IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 25, 2018, accepted August 6, 2018, date of publication August 24, 2018, date of current version October 8, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2867034

Queueing Models for Cognitive Radio Networks: A Survey

FILIP PALUNČIĆ[1], **(Member, IEEE), ATTAHIRU S. ALFA**[1, 2], **(Member, IEEE),**
B. T. MAHARAJ[1], **(Member, IEEE), AND HILARY M. TSIMBA**[1]
[1]Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa
[2]Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada

Corresponding author: Filip Palunčić (filip.paluncic@up.ac.za)

This work was supported in part by the National Research Foundation (NRF) and in part by the SENTECH Chair in Broadband Wireless Multimedia Communication, University of Pretoria.

ABSTRACT Cognitive radio networks (CRNs) are an emerging paradigm for next generation wireless communication systems allowing for more efficient radio spectrum utilization. In order to harness the full potential that CRNs may offer, many challenges and problems need to be overcome and addressed. One of the critical questions is the performance of secondary networks under primary user activity constraints. In this respect, queueing assumes a primary role in characterizing the delay, throughput and other performance metrics for secondary users, which in turn has implications for resource allocation, medium access control and quality of service provisioning. This survey presents an overview and classification of the various queueing models and techniques which have been proposed in the literature in the context of CRNs. Furthermore, open problems, future research directions and further potential applications related to queueing for CRNs are identified.

INDEX TERMS Cognitive radio networks, queueing models, queueing theory.

I. INTRODUCTION

The ever-growing demand for wireless services is leading to the overburdening of a precious, but limited, resource – the radio spectrum. Even though the available radio spectrum is a limited resource, the real culprit for this situation is the severe underutilization of the licensed spectrum [1]. In order to meet the demands of wireless communications in the near future, cognitive radio networks (CRNs) are a promising paradigm for greatly improved spectrum utilization, which leverage the power of software defined radios, whose operating parameters, such as transmission/reception frequency band, transmit power, modulation and coding schemes, etc., can be dynamically adjusted [2]. The premise of CRN is that the incumbent primary network(s) consisting of licensed users (or primary users – PUs) can coexist with secondary network(s), whereby the secondary users (SUs) opportunistically and dynamically access the spectrum in such a manner as not to interfere or degrade the performance of primary users. Therefore, primary users have priority to spectrum access and use.

While the immense benefits of CRN are self-evident, the realization of CRNs' full potential requires appropriate solutions to many non-trivial challenges and limitations. Since SUs' access and use of the spectrum is dependent on and limited by PU activity, the analysis and quantification of the secondary network's performance in terms of capacity, throughput and delay is of paramount importance. In this respect, since a SU cannot be guaranteed instant access to the network, some form of queueing modeling is necessary to reflect the realistic situation of delayed network access (queueing delay) and varying network conditions (varying service times) once access is achieved. In fact, queueing models and their analysis are the primary means by which the average throughput and delay of secondary networks are derived. Furthermore, the construction of efficient resource allocation and medium access control schemes and Quality of Service (QoS) provisioning is directly coupled to and influenced by metrics derived from queueing models. Therefore, it is fair to say that realistic evaluations of CRNs' performance need to incorporate a queueing model in one form or another.

CRNs have garnered immense interest within the academic community. This is demonstrated by the plurality of surveys related to CRNs, both older (a selection being [3]–[6])

and of recent date (a selection being [7]–[13]). Although some surveys have touched upon queueing models for cognitive radio networks in particular sections or contexts (e.g. [14, §5], where queueing models for PU activity are surveyed, or [15, §2.4], where a section is dedicated to CRN in the context of a survey on retrial queueing models), as far as we are aware, no survey in literature has been dedicated exclusively to queueing models for cognitive radio networks. In this survey, we fill this gap by providing a comprehensive overview and classification of queueing models as applied in the context of cognitive radio networks. We believe that this is timely given the fundamental role queueing theory and models play in the evaluation and description of key performance measures of cognitive radio networks, a role it is certain to continue to play in the foreseeable future. The importance of queueing theory and models as a fundamental tool in the analysis of cognitive radio networks in a plethora of scenarios is demonstrated by the exponential rise in publications on CRNs incorporating queueing models over the last decade as evidenced by the references herein. Furthermore, numerous open problems, future research directions, understudied aspects and other potential applications of queueing theory and models exist for cognitive radio networks.

The main objective of this article is to describe and classify various queueing models, both continuous-time and discrete-time, and queueing-theoretic tools applied in the analysis of cognitive radio networks. The rest of this article is organized as follows. Section II gives a short introduction to dynamic spectrum access and cognitive radio networks, while Section III gives a short overview of queueing theory. The relevant literature is surveyed in Section IV, where the primary classification is into continuous-time and discrete-time queueing models. Section V discusses open problems and possible future research directions. The conclusion is presented in Section VI.

II. DYNAMIC SPECTRUM ACCESS AND COGNITIVE RADIO NETWORKS

Dynamic spectrum access represents a paradigm shift in wireless communications and is the cornerstone of cognitive radio networks. Dynamic spectrum access entails the ability of network nodes to access the spectrum in a dynamic manner, meaning that such nodes must be capable of spectrum sensing, spectrum occupancy estimation and dynamically changing its operating parameters.

Cognitive radio networks leverage the concept of dynamic spectrum access to permit the temporal and spatial coexistence of primary networks, which represent the incumbent owners of the spectrum, and secondary networks, which access the primary network channels in an opportunistic manner. The primary network users have absolute priority in accessing and using the primary channels, while secondary network users are permitted to use these channels only when they are vacant or if the interference to the primary users is below a set threshold. The three types of spectrum sharing paradigms for cognitive radio networks are:

1) Interweave: The SUs access a channel only when the channel is not occupied by any PU. The SU traffic is ‘‘interweaved’’ with that of the PUs.
2) Underlay: The SU is permitted to transmit concurrently with a PU provided that the interference at the PU receiver due to SU transmissions is below a certain predefined threshold.
3) Overlay: The SU cooperatively relays PU transmissions in order to improve the SNR at the PU receiver, thereby allowing it to transmit its own data simultaneously.

Cognitive radio networks introduce numerous novel research challenges:

1) Spectrum Sensing: SUs must be able to perform spectrum sensing in order to determine/estimate the channel occupancy by PUs (and also other SUs). The accuracy of spectrum sensing affects the overall network performance – missed detections result in undesired PU interference, while false alarms result in actual spectrum opportunities being unutilized, affecting the secondary network throughput.
2) Spectrum Handoff: Once the channel being used by a SU is re-occupied by a PU, the secondary user needs to vacate the channel immediately. In order to complete its interrupted transmission, the SU may wait until that same channel becomes vacant again, or it may switch to another channel. Since there is overhead associated with channel switching, there is a trade-off between these two options.
3) Resource Allocation: Given the available resources (channels) at any given time, resource allocation tries to establish the optimal assignment and utilization of these resources by the SUs as a collective in order to maximize certain overall network performance parameter(s), e.g. throughput, transmission delay, etc.
4) Medium Access Control: A certain common protocol is needed to regulate how the SUs access the network.

Many of these elements are interlinked and interdependent.

III. QUEUEING THEORY

Queueing theory has its origins in the seminal work of Erlang [16] in the context of early telephone exchange networks. Since then, interest in queueing theory has been fueled by its application in diverse contexts, such as operations research, telecommunication and data networks, industrial engineering, etc. For a concise historical overview of queueing theory, see Bhat [17, §1.3].

Queueing systems consist of customers (e.g. persons, users, systems, packets, connections, etc.) that require some form of service and servers who provide the desired service. A queueing model is defined in terms of the probability distribution of the inter-arrival times of customers, the probability distribution of the service time, the number of servers and the queue capacity (if an infinite queue length is not

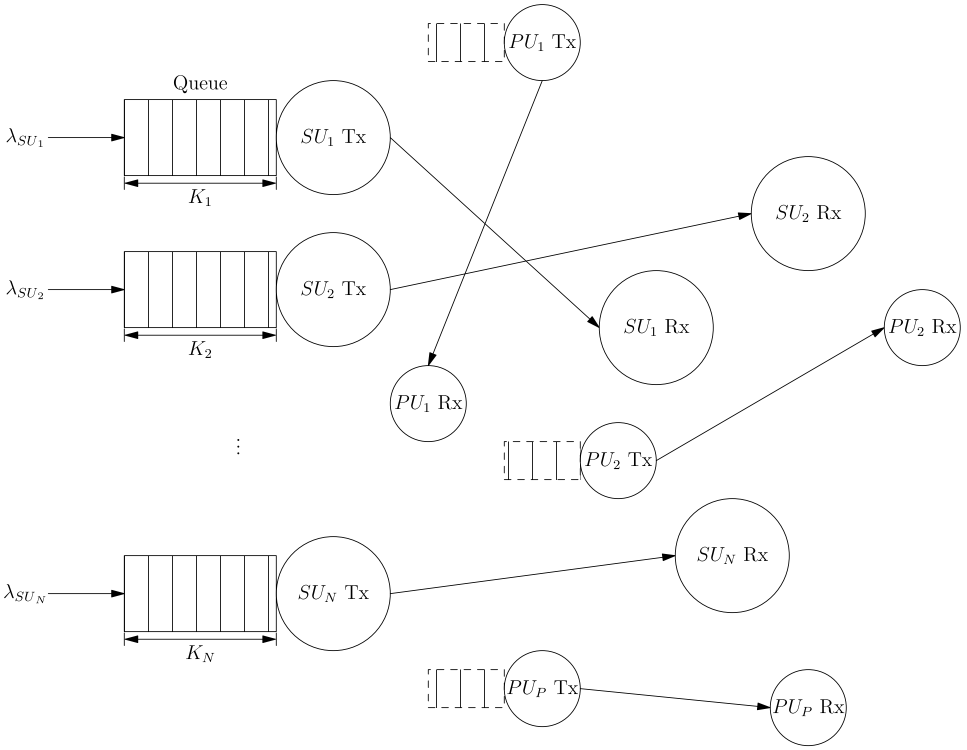

FIGURE 1. Queueing model for a decentralized SU network.

assumed). These queueing parameters are succinctly captured using Kendall's notation.[1]

Given the stochastic nature of queueing models, and the general complexity of obtaining the system characterization for any time, it is generally sufficient to obtain typical performance measures of interest under steady-state (equilibrium) conditions, provided that the system is stable. The mean values of the queueing delay time, system time (sum of queueing and service times) for a customer and the busy period (the time that a server is continuously busy) are the typical measures of interest from which other relevant measures can be derived.

Broadly speaking, queueing analysis can be divided into continuous-time and discrete-time. While continuous-time analysis allows for the characterization of the queueing system at any instant of time, discrete-time analysis gives the system state only at discrete time points. While continuous-time analysis is simpler in many cases, for digital systems which are time-slotted, discrete-time analysis is more appropriate.

Queueing theory and models have found wide applicability in telecommunications and data networks [18]–[20]. Therefore, it is no surprise that it has also found equal applicability in cognitive radio networks. The following queueing-theoretic and queueing related works are of particular significance with regard to queueing models in CRNs: [18], [20]–[29].

IV. QUEUEING MODELS FOR COGNITIVE RADIO NETWORKS

The prevalent queueing models used for cognitive radio networks are the priority queueing models, where higher-priority customers can preempt the service of lower-priority customers, and vacation models, where a server can go on ''vacation'' and cease providing service for some period. These are appropriate models for the CRN paradigm, where PUs can preempt the transmissions of SUs. The server(s) are the PU channel(s), which can be opportunistically and dynamically accessed by SUs under prescribed conditions. The customers are PUs' and SUs' data packets, sessions or connections, which are queued if they cannot obtain instant access to the required channel(s). In the most general setting, the service time of a SU customer is dependent on the channel transmission rate (which is time-varying), PU activity, data packet length, medium access control (MAC) mechanism, resource allocation scheme, number of SUs, number of PU channels, sensing errors, etc. The presence of so many inter-related and interacting factors which influence the queueing analysis for CRNs makes the extraction of relevant performance measures from such models, at least in the more generic contexts, extremely complex or even intractable. As a result, many researchers have made simplifying assumptions of one form or another: homogeneous channels, PU and/or SU homogeneity, perfect spectrum sensing, negligible spectrum sensing and channel contention time, etc. Another simplifying assumption commonly made is that of an infinite buffer size (queue capacity).

---

[1] For example, $M/G/1/K$ denotes a queueing model with exponentially distributed inter-arrival times, general distribution for service times, a single server and a queue with a maximum length of K.

Fig. 1 depicts a common queueing architecture as applicable to SU networks. Each of the N SU transmitters has a queue of size K_i (which may be infinite), $i = 1, 2, \ldots, N$, where the SU_i arrival rate (of data packets/sessions/connections) is λ_{SU_i}. While a decentralized network is depicted here, it can be easily modified for a centralized network. The servers are the M PU channels. Since various SU's data packets/sessions/connections may join the queue of any channel, the channel queues are "virtual" given that the actual queues are at the SU transmitters. This is depicted in Fig. 2, where the arrival rate λ_{CH_j} at channel j, where $j = 1, 2, \ldots, M$, encompasses the arrivals of data packets/sessions/connections from the N SUs. The service rates of the M servers (channels) are μ_j, where $j = 1, 2, \ldots, M$.

