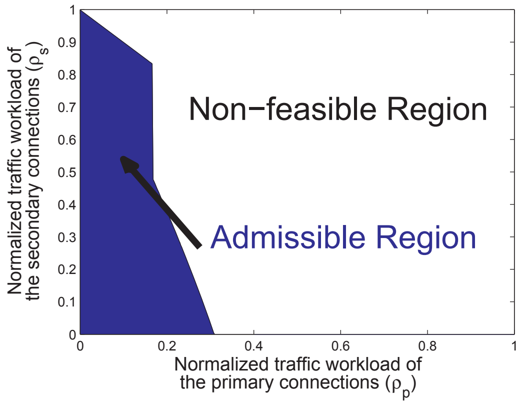


Fig. 8. Admissible region for the normalized traffic workloads (ρ_p, ρ_s), where the average cumulative delay constraint can be satisfied when $t_s = 0$ (slot), $\mathbf{E}[X_p] = 20$ (slots/arrival) and $\mathbf{E}[X_s] = 10$ (slots/arrival).

channel is selected with probability $1/3$. Moreover, based on the considered traffic parameters, the traffic-adaptive scheme will adopt the always-changing sequence and the always-staying sequence when $\lambda_p \leq 0.018$ (arrivals/slot) and $\lambda_p \geq 0.018$ (arrivals/slot), respectively. The three target channel selection schemes result in various target channel sequences. Based on the proposed analytical model, we can evaluate the average extended data delivery time resulting from these target channel sequences.

Figure 9 compares the extended data delivery time of the three target channel selection methods. We have the following three important observations. First, we consider $\lambda_p < 0.018$ (arrivals/slot). Because the probability of changing operating channel is higher than that of staying on the current operating channel for the interrupted secondary user in the random-based scheme, we can find that the average extended data delivery time for the random-based target channel selection scheme is similar to that for the traffic-adaptive target channel selection scheme, which adopts the always-changing sequence. Secondly, when $\lambda_p > 0.018$ (arrivals/slot), the traffic-adaptive scheme can shorten the average extended data delivery time because it adopts the always-staying sequence. For a larger value of λ_p, the traffic-adaptive scheme can improve the extended data delivery time more significantly. Thirdly, it is shown that the random-based and traffic-adaptive schemes can result in shorter extended data delivery time compared to the slot-based scheme. For example, when $\lambda_p = 0.018$, the random-based and traffic-adaptive schemes can improve the extended data delivery time by 35% compared to the slot-based scheme. This is because the slot-based scheme ignores the queueing behaviors of the secondary connections.

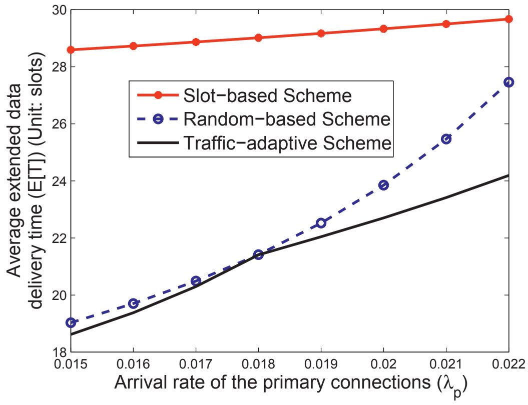


Fig. 9. Comparison of average extended data delivery time for different target channel selection sequences.

8 Conclusions

In this paper, we have proposed a PRP M/G/1 queueing network model to characterize the spectrum usage behaviors with multiple handoffs. We studied the latency performance of the secondary connections by considering the effects of (1) general service time distribution; (2) various operating channels; and (3) queueing behaviors of multiple secondary connections. The proposed model can accurately estimate the extended data delivery time of different proactively designed target channel sequences. On top of this model, we showed the extended data delivery time of the secondary connections based on the *always-staying* and the *always-changing* sequences used in the IEEE 802.22 standard. If the secondary users can adaptively adopt the better target channel sequence according to traffic conditions, the extended data delivery time can be improved significantly compared to the existing target channel selection methods, especially for the heavy traffic loads of the primary users.

Some interesting research issues can be extended from this paper. For example, we can consider other re-establishment policies rather than the resumption policy as in this paper. In this paper, we assumed that the interrupted secondary user can resume its unfinished transmission on the suitable channel. However, in other scenarios, the interrupted secondary user may need to retransmit the whole connection rather than resuming the unfinished transmission. In this situation, a CR network should be modeled by the preemptive *repeat* priority queueing network, and is worthwhile to investigate the latency performance resulted from this policy. Furthermore, how to apply the concept of priority queueing to other applications, such as electronic health (e-Health) applications [60], is also an important issue.

References

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

[2] S. Haykin, “Cognitive Radio: Brain-empowered Wireless Communications,” *IEEE Journal on Selected Areas in Communications*, 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] 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.