Queueing models have been proposed and analyzed for various CRN topics: resource allocation (e.g. [30]–[33]), medium access control (MAC) protocols (e.g. [34]–[36]), spectrum sensing (e.g. [37]), spectrum handoff (e.g. [38], [39]), underlay/overlay network (e.g. [40]), energy harvesting (e.g. [41]) and game-theoretic formulations (e.g. [42]–[44]), or a combination of some of the above topics (e.g. [45]).

Both continuous-time and discrete-time queueing models have been applied in the context of CRNs. This distinction is typically, though not exclusively, linked to whether or not time-slotting is assumed for the CRN. In the case of time-slotted CRN, synchronization is required between the primary and secondary network. Such a case is typically modeled using discrete-time queueing. However, there are also instances where time-slotted CRN queueing performance is approximated using continuous-time analysis [36].

In the analysis of proposed queueing models for CRNs, the two most common approaches to determining the queueing performance measures of interest are:

1) Leveraging and adapting known and existing results from queueing theory and literature with respect to the selected queueing model;
2) Using known and existing queueing-theoretic tools and techniques to analyze the proposed queueing model.

PU activity is almost ubiquitously modeled in the relevant literature using the ON-OFF process, a two-state Markov process where a PU occupies the transmission channel during the ON state and is idle during the OFF state. For an overview of PU activity models based on queueing theory, see the survey by Saleem and Rehmani [14, §5].

Before presenting an overview of the literature on queueing models for CRN, we give a few select examples of queueing models used in the context of CRN in order to give a sense of how queueing theory is applied in this field. The examples encompass both continuous- and discrete-time models. Naturally, it is not possible to go into as much detail as in these examples in the subsequent sections when describing the various contributions to queueing models in CRN.

Continuous-time queueing models consist of a continuous-time Markov chain (MC) such as $\{X(t) : t \geq 0\}$, where $X(t)$ denotes some state of interest (e.g. number of customers in the queueing system) at any time $t \geq 0$. Similarly, discrete-time queueing models consist of a discrete-time Markov chain such as $\{X(n) : n = 0, 1, \ldots\}$, where $X(n)$ denotes some state of interest at discrete time instances. The previous examples are single-state, one-dimensional Markov chains. However, in many queueing systems, multi-state, multi-dimensional Markov chains are required to accurately model the queueing dynamics. Generally, as the number of Markov chain states increases, so does the complexity of analysis.

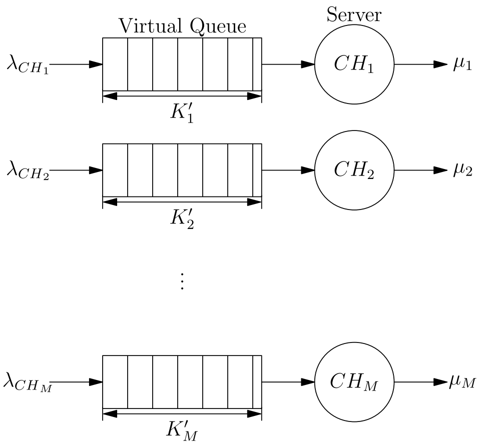

FIGURE 2. **Queueing model with *M* virtual queues corresponding to *M* PU channels.**

In what follows, we consider some specific queueing models from literature. We begin with the continuous-time queueing model by Doost-Mohammady *et al.* [33], where resource allocation in CRN with streaming and non-streaming SUs is considered (see also Sec. IV-A for a more detailed description of this work). Here, we only present the queueing model for the streaming case. In this case, the M streaming SUs receive a dedicated channel. However, since PUs can reclaim these channels at any time, a further N backup channels are needed to guarantee desired QoS. Queueing occurs if a SU's dedicated channel is occupied by a PU and all N backup channels are occupied by PUs/SUs. This scenario can be modeled by a continuous-time Markov process with a state space $\{(m(t), n(t)) : 0 \leq m(t) \leq M, 0 \leq n(t) \leq N\}$, where $m(t)$ is the number of dedicated channels occupied by PUs and $n(t)$ is the number of backup channels not occupied by PUs at an arbitrary time t. The state transition diagram is depicted in [33, Fig. 2]. Assuming a Poisson PU arrival rate λ_1 and service rate μ_1 for the M dedicated channels, and a corresponding arrival rate λ_2 and service rate μ_2 for the N backup channels, the generator matrix $\mathbf{Q}$ is

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A}_1^{(0)} & \mathbf{A}_0^{(0)} & \mathbf{0} & \cdots & & \\ \mathbf{A}_2^{(1)} & \mathbf{A}_1^{(1)} & \mathbf{A}_0^{(1)} & \mathbf{0} & \cdots & \\ & \ddots & \ddots & \ddots & & \\ \mathbf{0} & \cdots & \mathbf{A}_2^{(k)} & \mathbf{A}_1^{(k)} & \mathbf{A}_0^{(k)} & \cdots \\ & & & \ddots & \ddots & \ddots \\ \mathbf{0} & \cdots & & & \mathbf{A}_2^{(M)} & \mathbf{A}_1^{(M)} \end{bmatrix},$$

where the sub-matrices $\mathbf{A}_0^{(k)}, \mathbf{A}_1^{(k)}, \mathbf{A}_2^{(k)}, 0 \leq k \leq M$, are

$$\begin{aligned}\mathbf{A}_0^{(k)} &= (M-k)\lambda_1 \mathbf{I}_{N+1},\\ \mathbf{A}_1^{(k)} &= \mathbf{T} + \tilde{\mathbf{A}}_1^{(k)},\\ \tilde{\mathbf{A}}_1^{(k)} &= -((M-k)\lambda_1 + k\mu_1)\mathbf{I}_{N+1},\\ \mathbf{A}_2^{(k)} &= k\mu_1 \mathbf{I}_{N+1},\end{aligned}$$

where $\mathbf{I}_{N+1}$ is an $(N+1) \times (N+1)$ identity matrix and

$$\mathbf{T} = \begin{bmatrix} -N\mu_2 & N\mu_2 & & & \\ \ddots & \ddots & & & \\ & k\lambda_2 & \psi & (N-k)\mu_2 & \\ & & \ddots & \ddots & \ddots \\ & & & N\lambda_2 & -N\lambda_2 \end{bmatrix},$$

where $\psi = -k\lambda_2 - (N-k)\mu_2$. For this quasi-birth-death (QBD) process, at steady-state $\mathbf{\Pi Q} = \mathbf{0}$, where

$$\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_0 & \boldsymbol{\pi}_1 & \cdots & \boldsymbol{\pi}_M \end{bmatrix},$$

$\boldsymbol{\pi}_i = \begin{bmatrix} \pi_{i,0} \; \pi_{i,1} \cdots \pi_{i,N} \end{bmatrix}$, $0 \leq i \leq M$, is the steady-state probability vector (i.e. $\pi_{i,j}$ is the probability that in steady-state the Markov chain is in state $m(\infty) = i$ and $n(\infty) = j$). Therefore, it follows that

$$\sum_{i=0}^{M} \sum_{j=0}^{N} \pi_{i,j} = 1.$$

The equation $\mathbf{\Pi Q} = \mathbf{0}$ signifies that at steady-state the net flow rate into any given Markov state is equal to the outflow rate from that same state. By solving for $\mathbf{\Pi Q} = \mathbf{0}$ and $\sum_{i=0}^{M} \sum_{j=0}^{N} \pi_{i,j} = 1$, it is possible to obtain the steady-state probability vector $\mathbf{\Pi}$.

The next example includes a two-dimensional discrete-time Markov chain (DTMC) in the context of the proposed α-retry policy [46]. This policy states that when a SU is interrupted during transmission by an arriving PU, the SU rejoins the SU queue with probability α for later retrial or leaves the system with probability $\bar{\alpha} \triangleq 1 - \alpha$ (see [46, Fig. 1]). A time-slotted CRN is assumed and the slot boundaries are indicated by $u = 1, 2, \ldots$. The arrival of packets can occur only in the interval (u, u^+), while the departure of packets can only occur in the interval (u^-, u). The arrival and transmission intervals of PUs follow a geometric distribution with arrival rate λ_1 ($0 < \lambda_1 < 1, \bar{\lambda}_1 \triangleq 1 - \lambda_1$) and transmission rate μ_1 ($0 < \mu_1 < 1, \bar{\mu}_1 \triangleq 1 - \mu_1$), respectively (i.e. the probability of a packet arriving in any time-slot is λ_1 and the probability of a transmission completion in any time-slot is μ_1). Similarly, for SUs, the corresponding arrival rate is λ_2 ($0 < \lambda_2 < 1, \bar{\lambda}_2 \triangleq 1 - \lambda_2$) and transmission rate μ_2 ($0 < \mu_2 < 1, \bar{\mu}_2 \triangleq 1 - \mu_2$).

Assuming a finite SU queue buffer size K, let $L_u = i$ ($i = 0, 1, \ldots, K+1$) denote the number of SU and PU packets in the system at $t = u^+$ and let $R_u = j$ ($j = 0, 1$) denote whether the spectrum is occupied by a PU at $t = u^+$ ($j = 1$) or otherwise ($j = 0$). Using (L_u, R_u) as a state, a two-dimensional discrete-time Markov chain with a state space $(0,0) \cup \{(i,j) : 1 \leq i \leq K+1, j = 0, 1\}$ is obtained. The state transition probability matrix for this DTMC is

$$\mathbf{P} = \begin{bmatrix} \mathbf{C}_0 & \mathbf{B}_0 & \mathbf{A}_0 & & & \\ \mathbf{D}_1 & \mathbf{C}_1 & \mathbf{B}_1 & \mathbf{A}_1 & & \\ & \mathbf{D}_2 & \mathbf{C}_2 & \mathbf{B}_2 & \mathbf{A}_2 & \\ & & \ddots & \ddots & \ddots & \ddots \\ & & \mathbf{D}_{K-1} & \mathbf{C}_{K-1} & \mathbf{B}_{K-1} & \mathbf{A}_{K-1} \\ & & & \mathbf{D}_K & \mathbf{C}_K & \mathbf{E}_K \\ & & & & \mathbf{D}_{K+1} & \mathbf{F}_{K+1} \end{bmatrix}.$$

It is easily verified that the sub-matrices are $\mathbf{C}_0 = \begin{bmatrix} \bar{\lambda}_1 \bar{\lambda}_2 \end{bmatrix}$, $\mathbf{B}_0 = \begin{bmatrix} \bar{\lambda}_1 \lambda_2 & \lambda_1 \bar{\lambda}_2 \end{bmatrix}$, $\mathbf{A}_0 = \begin{bmatrix} 0 & \lambda_1 \lambda_2 \end{bmatrix}$, $\mathbf{D}_1 = \begin{bmatrix} \bar{\lambda}_1 \bar{\lambda}_2 \mu_2 & \bar{\lambda}_1 \bar{\lambda}_2 \mu_1 \end{bmatrix}^T$,

$$\mathbf{D}_i = \begin{bmatrix} \bar{\lambda}_1 \bar{\lambda}_2 \mu_2 & 0 \\ \bar{\lambda}_1 \bar{\lambda}_2 \mu_1 & 0 \end{bmatrix},$$

$1 < i \leq K+1$,

$$\mathbf{C}_i = \begin{bmatrix} \bar{\lambda}_1(\bar{\lambda}_2 \bar{\mu}_2 + \lambda_2 \mu_2) & \lambda_1(\bar{\lambda}_2 \mu_2 + \bar{\lambda}_2 \bar{\mu}_2 \bar{\alpha}) \\ \bar{\lambda}_1 \lambda_2 \mu_1 & \bar{\lambda}_2(\lambda_1 \mu_1 + \bar{\mu}_1) \end{bmatrix},$$

$1 \leq i \leq K$,

$$\mathbf{B}_i = \begin{bmatrix} \bar{\lambda}_1 \lambda_2 \bar{\mu}_2 & \lambda_1(\bar{\lambda}_2 \bar{\mu}_2 \alpha + \lambda_2 \mu_2 + \lambda_2 \bar{\mu}_2 \bar{\alpha}) \\ 0 & \lambda_2(\lambda_1 \mu_1 + \bar{\mu}_1) \end{bmatrix},$$

$1 \leq i \leq K-1$,

$$\mathbf{A}_i = \begin{bmatrix} 0 & \lambda_1 \lambda_2 \bar{\mu}_2 \alpha \\ 0 & 0 \end{bmatrix},$$

$1 \leq i \leq K-1$,

$$\mathbf{E}_K = \begin{bmatrix} \bar{\lambda}_1 \lambda_2 \bar{\mu}_2 & \lambda_1(\lambda_2 + \bar{\lambda}_2 \bar{\mu}_2 \alpha) \\ 0 & \lambda_2(\lambda_1 \mu_1 + \bar{\mu}_1) \end{bmatrix},$$

and

$$\mathbf{F}_{K+1} = \begin{bmatrix} \bar{\lambda}_1(\lambda_2 \mu_2 + \bar{\mu}_2) & \lambda_1 \\ \bar{\lambda}_1 \lambda_2 \mu_1 & \lambda_1 \mu_1 + \bar{\mu}_1 \end{bmatrix}.$$

If $\mathbf{\Pi} = \begin{bmatrix} \pi_{0,0} \; \pi_{1,0} \; \pi_{1,1} \; \cdots \; \pi_{K+1,0} \; \pi_{K+1,1} \end{bmatrix}$ is the steady-state probability vector where

$$\pi_{i,j} \triangleq \lim_{u \to \infty} P\{L_u = i, R_u = j\},$$

then the steady-state (equilibrium) equations are $\mathbf{\Pi P} = \mathbf{\Pi}$ subject to the normalization condition $\mathbf{\Pi 1} = 1$, where $\mathbf{1}$ is a column vector consisting of $2K+3$ 1s.

Another continuous-time example is the performance evaluation of an underlay CRN using an $M/G/1/K$ queueing model with finite queueing buffer size [40]. The model considers a single SU transmitter/receiver pair and a PU receiver, where the underlay model imposes a power constraint on the received power at the PU receiver due to the SU transmitter. The general service time of a packet is based on the instantaneous Shannon channel capacity (which is assumed to be constant during a transmission interval, but varies independently between intervals). Incorporating packet timeout, the PDF $f_{T_0}(t)$ (T_0 is a random variable representing the transmission time with timeout) of the transmission time is derived [40, eq. (11)]. Assuming a Poisson arrival process with arrival rate λ, if $\alpha(k)$ denotes the probability of k packets arriving during a service interval T_0, then

$$\alpha(k) = \int_0^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} f_{T_0}(t)\, \mathrm{d}t.$$

Consider an embedded Markov chain whose states denote the number of packets in the system immediately after service completion (see [40, Fig. 1]). If $p_{j,k}$ denotes the transition probability from state j to state k of the Markov chain, then

$$p_{0,k} = \begin{cases} \alpha(k), & 0 \leq k \leq K-2, \\ \sum_{i=K-1}^{\infty} \alpha(i), & k = K-1, \end{cases}$$

and

$$p_{j,k} = \begin{cases} \alpha(k-j+1), & j-1 \leq k \leq K-2, \\ \sum_{i=K-j}^{\infty} \alpha(i), & k = K-1, \end{cases}$$

where $1 \leq j \leq K-1$. If p_k, $k = 0, 1, \ldots, K-1$, denotes the steady-state probability that there are k packets in the system immediately after service completion, then it follows that $p_k = \sum_{j=0}^{K-1} p_j p_{j,k}$, $k = 0, 1, \ldots, K-1$, and $\sum_{k=0}^{K-1} p_k = 1$. Combining the above equations leads to the following K equations:

$$[\alpha(0) - 1]p_0 + \alpha(0)p_1 = 0$$

for $k = 0$,

$$\alpha(k)p_0 + \sum_{j=1}^{k-1} \alpha(k-j+1)p_j + [\alpha(1) - 1]p_k + \alpha(0)p_{k+1} = 0$$

for $k = 1, 2, \ldots, K-2$, and

$$p_0 + p_1 + \ldots + p_{K-1} = 1.$$

These K equations can be represented in matrix form as $\mathbf{Ap} = \mathbf{b}$, where $\mathbf{p} = \left[p_0\ p_1\ \cdots\ p_{K-1}\right]^T$ is a $K \times 1$ steady-state probability vector, $\mathbf{b} = \left[0\ \cdots\ 0\ 1\right]^T$ is a $K \times 1$ vector and

$$\mathbf{A} = \begin{bmatrix} \alpha(0)-1 & \alpha(0) & 0 & & \cdots & 0 \\ \alpha(1) & \alpha(1)-1 & \alpha(0) & & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \alpha(K-2) & \alpha(K-2) & \cdots & & \alpha(1)-1 & \alpha(0) \\ 1 & 1 & \cdots & & 1 & 1 \end{bmatrix}$$

is a $K \times K$ matrix. The steady-state probability vector is obtained as $\mathbf{p} = \mathbf{A}^{-1}\mathbf{b}$. Due to the PASTA (Poisson arrivals see time averages) property of Poisson processes [25], the steady-state probability that there are k packets in the system at an arbitrary time is

$$\bar{p}_k = \begin{cases} p_k/(p_0 + \lambda \bar{T}_0), & 0 \leq k \leq K-1, \\ 1 - 1/(p_0 + \lambda \bar{T}_0), & k = K, \end{cases}$$

where $\bar{T}_0$ is the average transmission time.

In all these examples, once the steady-state (equilibrium) probability vector is determined, various performance metrics of interest can be derived.

A few words are in order regarding the organization of this section. From the perspective of queueing theory, we have selected the primary classification as continuous-time versus discrete-time, given that both categories are well represented in the relevant literature. Within each category, literature is organized based on the following topics[2] in the given order:

1) Queueing model,
2) Resource allocation,
3) Medium access control,
4) Multi-class SUs,
5) Spectrum sensing,
6) Spectrum handoff,
7) Underlay/overlay paradigm,
8) Energy harvesting,
9) Game theory,
10) Cooperation/relaying schemes,
11) Other.

A. CONTINUOUS-TIME MODELS

A representative example of a continuous-time queueing model and its analysis is given at the beginning of Section IV.

1) QUEUEING MODEL

Chang and Jang [47] quantify the spectrum occupancy, delay and throughput for a two queue CRN (preemptive priority PU $M/M/1$ queue and a retrial SU $M/M/1$ queue), both being served by the same server (when a PU/SU is serviced, the entire radio spectrum is allocated to them). Arriving SUs' packets which find the spectrum occupied leave the serving facility (going to the "orbit"), whence it rearrives at the SU queue with a particular retrial rate. The steady-state probabilities and the average number of packets in the system and average delay are obtained based on results from [28].

Dudin *et al.* [48] propose and analyze a priority retrial queue model for CRN. Different types of PUs are assumed having different service time distributions (all types have same priority of access). There is only one type of SU and a SU can be serviced by a sub-channel/sub-server (channel/server is divided into identical sub-channels/sub-servers). PU, of any type, requires the entire channel/server. Arrivals of both PUs and SUs is modeled using a marked Markovian arrival process (MMAP). Service time distributions of PUs are modeled by phase-type distributions, while SU service times are exponentially distributed. PUs which find no server available are lost (PUs have preemptive priority over SUs), while SUs which find no available sub-channels/sub-servers move into the "orbit", from which they randomly attempt to access the system (with exponential distribution of inter-attempt times), or permanently leave the orbit due to impatience (with exponentially distributed leave time). SUs interrupted by arriving PUs also go to the orbit. Modeled as a continuous-time Markov chain, whose states include: 1) number of SUs in orbit; 2) number of PU customers in service;

---

[2] It is assumed that this represents the primary topic of a publication. Any given publication may subsume more than one of these topics.

3) number of SU customers in service; 4) state of underlying MMAP; 5) number of servers at a particular phase for PU customers. The generator matrix for the Markov chain is derived and it is shown that it belongs to class of continuous-time asymptotically quasi-Toeplitz Markov chains, which allows use of known results to derive the stationary probability vectors.

Heo *et al.* [49] analyze the performance of SU traffic in a priority preemption network. They consider the SU arrival time, the PU activity and preemption probability to mathematically model the waiting time of SU in the network. The network is set-up with a number of servers (channels) that are allocated to the SUs via a distributor. The authors also determine the blocking and forced termination probabilities of the SUs.

Liu *et al.* [50] discuss a spectrum trading model whereby PUs lease their idle spectrum to SUs. Each PU server has different service characteristics such as service state, time, content, area and price (STACP). Queues are classified based on STACP qualifiers and the SUs decide which queue to join depending on their service demands. The main performance objective of the SUs is the sojourn time.

Wang *et al.* [36] analyze the delay performance of an interweave, time-slotted CRN where the SUs employ a random access scheme.[3] For the scenarios of single and multiple PU channels, the queueing model is analyzed using a continuous-time fluid flow approximation based on Poisson driven stochastic differential equations. The derivation of moments for SU queue lengths assume a light traffic regime. Based on the queue analysis, the contention probability which minimizes queueing delay is determined. Furthermore, two packet generation control mechanisms (randomized and queue-length based) are proposed which can be adapted based on delay constraints.

Liu *et al.* [51] analyze a CRN while considering the traffic pattern of the PU. The motivation is that the generally assumed ON-OFF behavior is inadequate. Fading is assumed and a general Gaussian distribution for the channel capacity is imposed. The authors consider self-similar traffic and investigate network performance subject to fractional Brownian motion processes. Service decomposition is employed for the queueing analysis with priorities assigned to the users in the network. The PU network is given the highest priority while the SU network is the low priority network. The system is modeled using a simple single-server single-queue model.

Suliman and Lehtomäki [52] analyze the waiting time distribution of SUs in an interweave CRN. At the beginning of each time slot, SUs perform spectrum sensing of all channels to determine their occupancy/availability. The system is modeled as an $M/D/1$ queue with the service rate equal to the time slot length. The Pollaczek-Khintchine formula is exploited in the theoretical analysis which leads to the derivation of the waiting time for both the PU and SU queues. Monte Carlo simulations are performed to verify the results for the scenarios of perfect and imperfect sensing.

Usman *et al.* [53] focus on a detailed derivation and analysis of extended delivery time (EDT) for secondary packet transmissions. The major novel contributions are:

1) Detailed EDT formulation and analysis with consideration of periodic sensing. Considers scenario that secondary packet's transmission time is constant or random (depending on fading channel conditions) and spans over multiple secondary transmission slots under the work-preserving strategy (transmission resumes from point of interruption).
2) Complete and exact statistics of EDT for secondary packet transmission for both continuous and periodic sensing. The exact distribution function of EDT is derived (in distinction to previous literature which carried out the delay analysis based on moments of the delivery time).
3) Accurate queueing analysis of secondary packet transmission considering two different service time characteristics for arriving packets depending on buffer status (i.e. queue size).

The detailed EDT analysis is extended to analyze the service time of SU packets assuming a $M/G/1$ queue, where only a single channel is considered. The service times of two types of packets are analyzed: Type 1 – packet arrives while the queue is not empty, and so when the packet reaches transmission, PU is idle; Type 2 – packet arrives when the queue is empty, PU may be on or off. Closed-form expressions are derived for both types leveraging the PDF relating to EDT.

Bassoo and Khedun [54] evaluate the preemptive and non-preemptive priority queue waiting times for PUs and SUs where a PU delay is modeled (arriving PU packets wait a finite delay before attempting to access the channel). The motivation for such a delay is to prevent SU starvation due to PU activity. In non-preemptive case, a PU which finds the channel occupied by a SU cannot preempt it after its delay, whereas in the preemptive case, it can. Modeled as an $M/D/1$ queue for both PUs and SUs, being served by a single server.

Raspopovic *et al.* [55] determine the blocking probabilities of wideband (WB) and narrowband (NB) systems competing for the same spectrum. The WB users are defined as the primary users (type A) who occupy more than one NB channel when operating. NB users are defined as the secondary users (type B) and occupy only one NB channel when operating. It was found that the model closely follows a $M/M/K$ system for both user types since there is no priority assigned to any of the user types. The goal is to ensure, at least, that the blocking probability of type A users is less than that for type B users. Type A blocking probability was lower bounded by type B blocking probability and the authors found that by increasing the number of NB channels dedicated to type B users, the blocking probability for both types decreased. Another improvement is observed from limiting type A user population which reduces type A blocking probability relative to type B blocking probability.

---

[3] SUs randomly select a channel based on a uniform distribution and contend for the channel with a given probability.

Zhang *et al.* [56] determine the performance of a CRN network with a two-level queue for SUs. The aim is to improve performance by discarding SU packets that will stay in the network for too long without transmission. The CRN is in interweave mode with priority preemption resume capability. There are two parts to the queue, a delay part and a discard part. When no channel is available, SUs will join the delay part of the queue on FCFS basis. A threshold determines how many SUs are allowed in the delay queue, the rest will join the discard part. SUs will generate packets while in the queue. Packets generated by SUs in the delay queue will be saved in a buffer at the SU. When a channel becomes free, the SUs will be reconnected on a FCFS basis. Those in the discard queue will not be saved and will be lost. The authors then theoretically determine the ratio of packets generated to packets lost of an SU by utilizing a two-dimensional continuous-time Markov chain (CTMC). Performance is determined using statistical analysis and evaluated by varying the queue threshold and the idle times of the PUs and SUs in the network.