[5] T. A. Weiss and F. K. Jondral, “Spectrum Pooling: An Innovative Dtrategy for the Enhancement of Spectrum Efficiency,” *IEEE Radio Communications Magazine*, vol. 42, no. 3, pp. S8–S14, Mar. 2004.

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

[7] H.-J. Liu, Z.-X. Wang, S.-F. Li, and M. Yi, “Study on the Performance of Spectrum Mobility in Cognitive Wireless Network,” *IEEE Singapore International Conference on Communication Systems (ICCS)*, Jun. 2008.

[8] L.-C. Wang, C.-W. Wang, and K.-T. Feng, “A Queueing-Theoretical Framework for QoS-Enhanced Spectrum Management in Cognitive Radio Networks,” *accepted by IEEE Wireless Communications Magazine*, 2010.

[9] Q. Zhao, L. Tong, A. Swami, and Y. Chen, “Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[10] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, “Opportunistic Spectrum Access via Periodic Channel Sensing,” *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 785–796, Feb. 2008.

[11] O. Mehanna, A. Sultan, and H. E. Gamal, “Blind Cognitive MAC Protocols,” *IEEE International Conference on Communications (ICC)*, Jun. 2009.

[12] R.-T. Ma, Y.-P. Hsu, and K.-T. Feng, “A POMDP-based Spectrum Handoff Protocol for Partially Observable Cognitive Radio Networks,” *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2009.

[13] M. Hoyhtya, S. Pollin, and A. Mammela, “Performance Improvement with Predictive Channel Selection for Cognitive Radios,” *IEEE International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART)*, Feb. 2008.

[14] ——, “Classification-based Predictive Channel Selection for Cognitive Radios,” *IEEE International Conference on Communications (ICC)*, May 2010.

[15] H. Han, Q. Wu, and H. Yin, “Spectrum Sensing for Real-time Spectrum Handoff in CRNs,” *IEEE International Conference on Advanced Computer Theory and Engineering (ICACTE)*, Aug. 2010.

[16] L. Yang, L. Cao, and H. Zheng, “Proactive Channel Access in Dynamic Spectrum Networks,” *Physical Communication*, vol. 1, no. 2, pp. 103–111, 2008.

[17] S.-U. Yoon and E. Ekici, “Voluntary Spectrum Handoff: A Novel Approach to Spectrum Management in CRNs,” *IEEE International Conference on Communications (ICC)*, May 2010.

[18] Y. Song and J. Xie, “Common Hopping Based Proactive Spectrum Handoff in Cognitive Radio Ad Hoc Networks,” *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2010.

[19] A. Lertsinsrubtavee, N. Malouch, and S. Fdida, “Spectrum Handoff Strategies for Multiple Channels Cognitive Radio Network,” *ACM CoNEXT Student Workshop*, Mar. 2010.

[20] A. W. Min and K. G. Shin, “Exploiting Multi-channel Diversity in Spectrum-agile Networks,” *IEEE International Conference on Computer Communications (INFOCOM)*, Apr. 2008.

[21] X. Li, Q. Zhao, X. Guan, and L. Tong, "Optimal Cognitive Access of Markovian Channels under Tight Collision Constraints," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 746–759, Apr. 2011.

[22] S. Srinivasa and S. A. Jafar, "The Throughput Potential of Cognitive Radio: A Theoretical Perspective," *IEEE Communications Magazine*, pp. 73–79, May 2007.

[23] 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.

[24] 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.

[25] C.-W. Wang, L.-C. Wang, and F. Adachi, "Modeling and Analysis for Reactive-decision Spectrum Handoff in Cognitive Radio Networks," *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2010.

[26] 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.

[27] 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.

[28] 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.

[29] B. Wang, Z. Ji, K. J. Ray Liu, and T. C. Clancy, "Primary-prioritized Markov Approach for Dynamic Spectrum Allocation," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 1854–1865, Apr. 2009.

[30] C. Zhang, X. Wang, and J. Li, "Cooperative Cognitive Radio with Priority Queueing Analysis," *IEEE International Conference on Communications (ICC)*, Jun. 2009.

[31] H. Tran, T. Q. Duong, and H.-J. Zepernick, "Average Waiting Time of Packets with Different Priorities in Cognitive Radio Networks," *IEEE International Symposium on Wireless Pervasive Computing (ISWPC)*, May 2010.

[32] I. Suliman and J. Lehtomaki, "Queueing Analysis of Opportunistic Access in Cognitive Radios," *IEEE International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART)*, May 2009.

[33] H. Li, "Queuing Analysis of Dynamic Spectrum Access Subject to Interruptions from Primary Users," *International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, Jun. 2010.

[34] P. Zhu, J. Li, and X. Wang, "A New Channel Parameter for Cognitive Radio," *International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, Aug. 2007.

[35] ——, "Scheduling Model for Cognitive Radio," *International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, May 2008.