Oklander and Sidi [57] aim to model system dynamics in an interweave CRN. Matrix geometric analysis is used to determine the stationary probabilities of a CTMC. Matrix geometric analysis methods provide a fast solution to calculate the stationary probabilities from a transition matrix. Channel state estimation is also incorporated in the decision-making component. State estimation is achieved through sensing.

Jang and Chang [58] model an $M/M/1$ queue based on varying transmission rates. The channel is a Nakagami-m fading channel with Doppler shifts. The transmission rate is derived analytically using average fade duration and incorporating the Doppler shift. This transmission rate is then mapped to the exponential service rate. The general delay and throughput equations for an $M/M/1$ priority queue are derived.

Azarfar *et al.* [59] study the effect of server interruptions in single and multiple channel CRNs. The system is modeled using an $M/G/1$ queue. After a transmission interruption, the scenarios of partial transmission (resume transmission) and re-transmission are considered. Analytical probability is used to study the queue behavior and to derive queue performance parameters. In another paper, Azarfar *et al.* [60] study the effects of different queue priority disciplines on performance.

Further examples of publications focusing on continuous-time queueing models and their analysis include [61]–[78].

2) RESOURCE ALLOCATION

Doost-Mohammady *et al.* [33] consider the problem of channel allocation (with the aim of maximizing spectrum utilization) with QoS provisioning with heterogeneous PU activity and SU demands under streaming and non-streaming scenarios. Each SU that wishes to transmit sends a QoS vector to the SU-BS – QoS vector consists of SU packet arrival rate (assumed Poisson), required rate, packet delay constraint and packet length (which is assumed constant). BS groups SUs based on identical QoS vectors. For the streaming case, each SU (of the M SU streaming nodes) is assigned a channel for which it doesn't have to contend with other SUs. However, it can be preempted by a PU, so there are N backup channels per group to allow for continuous channel use. When any of the M streaming nodes is interrupted by a PU on its default channel, it will switch to a backup channel, if available. If none is available, it will queue to gain access to the spectrum (either its default/assigned channel or one of the backup channels). Hence, streaming nodes arrive randomly at the queue, which is served by N servers (corresponding to the backup channels). This scenario is modeled as a two-dimensional continuous-time Markov process, whose joint-state elements are the number of streaming nodes which are removed from their default channel by PU activity and the number of backup channels not occupied by PUs. The queueing is modeled by a quasi-birth-death (QBD) process based on PU and SU arrival and service rates, for which steady-state equations can be derived. The average queue length follows directly from the steady-state probabilities. The aim is to identify the lowest N for which the average delay is below or equal to the delay constraint.

Awoyemi *et al.* [79] consider the optimization of resource allocation with heterogeneous SUs having different time delay and QoS requirements. The secondary network is divided into concentric rings around the SU base station (BS), with rings closer to the BS providing higher data rates (based on adaptive modulation and coding). SUs are divided into DS (Delay-Sensitive) and DT (Delay-Tolerant) users. The model allows for a certain fraction θ of DT SUs' packets to change queues, i.e. from a lower-rate to a higher-rate queue. The parameter θ is selected so as to optimize the overall system throughput. The arrival rate is Poisson and the service rate is exponential at each queue. There are N sub-channels, thus each queue is $M/M/N$. Continuous-time Markov chain is used to model the queue dynamics. Based on the transition flow diagram, the balance equations are derived, whereby the steady-state probability vector is obtained using the state reduction of GTH (Grassmann, Taksar and Heyman) algorithm. Optimal value of θ is obtained using Newton's method. Blocking probability and throughput are derived from the steady-state probabilities.

Canberk *et al.* [80] develop a novel framework for QoS in an interweave heterogeneous CRN. The aim is to improve throughput and fairness among the SUs. The network is divided into different modules with the PU activity module defining the opportunity index. The SU-BS will then allocate spectrum bands to the appropriate SUs. Four different types of users are defined which are modeled by different queue types. Type 1 are E1/T1 users which are given the highest priority as their requirement is a constant bit rate. The queue is modeled as $D/G/1$. Type 2 are video conference users with second highest priority and are modeled by a $G/G/1$ queue. Type 3 are Voice-over-IP (VoIP) users modeled using a two state Markov modulated Poisson process $MMPP/G/1$ queue. Type 4 are best effort users which have the lowest priority and are modeled as an $M/G/1$ queue. Spectrum decisions are further categorized into perfect, aggressive and smooth decisions. A new metric, the request index, is defined to aid with admission control. Pseudo-code algorithms for the admission control, spectrum decision and spectrum mobility are given.

Shiang and van der Schaar [30] propose a virtual prioritized $M/G/1$ queue for heterogeneous delay-sensitive SUs and devise a dynamic strategy learning algorithm for channel selection which aims to optimize SU utility functions (end-to-end delay or throughput). PUs are top priority users by default, while the SUs are assigned to different priority classes based on the SU's delay-sensitivity and multimedia application. In the queueing model, the servers are the frequency channels and each channel has a virtual prioritized queue for PUs and SUs. These queues are virtual because they correspond to channels, whereas each SU possesses its own physical queues as destined for each channel. Therefore, this virtualization requires information exchange (via control channel) between decentralized SUs. The PU traffic[4] is modeled as $M/G/1$, as are the multiple virtual prioritized queues. The delay derivation is based on results from [21].

Jashni *et al.* [81] propose a distributed resource allocation algorithm that dynamically selects the proper channels for delay-sensitive applications in a multi-hop CRN. The key contribution of the paper is the consideration of the queueing delay in addition to the transmission delay as the routing criterion. The SUs are classified and assigned priorities. SUs of the same priority are served FCFS. The algorithm selects the path with minimum delay.

Zhao *et al.* [82] analyze the queueing delay in CRN with heterogeneous services and channels. The main criteria for heterogeneity is the delay requirement, whereby SU packets are differentiated based on their delay sensitivity. "Transmission window" scheme is proposed to prevent the starvation of SU delay-insensitive packets. The delay of PUs is analyzed using an $M/D/1$ queue.

AlQahtani [83] considers resource allocation where M2M (Machine-2-Machine) and H2H (Human-2-Human) devices co-exist in a single LTE cell with multiple resources (resource blocks). H2H devices assume the role of primary users and thus can preempt M2M communications. M2M is divided into two categories: 1) RT (Real-Time) which have a higher priority; 2) NRT (Non-Real-Time) which have lower priority. Poisson arrivals and exponential service times are assumed for all users (H2H, M2M-RT and M2M-NRT). Queueing analysis consists of continuous-time Markov chain, whose states include number of H2H, M2M-RT and M2M-NRT currently being served and size of M2M-RT and M2M-NRT queues. Balance equations are obtained based on the flow rate between MC states, from which the steady-state probability vector is obtained.

Wu *et al.* [84] consider sensing errors present in a multi-user multi-channel CRN. The impact of false and missed detection on the mean delay of the system is investigated. A vacation queueing system is employed to model the system. Upon a missed detection, both the SU and PU vacate the channel, the extra delay for PU re-transmission must be added to the total delay for the SUs. Matrix geometric solutions are used to determine the stationary probabilities and hence the expected queue delay. Given that there are many SUs, the problem of finding the optimal number of channels to allocate to SUs is transformed into a bipartite graph matching problem. The problem is then solved using the Kuhn-Munkres (KM) algorithm.

Do *et al.* [85] study the average waiting time of a multi-channel CRN in a $M/G/1$ preemptive priority scheme. The general distribution for the service time is derived through mathematical analysis. Convex optimization is employed to find the optimal method for the SU to distribute the packets over all the channels.

Resource allocation based on queueing analysis also appears in [86]–[91].

3) MEDIUM ACCESS CONTROL

Zhang *et al.* [92] analyze the optimal spectrum access strategy of SUs assuming multiple physical considerations, such as path loss, Nakagami-m fading and imperfect spectrum sensing. The CRN employs multiple channels and ARQ is implemented in the link layer. A $M/M/1/K$ queue model is employed. An optimal probability access vector is derived to minimize the average total packet delay.

Ni *et al.* [93] propose a time-slotted call admission scheme whereby SUs are only admitted at the beginning of a time slot. SUs that arrive after the beginning of the time slot have to join a waiting queue and may be admitted during the next slot. The system is modeled using an $M/M/K$ queue model and matrix-geometric methods are used to analyze the system and derive performance measures such as blocking probability and queueing delay.

Chouhan and Trivedi present a multi-channel priority based CRN architecture for MAC with three different priorities: preemptive, non-preemptive resume and pooling priority. Analytical probability is used to analyze the queue performance. For the single channel case, an $M/M/1$ queue model is utilized.

Su and Zhang [34] derive aggregate SU throughput (data rate) and transmission delay for two sensing policies by modeling the processing of SUs as a bulk-serving $M/G^Y/1$ queue for the non-saturated network case. The sensing policies are: random sensing policy – each SU randomly chooses a channel to sense; negotiation-based sensing policy – by overhearing control packets, SUs attempt to sense distinct channels. During each time slot, the SU who reserves channel use transmits over all the unused channels. The same $M/G^Y/1$ queue is also used by Zhang and Su [35] to evaluate their proposed MAC protocol, which considers channel contention, channel negotiation and sensing errors.

Feizi-Khankandi *et al.* [95] consider a SU contention-based MAC scheme model using a closed BCMP queueing

---

[4] The PUs at each channel are aggregated into a single virtual PU.

network. The contention mechanism consists of a random wait time before transmission is attempted. The scheme is time-slot based, whereby during each time-slot, SUs rescan the channels to determine their status. The queueing network consists of two nodes, an active node and an idle node. The active node includes all SUs that are under service. Since, the number of channels is fixed, N, the queueing model for the active node is $M/D/N/N$. The idle node queueing model is $M/G(n_I)/\infty$, where the general service time corresponds to random wait time and incorporates the probability of collision between two or more SUs (n_I is the number of SUs in the idle node). The service time distribution is derived by considering the available channels and the activity factor. The probability that an SU will succeed in accessing a free channel is determined and the spectrum utilization computed.

Jiang *et al.* [96] propose a MAC protocol scheme that makes use of cloud services to determine the behavior of the PUs. The SUs cooperate and report channel measurements to access points which then send the information to cloud servers. SUs are mobile and use GPS to report their positions. The information is stored in a periodically updated channel preference matrix. An analytical queueing model is used to derive the successful transmission probability, sensing time and transmission quota for each data channel.

Shankar [97] determines the capacity of static and dynamic cognitive radio networks. The paper primarily focuses on the MAC layer with the PHY layer providing information on primary user activity. Shannon capacity is exploited to determine the transmission times of the networks. Based on this and the ON-OFF behavior of the primary user, the channel utilization is determined. Comparison of spectrum utilization between static and dynamic CRNs is performed. Furthermore, the multiplexing gains and delays of the CRNs are derived. The queue is modeled as an $M/G/1$ system with preemptive priority. The paper concludes that dynamic CRNs have higher multiplexing gain and lower delay than static CRNs.

Zhu *et al.* [98] present a three-dimensional CTMC of carrier sense multiple access scheme in an unslotted multichannel CRN under non-saturation condition. The SU generates transmission files with geometrically distributed number of packets after an exponentially distributed amount of time has passed. Matrix-analytic methods are employed to solve the modeled system which is a QBD process and thereby obtain performance measures such as throughput and loss probability.

Medium access control incorporating queueing is also investigated in [99]–[102].

4) MULTI-CLASS SUs

Queueing model for multi-class SUs is investigated in [103].

5) SPECTRUM SENSING

Hoang *et al.* [37] aim to minimize the packet loss rate, where a packet loss occurs if the packet deadline delay is not met, when spectrum sensing can be dynamically scheduled within a fixed time frame period. The motivation is that CRN throughput can be maximized when sensing is performed during periods of bad channel conditions. The queueing model is considered in the case where data packets arrive according to a stochastic model (Poisson distribution). A parallel queueing model is adopted where the sensing periods are modeled as virtual sensing nodes (with virtual packets) for each channel, whose delay deadline is equal to the frame period. No queue-theoretic analysis is performed, rather known scheduling schemes for heavy traffic queues are adapted.

Queueing analysis in context of spectrum sensing is also considered in [104].

6) SPECTRUM HANDOFF

Wang *et al.* [32] use a preemptive resume priority (PRP) $M/G/1$ queue per channel with high-priority (PUs) and low-priority (SUs) queues in order to optimize the overall system time of two spectrum decision schemes (probability- and sensing-based). The model permits multiple PU interruptions of SUs, incorporates sensing errors and heterogeneous channel capacities. Once a SU selects a channel, its connection request is queued in the low-priority queue of that channel and it cannot change queues thereafter. The waiting time of both spectrum decision schemes is obtained by leveraging existing results on the PRP $M/G/1$ queue. The same queue model is used in [39] for the evaluation of channel utilization and extended delay time of reactive spectrum handoff incorporating heterogeneous arrival rates of PUs, various arrival rates of SUs and handoff processing time. However, the resulting queueing model is a queue network[5] because interrupted SUs, who perform wideband sensing on demand, can join a low-priority queue of another channel (spectrum hopping).

Wu *et al.* [105] propose a QoE (Quality-of-Experience) driven spectrum handoff scheme which maximizes the quality for prioritized multimedia users. A hybrid spectrum handoff scheme is considered – selection of target channel candidate set is performed proactively and the spectrum handoff action is performed reactively. A mixed PRP (preemptive resume priority)/NPRP (non-preemptive resume priority) $M/G/1$ queueing model is presumed, where PRP applies to PUs, while NPRP to SUs. PUs can preempt SUs, while SUs (which have different priorities) cannot preempt other SUs in service, irrespective of priority. This is to overcome the problem of overly frequent spectrum handoffs for SUs. A reinforcement learning-based QoE-driven spectrum handoff scheme is proposed which learns from previous spectrum handoffs and past channel conditions in order to maximize/optimize long term rewards rather than greedily choosing the spectrum with the maximum immediate rewards. In [106], a hybrid PRP/NPRP model is considered – based on a decision rule, a higher priority SU can preempt a lower priority SU if its service time is not less than a fixed value from completion, otherwise it cannot preempt. The residence time, completion

---

[5] Such a queue network model was originally proposed by Wang *et al.* [38].

time, expected handoff delay and expected delivery time are derived based on the Laplace transform. Multi-teacher apprentice learning is proposed where a new SU learns from neighboring experienced SUs regarding the CRN environment.

Koushik *et al.* [107] propose a "channel+beam" handoff scheme in CRN with multi-beam smart antennas based on a mixed preemptive/non-preemptive $M/G/1$ queue model. Packets in interrupted beams can be detoured through neighboring beams. A discretion rule based on the remaining service time of active SU determines whether preemption or non-preemption applies. The service times for each node are determined by deriving the SINR at each node (PU, high priority SU and low priority SU). Performance measures are determined analytically.

Zhang *et al.* [108] consider the delay analysis for proactive spectrum handoff schemes in multi-channel CRN, where the SUs are classified into two priority classes. Each channel is modeled as a PRP $M/G/1$ queue where PUs can interrupt SUs, while a higher-priority SU can interrupt a lower-priority SU if it has experienced at least one interruption. Each channel has three virtual queues: PU, higher- and lower-priority SU queues. For purposes of cumulative handoff delay for both SU classes, two handoff schemes are considered: 1) SU always stays on the same channel; 2) SU always switches channel. With simulations, it is shown that the two handoff strategies perform better under different PU traffic conditions: "always-stay" at low PU utilization and "always-change" at higher PU utilization.

Tayel *et al.* [109] consider the optimization of proactive spectrum handoff (both fixed and probabilistic sequencing) based on extended data delivery time (modeled using PRP $M/M/1$ queue) for the case of non-identical channels. Depending on the number of times spectrum handoff occurs, the average extended data delivery time for both fixed and probabilistic sequence approaches is derived. The optimization problem is to find the fixed and probabilistic sequence transition matrices which achieve the minimum extended data delivery time. The first uses a genetic algorithm for integer-valued optimization, while the second uses particle swarm for real-valued optimization.

Talat *et al.* [110] analyze a preemptive resume priority $M/G/1$ queueing network for channel hopping in CRN. A channel selection probability matrix based on channel parameters and PU behavior is derived and analytic methods are employed to determine the system time. An optimization problem that minimizes overall system time is formulated to determine the optimal channel selection probabilities.

Spectrum handoff schemes based on queueing models can also be found in [111]–[115].

7) UNDERLAY/OVERLAY PARADIGM

Sibomana *et al.* [116] qualify the SU packet queue wait time and total system time for underlay CRN in a point-to-multipoint configuration (single SU-Tx, same configuration for PU network) with varying transmission rates and various priority classes for SUs, with non-preemptive priority. Analysis incorporates packet timeout period, interference constraint to PU-Rxs and target BER. Two scheduling schemes are considered: 1) OS (Opportunistic Scheduling) – select user with best channel conditions; 2) MS (Multicast Scheduling) – group-oriented transmission based on least SINR of users. High priority classes use OS, while low priority classes use MS. A $GI/G/1$ queueing model is adopted at the SU-Tx.

Agrawal *et al.* [117] provide a queueing analysis of an underlay CR with a single SU and multiple PUs, where the SU BS, using multiple antennas with zero-forcing beamforming, results in zero (or near zero) interference at PU receivers. Closed-form expressions for the PDF of the packet transmission times assuming a quasi-static block Rayleigh fading channel are derived. The waiting time derivation is based on the Pollaczek-Khinchine formula for $M/G/1$ queue which uses the obtained first and second moments of the transmission time.

Farraj *et al.* [118] investigate queue performance measures in an underlay CRN consisting of a single PU and a single SU. The main criterion for network access is that the SU must satisfy the outage probability requirement of the PU. An $M/G/1$ queue model is found to be appropriate for the cognitive user. Shannon's theorem is used to determine the channel capacity of the SU. The mean waiting times and server utilization are analyzed using statistical methods and the effect of various outage probability requirements investigated.

Tsimba *et al.* [119] propose an increased spectrum utilization approach by enabling SUs to share capacity in an underlay CRN. Two or more SUs are allowed to simultaneously transmit with the PU present in the system. The control criteria is the interference temperature limit which is shared between the SUs. An $M/M/1$ queue with head of line processor sharing capabilities is implemented. Packets belonging to a particular SU will join that particular queue and wait for service. Two disciplines are proposed to manage the queues: 1) a preemptive discipline that will immediately adjust the transmission power to accommodate an arriving packet from another SU; 2) a non-preemptive discipline that will hold off servicing of an arriving packet until the packet being served has left the system and will only thereafter adjust the service rates. Furthermore, the SUs are weighted, with the higher weighted SUs enjoying higher transmit powers and hence better service rates.

Chu *et al.* [40] investigate an understudied topic: queueing performance for underlay CRN. They consider the scenario where communication takes place between node pairs over Nakagami-m fading channels. The queueing model is $M/G/1/K$ and is analyzed using an embedded Markov chain, where the states are the number of packets in the system at the end of a service time. The service time of a packet is derived based on Shannon's capacity of a channel and also considers packet timeout. The transition probabilities of the embedded Markov chain are combined with steady-state balance equations to determine the equilibrium probabilities for the number of packets in the system. The result is continuous-time due to the PASTA (Poisson arrivals see time averages) property.

Continuous-time queue performance measures for the underlay CRN paradigm are derived in [120].

8) GAME THEORY

Do *et al.* [42] use queueing in their game-theoretic formulation where non-cooperative SUs individually maximize their own benefit based on a probability of joining the queue at the CR base station (BS) assuming that a single PU band is available. The benefit is defined as the difference between a fixed reward and a cost proportional to the SU system sojourn time. The sojourn time, which is a function of the SUs arrival rate, is derived for a $M/M/1$ queue with service breakdowns. The equilibrium state is obtained using transition-rate diagram, from which steady-state equations are obtained and corresponding probability generating function (PGF) derived. By incorporating an admission fee at CR-BS, SUs can be forced to adopt socially optimal arrival rates. These ideas are further extended in [44] (cf. also [43]), where pricing for spectrum access control for competitive and cooperative operator duopolies are considered. Two DSA models are considered: Opportunistic-DSA (O-DSA) – SUs access a single PU channel when there is no PU activity; Dedicated-DSA (D-DSA) – spectrum is dedicated to SUs and there is no PU traffic on these bands. The O-DSA scenario is modeled using a $M/G/1$ queue with server breakdown. The average waiting time in the queue is derived using the Pollaczek-Khinchin formula.

Chang *et al.* [121] consider a non-cooperative game to determine the optimal pricing policy at BS and derive individual and social equilibrium strategies based on whether the SU decides to join or not to join the SU queue at the BS. A single BS serves both PUs and SUs, and the BS empties the SU queue whenever a PU appears and occupies the spectrum. A partially observable queue is assumed: all system and profit model parameters are known to the SU considering joining the system, except the queue length, and the SU can observe whether the BS is serving a PU or SU at the moment of arrival. Data arrival rates of PUs and SUs follow a Poisson process, and both PUs and SUs have service times with exponential distributions. The queue model states consist of two elements: 1) number of SUs in the queue; 2) whether the BS is serving a PU or SU. Based on transition rate flow between states, balance equations are obtained, from which steady-state probabilities are derived using the theory of homogeneous linear difference equations.

Guijarro *et al.* [122] develop a game theoretic approach to a priority queueing model in a CRN consisting of a primary and a secondary operator. Using an $M/M/1$ queue and FCFS discipline, the cost of transmitted packets is used as the main criteria for QoS. Both the monopoly and duopoly cases are evaluated and compared. Queueing theory is used to provide values and parameters for the determination of the game model equilibrium.

Wang *et al.* [123] analyze the equilibrium threshold strategy for CRN with SU balking and server interruptions. An observable queueing system is assumed; a tagged SU decides to join or balk depending on the number of SUs waiting in the queue. Individual and social welfare scenarios are considered. Matrix-geometric methods are employed to determine performance measures such as waiting time.

Similar to [42], Tran *et al.* [43] use game-theoretic formulations to determine revenue and social maximization pricing mechanisms for shared-use and exclusive-use dynamic spectrum access (DSA) for monopoly and duopoly operator scenarios. Operators lease spectrum (opportunities) from spectrum owners. Here, cost depends not only on delay, but also on the delay-sensitivity of the SU (hence SU heterogeneity). Shared-use – multiple SUs contend for a single PU channel and can be interrupted by PUs; modeled as $M/G/1$ queue with server breakdown; waiting time in system derived using Pollaczek-Khinchin formula; the first and second order moments of extended service time (including multiple SU handoffs due to PU activity) are derived using renewal theory. Exclusive-use – each SU is assigned a separate PU channel/band, and for the duration of leasing, there is no PU activity; modeled as $M/G/\infty$ queue.

Wang *et al.* [124] apply a retrial queueing model for CRN with a single PU and corresponding band which multiple SUs attempt to access, and determine the Nash equilibrium for the cases of non-cooperative SU medium access control which aims to maximize individual benefit and cooperative SU medium access control which aims to maximize social welfare. With regard to system observability, two scenarios are considered: 1) unobservable case – SU has no information on the system state, joins the system with probability q, otherwise balks; 2) partially observable case – SU knows whether the channel is idle or occupied; if idle, it joins with probability 1, if busy, joins with probability q. Retrial queueing model assumes that joining SUs which find the channel occupied enter the orbit, from where the SU retries to enter the system with exponentially distributed inter-retrial times. The PU and SUs arrive according to a Poisson process and the transmission times of both have general distributions. The retrial queue is modeled by a continuous-time multi-dimensional Markov process. Based on the transition rate diagram, balance equations for the system are derived which are solved using generating functions.

Safwat [125] studies the decision-making process of SUs regarding joining or balking in the presence of PUs and higher priority SUs. Perfect sensing is assumed and the lower priority SU can only transmit in the absence of both the PU and the higher priority SU. The higher priority SU is delay sensitive and will balk if the PU is present. The lower priority SU will first consider the delay of the queue through knowledge of the PU and the higher priority SU behavior and the length of its own queue before deciding whether to balk or join the queue. The base model is an $M/M/1$ queue. Analytical methods are used to determine the number of packets in the queues and to estimate the expected delay. Nash equilibrium methods are employed to determine the equilibrium decision process.

Nissel and Rupp [126] calculate the throughput in various CRN models with different priority groups. The base queueing model is an $M/M/1$ model. The analysis is extended to the case of joint spectrum sharing where an SU's packets can use another SU's spectrum provided that the second SU is idle.

Other examples of the application of game theory in the context of queueing for CRN include [127]–[134].

9) COOPERATION (RELAYING) SCHEMES

Zhang *et al.* [135] propose a cooperative-based network with preemptive resume priority. Many SUs compete for a single PU channel on a time share basis. Initially, the authors determine the performance of a CRN network where the SU does not relay PU's transmissions. An $M/G/1$ queue model is deployed with the outage probability of the SUs being incorporated in the service time distribution. The throughput is analytically determined. Then, relaying of the primary transmission via a SU through amplify-and-forward protocol is considered. The $M/G/1$ model is adapted to fit this scenario and the corresponding throughput is determined. Cooperative diversity is also analyzed and it is concluded that the secondary user improves throughput in a poor PU channel and the highest throughput gain for the SU occurs when the PU is at medium load.

Chang *et al.* [136] propose the use of redundant transmission through relay links. A SU's transmission may flow into several opportunistic paths simultaneously. Some transmissions will be duplicated. Using a time-slotted system, the authors statistically derive performance measures such as delay. Initially, the system is modeled using an $M/D/1$ queue, then it is developed into an $M/Geo/1$ system in order to incorporate the availability of a relay link. Delay is the chosen QoS requirement. Optimization is performed to maximize the number of opportunistic paths. The optimization problem is shown to be mathematically equivalent to the bin covering problem with NP-hard complexity.