[36] S. Wang and H. Zheng, "A Resource Management Design for Cognitive Radio Ad Hoc Networks," *IEEE Military Communications Conference (MILCOM)*, Oct. 2009.

[37] F. Borgonovo, M. Cesana, and L. Fratta, "Throughput and Delay Bounds for Cognitive Transmissions," *Advances in Ad Hoc Networking*, vol. 265, pp. 179–190, Aug. 2008.

[38] H.-P. Shiang and M. van der Schaar, "Queuing-based Dynamic Channel Selection for Heterogeneous Multimedia Applications Over Cognitive Radio Networks," *IEEE Transactions on Multimedia*, vol. 10, no. 5, pp. 896–909, Aug. 2008.

[39] Y.-C. Liang, Y. Zeng, E. C. Peh, and A. T. Hoang, "Sensing-throughput Tradeoff for Cognitive Radio Networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, Apr. 2008.

[40] 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.

[41] W.-Y. Lee and I. F. Akyildiz, "Optimal Spectrum Sensing Framework for Cognitive Radio Networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3845–3857, Oct. 2008.

[42] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: The First Cognitive Radio Wireless Regional Area Network Standard," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130–138, Jan. 2009.

[43] IEEE 802.11, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, supplement to IEEE 802.11 Standard, Sep. 1999.

[44] L.-C. Wang, C.-W. Wang, and F. Adachi, "Load-Balancing Spectrum Decision for Cognitive Radio Networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 757–769, Apr. 2011.

[45] 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.

[46] J. Mo, H.-S. W. So, and J. Walrand, "Comparison of Multichannel MAC Protocols," *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 50–65, Jan. 2008.

[47] C.-W. Wang, "Queueing-Theoretical Spectrum Management Techniques for Cognitive Radio Networks," Ph.D. dissertation, National Chiao-Tung University, Sep. 2010.

[48] 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.

[49] Chee-Hock Ng and Boon-Hee Soong, *Queueing Modelling Fundamentals with Applications in Communication Networks, 2nd*. John Wiley & Sons Inc., 2008.

[50] R. W. Wolff, "Poisson Arrivals See Time Averages," *Operations Research*, vol. 30, no. 2, pp. 223–231, Mar./Apr. 1982.

[51] S. K. Bose, *An Introduction to Queueing Systems*. Kluwer Academic/Plenum Publishers, New York, 2002.

[52] W. Hu, D. Willkomm, G. Vlantis, M. Gerla, and A. Wolisz, "Dynamic Frequency Hopping Communities for Efficient IEEE 802.22 Operation," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 80–87, May 2007.

[53] L. Kleinrock, *Queueing Systems - Volume 2: Computer Applications*. John Wiley & Sons Inc., 1975.

[54] *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.

[55] ETSI, "Universal Mobile Telecommunications System (UMTS); Selection Procedures for The Choice of Radio Transmission Technologies of The UMTS," *Technical Report UMTS 30.03 version 3.2.0*, April 1998.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] P. Phunchongharn, D. Niyato, E. Hossain, and S. Camorlinga, "An EMI-Aware Prioritized Wireless Access Scheme for e-Health Applications in Hospital Environments," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1247–1258, Sep. 2010.

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.

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 grant from IEEE ICC 2009. His current research interests include cross-layer optimization, MAC protocols design, and radio resource management in wireless sensor networks, ad hoc networks, and cognitive radio networks.

Chung-Ju Chang (F'08) was born in Taiwan, ROC, in August 1950. He received the B.E. and M.E. degreesin electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1972 and 1976, respectively, and the Ph. D degree in electrical engineering from National Taiwan University, Taiwan, in 1985. From 1976 to 1988, he was with Telecommunication Laboratories, Directorate General of Telecommunications, Ministry of Communications, Taiwan, as a Design Engineer, Supervisor, Project Manager, and then Division Director. He also acted as a Science and Technical Advisor for the Minister of the Ministry of Communications from 1987 to 1989. In 1988, he joined the Faculty of the Department of Communication Engineering, College of Electrical Engineering and Computer Science, National Chiao Tung University, as an Associate Professor. He has been a Professor since 1993 and a Chair Professor since 2009. He was Director of the Institute of Communication Engineering from August 1993 to July 1995, Chairman of Department of Communication Engineering from August 1999 to July 2001, and Dean of the Research and Development Office from August 2002 to July 2004. Also, he was an Advisor for the Ministry of Education to promote the education of communication science and technologies for colleges and universities in Taiwan during 1995 - 1999. He is acting as a Committee Member of the Telecommunication Deliberate Body, Taiwan. Moreover, he once served as Editor for IEEE Communications Magazine and Associate Editor for IEEE Transactions Vehicular Technology. His research interests include performance evaluation, radio resources management for wireless communication networks, and traffic control for broadband networks. Dr. Chang is members of the Chinese Institute of Engineers (CIE) and the Chinese Institute of Electrical Engineers (CIEE).