Shahrasbi and Rahnavard [137] aim to maximize throughput by using rateless coding error schemes (RLC) [138]. The work assumes a Rayleigh fading model for all wireless links and a queueing model that incorporates the cooperative nature of the network. A cooperative framework where the RLC is implemented at packet level is assumed. A message is made up of a certain number of packets and the PU will generate and send those packets until it receives an ACK message from the primary receiver. The secondary user is assumed to be able to receive and decode the PU transmission. The PU receiver will then use the combination of packets received from the primary transmitter and those from the secondary transmitter to decode the originally transmitted message from the PU. Queueing theory is used to determine the mean number of packets required to be transmitted by the PU before the SU can aid in decoding the message.

A cooperative scheme with a buffered relay is also investigated in [139].

10) OTHER

Balieiro *et al.* [140] investigate the collision probability of PUs with SUs in a wireless virtualization scenario, where multiple PVNs (Primary Virtual Networks) coexist with multiple SVNs (Secondary Virtual Networks) using the same set of channels. Assuming Poisson arrivals and exponential service times for both PUs and SUs, the collision probability is derived based on the steady-state probabilities of a two-dimensional Markov chain.

Da Silva and Brito [141] determine optimum switching points for an adaptive modulation scheme assuming Nakagami-m fading. A priority $M/G/1$ queue model is adopted and both unlimited and limited retransmission scenarios are analyzed.

B. DISCRETE-TIME MODELS

A representative example of a discrete-time queueing model and its analysis is given at the beginning of Section IV.

1) QUEUEING MODEL

Rashid *et al.* [31] derive various performance measures of a queueing model for a centralized interweave time-slotted network. The channel gains vary with time and the opportunistic spectrum scheduling scheme assigns an idle channel to a SU with the highest transmission rate for that channel. The bursty nature of packet arrivals at SUs is modeled as a batch Bernoulli arrival process and the SUs have finite buffer sizes. The queueing model is a quasi-birth-death (QBD) process based on finite state Markov chain consisting of joint states of queue length and throughput at a given time slot. The steady-state probability vector for the QBD process is obtained using the matrix-analytic method [27].

Wang *et al.* [45] consider queueing analytics for multi-SU and multichannel time-slotted interweave CRN incorporating: sensing errors; contention-based and contention-free MAC; link adaptation techniques – AMC (Adaptive Modulation and Coding) and ARQ; finite buffer size. The channel condition changes each time slot and is modeled by a Markov chain whose states correspond to different modulation and coding schemes (MCS) based on the required average packet error rate at a given SNR. The queue model is $G/G/1/K$ which is modeled as finite state Markov chain (FSMC) consisting of joint states combining sensing results, medium access results, capacity (transmission rate) based on MCS selection for each PU channel and the SU queue length. Based on the steady-state probabilities, the average queue length, packet-drop and packet-collision rates are derived.

Adem and Hamdaoui [142] provide closed-form expressions for the SU waiting delay and service delay where the availability of PU channels is modeled by a two state Markov chain. The waiting delay is derived using the residual time concept. Based on a $Geo/G/1$ queue model, statistical methods are used to derive desired performance measures. Using probability generating functions, Zaman *et al.* [143] derive closed-form expressions for various performance parameters, including residual service.

Jin *et al.* [144] consider a CRN spectrum access mechanism taking into account SU impatience and imperfect sensing. A partially observable priority queue model is assumed. SUs operate under an admission threshold; when reached SUs will be blocked from entering the network. Matrix-geometric methods are employed to derive the stationary distribution.

Jeon *et al.* [145] consider the stability region of a single primary and secondary source-destination pair assuming the capture effect, whereby a SU may transmit concurrently with a PU and successful decoding can be performed if the received SINR is sufficiently high. If the PU is sensed to be idle, SU transmits with probability 1 if its queue is non-empty, otherwise, if the PU is active, the SU transmits with probability p. Under imperfect sensing by SU, the aim is to determine the optimum access probability by the SU when the PU is active to maximize its own stable throughput but which ensures PU stability at a given input rate demand. The stability is defined as the queueing stability, which ensures that a steady state is reached and the queue doesn't grow indefinitely. Formally, queueing stability is defined so that the probability that the queue size is less than some value at some time has a proper limiting cumulative probability distribution as time tends to infinity. Infinite queue size is assumed. The arrivals and services are modeled as stochastic processes which are sequences of binary random variables (the arrival process is modeled as a Bernoulli process). The service rates of the PU and SU depend on the queue size of the other, hence the queues are interacting, which makes their analysis challenging. To decouple this interaction, stochastic dominance technique is employed, where the SU/PU is considered to send dummy packets when its queue is empty. Loynes' theorem is applied for queueing stability, which states that a queue is stable if its arrival rate is less than or equal to its service rate. In the case where the SU transmits dummy packets, the service rate of PU is not dependent on the SU queue size, and assuming a discrete-time $Geo/Geo/1$ queue model at the PU, gives the probability that the PU queue is empty, from which the SU service rate can be obtained.

Zhao *et al.* [46] propose a spectrum access strategy based on the so-called α-Retry Policy. This model is proposed as a compromise between two extremes: 1) whenever a SU is interrupted by a PU packet, the interrupted SU packet returns to SU queue with probability 1, which causes greater delay for SU packets; 2) all interrupted SU packets leave the system permanently when there are no idle channels available, which causes the packet drop rate to increase. With the α-Retry Policy, an interrupted SU packet returns to the queue with retrial probability α and leaves the system with probability $1 - \alpha$. Assuming a finite buffer size and a geometric distribution for PU and SU arrivals, the queue is modeled using a two-dimensional discrete-time Markov Chain (DTMC), whose states consist of: 1) total number of packets, including PUs and SUs, in the system; 2) whether the spectrum is occupied by PU packet. The transition probability matrix using a block-structure is derived, which is used for equilibrium equations. The steady-state probability vector is obtained using Gaussian elimination. The retrial probability is maximized. This is possible because an increase in retrial probability results in higher throughput, but also in longer delays. Optimization is based on a cost function which incorporates both throughput and delay.

Luís *et al.* [146] derive the characteristic function (Fourier transform of probability distribution function) of SU service time under general conditions for a single PU and single SU, where the SU service time is variable and may be interrupted by the PU multiple times. It encompasses both SU saturation and non-saturation cases. Thereafter, based on simulations, it is discovered that the probability mass function is closest to a generalized Pareto distribution. The derived characteristic function is then used to parameterize the generalized Pareto distribution, thereby obtaining an approximation of the probability mass function of the SU service time. The derivation of the characteristic function of the SU service time depends on the probability of the SU's queue being empty. The derived characteristic function is applied to an $M/G/1$ queueing model, where known results are used in conjunction with the expected service time derived using the characteristic function to setup a set of non-linear equations, from which the probability of SU's queue being empty and the expected service time can be obtained. The approach represents a novel way of obtaining important statistical information in a general setting.

Further examples of publications whose primary focus is on queueing models for CRN and their analysis are [147]–[151].

2) RESOURCE ALLOCATION

A distributed algorithm for scheduling and resource allocation in decentralized, time-slotted cognitive wireless mesh networks (allowing for multiple intermediary nodes between source and destination) with the objective of maximizing the network's throughput under delay constraint is presented by El-Sherif and Mohamed [152]. Multiple data streams (source-destination routes) may exist. Packet arrivals at the source node of a data stream is modeled as a Bernoulli process. The probability that a packet from a stream enters the queue of some node in the stream is based on:

1) That a resource element (at time t, channel c) is allocated to one of the node's incoming edges;
2) The primary node owning channel c is either idle or the cognitive node is outside the primary node's interference range;
3) The preceding cognitive node in the route has at least one packet in its queue to transmit.

As a simplification, the average arrival and service rate at a node is calculated as the time average over a time frame. Each node's queue, based on the average arrival and service rates, is modeled as a discrete-time $Geo/Geo/1$ queue. In essence, the overall model is a queueing network, however the queueing analysis is performed on a per node basis. The resource allocation aims to maximize the sum of utility functions, which are a function of the minimum service rate along the data stream, for all data streams. The problem is formulated as an integer programming optimization problem, where the decision variables are relaxed to allow real values and an algorithm is developed for resource allocation based on the real-valued solution. Optimization is based on dual decomposition and Lagrangian multipliers. However, to perform the optimization, specific information exchange between nodes in range needs to occur.

For single/multiple SU(s) and multiple PU channels, Cao *et al.* [153] consider the problem of optimal channel selection policy (in the sense of minimizing the average queue delay) for dynamic load-balancing given that SU(s) have two service classes: 1) DS – delay sensitive packets, which have priority over 2) BE – best effort packets. A time-slotted CRN is assumed with heterogeneous channel characteristics (block fading) across time slots. The queue dynamics are modeled using discrete-time Markov chains (DTMC). Three DTMCs are developed: 1) primary user activity with single component state (PU queue length); 2) SU DS service with two component states (DS queue length and occupancy state of each channel); 3) SU BE service with three component states (BE and DS queue lengths and occupancy state of each channel). The steady-state probability vector is obtained using the matrix-analytic approach. Based on DTMC modeling of the queueing process, a Markov Decision Process (MDP) is formulated to obtain the optimal selection policy. For the case where traffic and channel state information is unknown, reinforcement learning is proposed to find the optimal policy.

Further examples of resource allocation which take discrete-time queueing into account include [154]–[156].

3) MEDIUM ACCESS CONTROL

Azarfar *et al.* [157] provide a detailed delay analysis of homogeneous multi-user multi-channel ad hoc CRNs for two baseline MAC protocols: 1) buffering policy – SU stays on the same channel; 2) switching policy – SU switches to a new channel upon PU appearance. These two policies form the extreme boundaries, with other possible protocols being some form of mixture of the two. The SUs compete for channel access by transmitting a reservation request over the control channel with some probability. A geometric arrival process is assumed for both PUs and SUs, the service times also assume a geometric distribution. For both buffering and switching MAC protocols, two composite-state Markov chains are constructed for occupancy (state consisting of number of packets in system for each node and whether a channel is reserved or not by the nodes) and busy/idle status (state consisting of number of busy nodes and number of idle nodes with non-empty queue). The transition probabilities for MCs and delay time (consisting of channel reservation and transmission time) are derived.

Medium access based on queueing analysis is also the topic of [158].

4) SPECTRUM SENSING

Spectrum sensing with queueing analysis is the primary topic in [159].

5) SPECTRUM HANDOFF

Lee *et al.* [160] investigate optimal channel-hopping sequence (OCS) which maximizes the aggregate throughput of SUs while maintaining PU QoS requirement using dynamic programming. Imperfect sensing and synchronization is assumed. In a multi-channel CRN, each channel is modeled as a $Geo/G/1$ queue. An analysis on the probability of channel availability and frame delay for PU is conducted.

Spectrum handoff incorporating discrete-time queueing analysis also appears in [161].

6) UNDERLAY/OVERLAY PARADIGM

Cooperation in an underlay CRN is investigated in [162].

7) ENERGY HARVESTING

Zhang *et al.* [163] consider the problem of aggregate network utility optimization for energy harvesting CR sensor networks. Sensor network consists of multiple energy harvesting sensors which use multiple licensed channels opportunistically to send sensed data to a single sink. Three queues are considered: 1) sensor data queue; 2) energy harvested queue; 3) a virtual collision queue – counts number of collisions that occur due to imperfect spectrum sensing; the queue is included to analyze acceptable interference to PUs due to collisions. Lyapunov optimization is employed to decompose the optimization problem into three sub-problems: 1) battery management; 2) sampling rate control; 3) channel and data rate allocation.

Amer *et al.* [164] propose an energy harvesting model for both PUs and SUs in a cluster based topology. Multiple SUs are grouped in a cluster, which are managed by a single cluster supervision block controller. The cluster controller manages both the packet and harvested energy queues. The stability of the throughput region is determined through two dominant systems: 1) SU cluster transmits dummy packets when any queue is empty; 2) SU transmits dummy packets whenever the relay queue is empty. The performance for both cooperative and non-cooperative systems is studied. Abd-Elmagid *et al.* [165] also study the stability region of a CRN with energy harvesting capabilities. The problem is relaxed by focusing on two cases: 1) finite battery queue with infinite relay queue; 2) finite relay queue with infinite battery queue. The optimization problem seeking to maximize the SU service rate, which is non-convex, is linearized and solved using standard linear programming techniques.

El Shafie and Sultan [41] consider the scenario of a single energy harvesting SU using random access over a single PU channel. Both the PU and the SU have data queues,[6] where the arrivals of data packets are modeled using Bernoulli

---

[6] The SU also has an energy queue.

random variables. The primary receiver sends an ACK/NACK using a feedback channel and the possibility of multiple-packet reception is assumed whereby concurrently transmitted packets can be decoded if the SINR exceeds a threshold. For the primary feedback-based access scheme, the PU's queue is analyzed using a Markov chain whose states represent the number of packets in the system depending on whether it's a first transmission or a retransmission. The solution of the state balance equations gives the steady-state probabilities of Markov chain states. In order to decouple interactions between queues, it is assumed that the SU sends dummy packets when its queue is empty. Given a constraint on the PU mean queue delay, the optimal value of the random-access probability that maximizes SU service rate is determined.

Lu *et al.* [166] investigate random access in a CRN network where SUs harvest energy from PU RF radiation. The focus is on SU throughput analysis with queueing theory employed to analyze the SU delay. Furthermore, an optimization problem is formulated to maximize SU throughput under the constraint of PU queue stability. Statistical analysis is employed to derive performance measures. Niyato *et al.* [167] study a similar RF energy harvesting problem, but use matrix-geometric methods to analyze the queue performance.

Nobar *et al.* [168] consider energy harvesting cognitive radio sensor networks, which contain a dedicated power beacon (PB) which harvests ambient energy (solar, wind, RF, etc.) and wirelessly transmits this energy to cognitive sensor nodes (CSN) which contain an RF energy harvesting unit. Energy harvesting at PB is modeled by an energy queue – energy is harvested into a battery when energy is not being transmitted to CSN. Energy harvesting in a time slot is exponentially distributed, one energy packet is harvested if the energy is greater than the PB transmit power. The number of energy packets in PB's battery is modeled by a Markov chain. Similarly, CSNs contain an energy queue, modeled as a $Geo/Geo/1$ queue. Optimization problem is to maximize CSN service rate, which is a function of the probability that the PB chooses energy transmission mode and the probability that the CSN chooses transmission mode, whilst maintaining the PU's QoS.

Energy harvesting in CRN with discrete-time queueing analysis is also considered in [169].

8) GAME THEORY

An energy-saving strategy for centralized CRN with LTE-Advanced structure is considered by Jin *et al.* [170]. Energy-saving is implemented through the introduction of sleeping and listening modes which the base station (BS) enters when PUs and SUs have no data packets to transmit. The BS exits the sleep mode if a PU packet arrives or the sleep timer expires, while the BS exits the listening mode (which follows the sleep mode under certain conditions) if a SU data packet arrives. The listening mode is intermediate between sleep and transmission modes in terms of energy consumption. Queueing is modeled using a two-dimensional DTMC (Discrete-Time Markov Chain) whose composite states consists of: 1) number of SU data packets in the system; 2) BS mode - sleep mode, PU transmit, SU transmit or listening mode. Geometric distribution is assumed for both arrival and service rates of PUs and SUs. Steady-state probabilities are obtained using a matrix geometric solution. Based on a utility function consisting of a reward and cost (based on SU waiting time), the Nash equilibrium is investigated, which leads to a pricing policy for spectrum access of SUs which encourages SUs to adopt a socially optimal arrival rate.

Other examples of publications based on game theory include [171], [172].

9) COOPERATION (RELAYING) SCHEMES

Homayounzadeh and Mahdavi [173] propose an efficient cooperative transmission strategy for SUs where the quality of the channels is considered in the channel allocation and relay selection strategies. Primary channel occupancy is modeled by an ON-OFF process, while SU aggregate traffic is modeled by a discrete-time batch Markovian arrival process. The work also takes into account sensing errors. The probability that packets will be transmitted over a particular channel during a particular time slot is determined. Matrix-geometric methods are employed to find the performance measures.

Feng *et al.* [174] consider a cooperative spectrum leasing system whereby PUs lease their licensed spectrum to SUs who act as relay nodes for that PU in order to minimize its transmit power and improve its transmit rate. SUs provide cooperative assistance in order to gain access to the licensed spectrum. SU aims to minimize its own transmit power during cooperation. A Distributed Win-Win Reciprocal-Selection-based Medium Access Scheme (DWWRS-MAS) is developed to distributively select the best cooperative pairs for the leasing system. Matching theory is used to analyze the algorithmic stability of DWWRS-MAS, while queueing theory is used to analyze the queueing stability (Loynes' theorem).

El Shafie *et al.* [175] consider the queueing performance of using N relays in a single PU and single SU communication pair network. Based on the PU/SU ARQ feedback messages, the relay nodes can determine if they should attempt to decode received PU/SU packets and, if successfully decoded, they accept the packet for retransmission with some probability. There are three relay decoding strategies:

1) Ordered acceptance strategy: The N relay nodes are ranked and they attempt decoding of a packet with a NACK in sequential order based on their rank.
2) Random assignment decoding strategy: Only a single relay node during a time slot attempts to decode a received packet – the relay node is selected based on a random distribution.
3) Round robin decoding strategy: Again, only a single relay node attempts decoding – relay nodes are selected on a round robin basis.

The PU and SU source nodes have $Geo/Geo/1$ queues. Based on the decoding strategies, the service rates at the PU and

SU source node queues are derived using probabilistic arguments. It is proven that ordered decoding is guaranteed to have greater service rates than the other two strategies.

Elmahdy *et al.* [176] perform optimization for SU (throughput and average delay) under PU QoS constraints for cooperative CRN with relaying. Model assumes a single PU (with a single queue) and a single SU (with two queues – SU packets and successfully decoded PU packets) communicating to a common destination. PU packets which are not acknowledged by the destination, are admitted to the PU queue at SU with some probability if the SU successfully decoded the packet. Based on a probability value, the SU transmits from its own queue or the PU queue. Therefore, the incentive for the SU to act as a relay is greater access to the licensed spectrum. The queue at the PU is described as an $Geo/Geo/1$ queue. The service rates of PU and SU packets are derived using a probabilistic analysis based on the described relaying scheme.

Kiwan *et al.* [177] characterize the stability region of a dedicated relay system under two different MAC schemes: perfect sensing and random access. Queues are implemented at the PU, SU and relay node. A collision channel model is adopted which assumes that if two packets arrive at a receiver at the same time, then both are lost. A flat block fading channel is also assumed. The service times for the PU and SU for both schemes as well as the service times for the relay queue are derived.

Cooperative schemes based on queueing theory are also investigated in [178]–[186].

10) OTHER

Jang and Chang [187] consider the cross-layer design of CRN that combines MIMO AMC (Adaptive Modulation and Coding) and adaptive link layer over Nakagami-m fading channels with finite queues at the primary and cognitive transmitters. A single PU channel, occupied by a single PU, is contested by multiple SUs. The channel experiences flat fading with Nakagami-m that varies from frame to frame. The AMC mode can only change by a single step from frame to frame. A two-class priority queue is modeled where the PU has higher priority. Poisson arrivals are assumed for both PU and SUs. The number of packets transmitted per frame time is dependent on transmission mode (AMC). Queueing dynamics are modeled using FSMC (Finite State Markov Chain) with composite states consisting of the queue size and channel conditions at discrete times. The transition probabilities are derived, and based on the stationary distribution (steady-state), various performance measures are obtained: spectral occupancy rate, the average packet drop rate and average spectral efficiency.

Rahimzadeh and Ashtiani [188] consider the use of secondary WLAN with a primary OFDMA based network in TDD mode. Empty RBs (Resource Blocks – consisting of a number of sub-carriers [frequency] and OFDMA symbols [time]) are used opportunistically by SUs. OFDMA frames consist of downlink (DL) and uplink (UL) sub-frames, both of which are utilized by SUs when idle. For the primary network, DL and UL sub-frame occupations are modeled separately using Markov chains (state – number of packets in DL/UL queue). Transition probability matrix for MCs are derived. The two MCs for DL and UL are combined into a single two-dimensional MC which models the occupancy of RBs by PUs. The analytical approach for the secondary network is to use two queueing networks (QN). The first QN represents the packet transmission process based on the IEEE 802.11 MAC protocol, with each node representing RTS/CTS/TR/ACK stages. Embedded within the TR node is another QN which encompasses the transmission time based on the two-dimensional MC for the primary network.

V. OPEN PROBLEMS AND FUTURE DIRECTIONS RELATED TO QUEUEING MODELS IN CRN

In this section, we highlight some of the open problems and possible future directions for research with regard to queueing models for CRN. It is not a comprehensive list, it includes those topics we deem most worthy of further investigation.

Despite the sheer volume of literature dedicated to queueing models for CRN, it is still possible to identify numerous open problems and future directions related to this topic. Realistic cognitive radio network models are complex due to multifarious factors, often interacting and interdependent. Individually, these factors can lead to complicated queueing models which are difficult to analyze. By considering many of these factors concurrently, the problems can be extremely complex and even intractable. Therefore, it is understandable that many works adopt simplifying assumptions in order to render the resulting problem manageable and/or tractable. All the literature cited in this survey made simplifying assumptions of one sort or another, to a greater or lesser extent. It is fair to state that the works with the greatest contribution are those with the most general models and assumptions. This is because such models approach real-world scenarios to a greater degree and their applicability covers a greater range. The extensions of models in terms of their scope and generality, and the corresponding application of queueing models, still provide ample scope for research.

As already indicated, PU channel occupancy is almost universally modeled by the ON-OFF process, where the ON (busy) and OFF (idle) times are uncorrelated. As indicated in [189], citing works based on empirical evidence, more realistic models of PU activity require correlation between consecutive ON (busy) periods and/or correlation between ON (busy) and OFF (idle) periods. Investigation of the effects of such alternative PU activity models on queueing models for CRN is a desideratum.

Although the overlay and underlay paradigms for CRN introduce their own set of challenges, they provide greater capacity for the secondary network as compared to the interweave paradigm [190, Ch. 2]. Yet, the application of queueing models for the overlay/underlay scenario is underrepresented in the literature, with the vast majority focusing on the interweave paradigm. This is undoubtedly due to the greater complexity of the queue modeling and analysis for the overlay/underlay case. Nevertheless, the capacity gain of the overlay/underlay paradigms warrant further investigation of queueing models for these particular types of CRN.

Queueing networks have been sparsely considered for cognitive radio networks, primarily in the context of spectrum handoff schemes. This is because the majority of the relevant literature assume direct communication links, such as between cognitive nodes and the secondary base station or between a pair of cognitive nodes, and therefore a single queue (with possibly multiple servers corresponding to multiple channels) suffices. However, queueing networks, which are more complex to analyze, are a more appropriate model when the communication links consist of multiple intermediary nodes between the source and destination, each having its own queue. This topic has been ignored in the literature, but may be particularly appropriate for ad hoc CRNs.

VI. CONCLUSION

In this paper, we have surveyed the literature on queueing models for cognitive radio networks. Queueing theory and models assume a fundamental role in the performance evaluation of CRNs. Queueing models are necessary due to the nature of CRNs: service interruptions of secondary users (as in the interweave paradigm) or varying service rates (as in the case of the overlay/underlay paradigm). Key performance indicators such as throughput, average customer waiting time, system utilization, etc. are derived using queueing-theoretic techniques applied to the selected queueing model.

Priority based queueing models are the most frequently adopted models for CRNs as they can capture the different priority requirements of the PUs versus the SUs. Furthermore, such models have also been applied to scenarios where the SUs are differentiated into separate classes, such as SUs which have stringent delay constraints (real-time) versus SUs with higher delay constraints (non-real-time). Retrial queues have also been considered for CRNs. In this case, customers in the orbit attempt to access/re-access the system in a random order, rather than, for example, the order in which they arrived at the queue. The selection of the queueing model needs to be aligned with the CRN scenario under consideration. Queueing models can be broadly classified into continuous-time and discrete-time models. Various techniques and methods are utilized to derive the steady-state probability vector, from which numerous performance measures of interest can be obtained.

Queueing models for CRNs have been applied in the context of numerous topics, including resource allocation, MAC protocols, admission control, spectrum handoff, energy harvesting, wireless sensor networks, ad hoc networks, etc. In many cases, optimization problems are formulated which utilize parameters which are obtained from the queueing model analysis. Hence, queueing model analysis plays a crucial role in various optimizations occurring in diverse scenarios.

As already highlighted, although a voluminous literature on queueing models for CRNs exists, numerous open problems and future research directions can be identified. This includes, but is not limited to, more realistic modeling of PU activity, more general modeling of CRN with fewer simplifying assumptions, further analysis of queueing models for CRN underlay/overlay paradigms, extension of queueing stability region analysis, networking queues for CRNs, network capacity combining information theory and queueing theory, simulator development for various scenarios as a real-time test-bed for different circumstances and traffic types, etc. The last decade has seen an exponential increase in the number of publications on queueing models for CRNs. The crucial role that queueing models play in the performance evaluation of CRNs, combined with the existence of a multitude of open problems and research directions relating to this topic, permits us to predict that this trend in the publication rate on queueing models for CRNs is going to continue for the foreseeable future.

REFERENCES

[1] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[2] J. Mitola, III, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, KTH Royal Inst. Technol., Stockholm, Sweden, 2000.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.*, vol. 50, pp. 2127–2159, Sep. 2006.

[4] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[5] D. Niyato and E. Hossain, "Medium access control protocols for dynamic spectrum access in cognitive radio networks: A survey," in *Cognitive Radio Networks*, Y. Xiao and F. Hu, Eds. Boca Raton, FL, USA: CRC Press, 2009, ch. 7, pp. 179–214.

[6] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, 1st Quart., 2009.

[7] A. S. Alfa, B. T. Maharaj, S. Lall, and S. Pal, "Mixed-integer programming based techniques for resource allocation in underlay cognitive radio networks: A survey," *J. Commun. Netw.*, vol. 18, no. 5, pp. 744–761, Oct. 2016.

[8] M. El Tanab and W. Hamouda, "Resource allocation for underlay cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1249–1276, 2nd Quart., 2017.

[9] A. Ali and W. Hamouda, "Advances on spectrum sensing for cognitive radio networks: Theory and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1277–1304, 2nd Quart., 2016.

[10] T. Elderini, N. Kaabouch, and H. Reyes, "Channel quality estimation metrics in cognitive radio networks: A survey," *IET Commun.*, vol. 11, no. 8, pp. 1173–1179, Jun. 2017.

[11] M. Amjad, F. Akhtar, M. H. Rehmani, M. Reisslein, and T. Umer, "Full-duplex communication in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2158–2191, 4th Quart., 2017.

[12] M. R. Hassan, G. C. Karmakar, J. Kamruzzaman, and B. Srinivasan, "Exclusive use spectrum access trading models in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2192–2231, 4th Quart., 2017.

[13] A. Naeem, M. H. Rehmani, Y. Saleem, I. Rashid, and N. Crespi, "Network coding in cognitive radio networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1945–1973, 3rd Quart., 2017.

[14] Y. Saleem and M. H. Rehmani, "Primary radio user activity models for cognitive radio networks: A survey," *J. Netw. Comput. Appl.*, vol. 43, pp. 1–16, Aug. 2014.

[15] T. Phung-Duc, "Retrial queueing models: A survey on theory and applications," in *Stochastic Operations Research in Business and Industry*, T. Dohi, K. Ano, and S. Kasahara, Eds. Singapore: World Scientific, 2017. [Online]. Available: http://infoshako.sk.tsukuba.ac.jp/~tuan/papers/Tuan_chapter_ver3.pdf

[16] A. K. Erlang, "The theory of probabilities and telephone conversations," *Nyt Tidsskrift Mat.*, vol. 20, no. B, pp. 33–39, 1909.

[17] U. N. Bhat, *An Introduction to Queueing Theory: Modeling and Analysis in Applications*. Boston, MA, USA: Birkhäuser, 2008.

[18] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1992.

[19] G. Giambene, *Queuing Theory and Telecommunications: Networks and Applications*, 2nd ed. New York, NY, USA: Springer, 2014.

[20] A. S. Alfa, *Applied Discrete-Time Queues*, 2nd ed. New York, NY, USA: Springer, 2016.

[21] L. Kleinrock, *Queueing Systems: Theory*, vol. 1. New York, NY, USA: Wiley, 1975.

[22] L. Kleinrock, *Queueing Systems: Computer Applications*, vol. 2. New York, NY, USA: Wiley, 1976.

[23] M. L. Chaudhry and J. G. C. Templeton, *A First Course in Bulk Queues*. New York, NY, USA: Wiley, 1983.

[24] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation: Vacation and Priority Systems*, vol. 1. Amsterdam, The Netherlands: North Holland, 1991.

[25] H. Takagi, *Queueing Analysis: Finite Systems: A Foundation of Performance Evaluation*, vol. 2. Amsterdam, The Netherlands: North Holland, 1993.

[26] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation: Discrete-Time Systems*, vol. 3. Amsterdam, The Netherlands: North Holland, 1993.

[27] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. New York, NY, USA: Dover, 1995.

[28] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*, 4th ed. Hoboken, NJ, USA: Wiley, 2008.

[29] C.-H. Ng and B.-H. Soong, *Queueing Modelling Fundamentals: With Applications in Communication Networks*, 2nd ed. Chichester, U.K.: Wiley, 2008.

[30] 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.

[31] M. Rashid, M. Hossain, E. Hossain, and V. Bhargava, "Opportunistic spectrum scheduling for multiuser cognitive radio: A queueing analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5259–5269, Oct. 2009.

[32] 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.

[33] R. Doost-Mohammady, M. Y. Naderi, and K. R. Chowdhury, "Spectrum allocation and QoS provisioning framework for cognitive radio with heterogeneous service classes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3938–3950, Jul. 2014.

[34] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.

[35] X. Zhang and H. Su, "CREAM-MAC: Cognitive radio-enabled multichannel MAC protocol over dynamic spectrum access networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 110–123, Feb. 2011.

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

[37] A. T. Hoang, Y.-C. Liang, and Y. Zeng, "Adaptive joint scheduling of spectrum sensing and data transmission in cognitive radio networks," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 235–246, Jan. 2010.

[38] 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.*, vol. 18, no. 6, pp. 18–26, Dec. 2011.

[39] C.-W. Wang and L.-C. Wang, "Analysis of reactive spectrum handoff in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 2016–2028, Nov. 2012.

[40] T. M. C. Chu, H. Phan, and H.-J. Zepernick, "On the performance of underlay cognitive radio networks using M/G/1/K queueing model," *IEEE Commun. Lett.*, vol. 17, no. 5, pp. 876–879, May 2013.

[41] A. El Shafie and A. Sultan, "Optimal random access for a cognitive radio terminal with energy harvesting capability," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1128–1131, Jun. 2013.

[42] C. T. Do, N. H. Tran, M. V. Nguyen, C. S. Hong, and S. Lee, "Social optimization strategy in unobserved queueing systems in cognitive radio networks," *IEEE Commun. Lett.*, vol. 16, no. 12, pp. 1944–1947, Dec. 2012.

[43] N. H. Tran, C. S. Hong, Z. Han, and S. Lee, "Optimal pricing effect on equilibrium behaviors of delay-sensitive users in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2566–2579, Nov. 2013.

[44] C. T. Do, N. H. Tran, Z. Han, L. B. Le, S. Lee, and C. S. Hong, "Optimal pricing for duopoly in cognitive radio networks: Cooperate or not cooperate?" *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2574–2587, May 2014.

[45] J. Wang, A. Huang, L. Cai, and W. Wang, "On the queue dynamics of multiuser multichannel cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1314–1328, Mar. 2013.

[46] Y. Zhao, S. Jin, and W. Yue, "A novel spectrum access strategy with α-retry policy in cognitive radio networks: A queueing-based analysis," *J. Commun. Netw.*, vol. 16, no. 2, pp. 193–201, Apr. 2014.

[47] W. S. Chang and W. M. Jang, "Spectrum occupancy of cognitive radio networks: A queueing analysis using retrial queue," *IET Netw.*, vol. 3, no. 3, pp. 218–227, Sep. 2014.

[48] A. N. Dudin, M. H. Lee, O. Dudina, and S. K. Lee, "Analysis of priority retrial queue with many types of customers and servers reservation as a model of cognitive radio system," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 186–199, Jan. 2017.

[49] 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," in *Proc. IEEE 68th Veh. Technol. Conf.*, Calgary, BC, Canada, Sep. 2008, pp. 1–5.

[50] L. Liu, G. Hu, M. Xu, and Y. Peng, "Server-uncertain spectrum trading in cognitive radio networks: A queueing-theoretic modeling approach," in *Proc. 7th Int. ICST Conf. Cogn. Radio Oriented Wireless Netw. Commun. (CROWNCOM)*, Stockholm, Sweden, Jun. 2012, pp. 320–325.

[51] L. Liu, X. Jin, G. Min, J. Hu, and K. Li, "Modelling and analysis of dynamic spectrum access in cognitive radio networks with self-similar traffic," in *Proc. IEEE Global Telecommun. Conf.*, Miami, FL, USA, Dec. 2010, pp. 1–5.

[52] I. Suliman and J. Lehtomäki, "Queueing analysis of opportunistic access in cognitive radios," in *Proc. 2nd Int. Workshop Cogn. Radio Adv. Spectr. Manage.*, Aalborg, Denmark, May 2009, pp. 153–157.

[53] M. Usman, H.-C. Yang, and M.-S. Alouini, "Extended delivery time analysis for cognitive packet transmission with application to secondary queuing analysis," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5300–5312, Oct. 2015.

[54] V. Bassoo and N. Khedun, "Improving the quality of service for users in cognitive radio network using priority queueing analysis," *IET Commun.*, vol. 10, no. 9, pp. 1063–1070, 2016.

[55] M. Raspopovic, C. Thompson, and K. Chandra, "Performance models for wireless spectrum shared by wideband and narrowband sources," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, vol. 3, Oct. 2005, pp. 1642–1647.

[56] M. Zhang, S. Jiang, G. Wei, and H. Wang, "Performance analysis of the cognitive radio network with a call level queue for secondary users," in *Proc. 5th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Beijing, China, Sep. 2009, pp. 1–4.

[57] B. Oklander and M. Sidi, "Modeling and analysis of system dynamics and state estimation in cognitive radio networks," in *Proc. IEEE 21st Int. Symp. Pers., Indoor Mobile Radio Commun. Workshops*, Istanbul, Turkey, Sep. 2010, pp. 49–53.

[58] W. M. Jang and W. S. Chang, "Queueing with transmission rate selection for cognitive radio networks in Nakagami-*m* fading," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, San Diego, CA, USA, Jan. 2013, pp. 247–251.

[59] A. Azarfar, J.-F. Frigo, and B. Sansò, "Analysis of cognitive radio networks based on a queueing model with server interruptions," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 1703–1708.

[60] A. Azarfar, J.-F. Frigon, and B. Sansò, "Dynamic selection of priority queueing discipline in cognitive radio networks," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Québec, QC, Canada, Sep. 2012, pp. 1–5.

[61] P. Kaur, A. Khosla, and M. Uddin, "Markovian queuing model for dynamic spectrum allocation in centralized architecture for cognitive radios," *IACSIT Int. J. Eng. Technol.*, vol. 3, no. 1, pp. 96–101, Feb. 2011.

[62] K. Kim, "T-preemptive priority queue and its application to the analysis of an opportunistic spectrum access in cognitive radio networks," *Comput. Oper. Res.*, vol. 39, no. 7, pp. 1394–1401, Jul. 2012.

[63] H. Tran, T. Q. Duong, and H.-J. Zepernick, "Delay performance of cognitive radio networks for point-to-point and point-to-multipoint communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 1–9, Dec. 2012, doi: 10.1186/1687-1499-2012-9.

[64] S. Gunawardena and W. Zhuang, "Service response time of elastic data traffic in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 559–570, Mar. 2013.

[65] J. Wang, A. Huang, W. Wang, and T. Q. S. Quek, "Admission control in cognitive radio networks with finite queue and user impatience," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, pp. 175–178, Apr. 2013.

[66] I. A. M. Balapuwaduge, L. Jiao, V. Pla, and F. Y. Li, "Channel assembling with priority-based queues in cognitive radio networks: Strategies and performance evaluation," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 630–645, Feb. 2014.

[67] F. A. Khan, K. Tourki, M.-S. Alouini, and K. A. Qaraqe, "Delay performance of a broadcast spectrum sharing network in Nakagami-*m* fading," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1350–1364, Mar. 2014.

[68] H. Al-Mahdi, M. A. Kalil, F. Liers, and A. Mitschele-Thiel, "Increasing spectrum capacity for ad hoc networks using cognitive radios: An analytical model," *IEEE Commun. Lett.*, vol. 13, no. 9, pp. 676–678, Sep. 2009.

[69] S. Chen, A. M. Wyglinski, S. Pagadarai, R. Vuyyuru, and O. Altintas, "Feasibility analysis of vehicular dynamic spectrum access via queueing theory model," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 156–163, Nov. 2011.

[70] Y. Lee, C. G. Park, and D. B. Sim, "Cognitive radio spectrum access with prioritized secondary users," *Appl. Math. Inf. Sci.*, vol. 6, no. 2S, pp. 595S–601S, 2012.

[71] A. K. Farraj, "Queue model analysis for spectrum-sharing cognitive systems under outage probability constraint," *Wireless Pers. Commun.*, vol. 73, no. 3, pp. 1021–1035, Dec. 2013.

[72] A. Azarfar, J.-F. Frigon, and B. Sansò, "Priority queueing models for cognitive radio networks with traffic differentiation," *EURASIP J. Wireless Commun. Netw.*, vol. 2014, p. 206, Dec. 2014, doi: 10.1186/1687-1499-2014-206.

[73] B. Sun, M. H. Lee, S. A. Dudin, and A. N. Dudin, "Analysis of multiserver queueing system with opportunistic occupation and reservation of servers," *Math. Problems Eng.*, vol. 2014, May 2014, Art. no. 178108, doi: 10.1155/2014/178108.

[74] B. Sun, M. H. Lee, A. N. Dudin, and S. A. Dudin, "$MAP + MAP/M_2/N/\infty$ queueing system with absolute priority and reservation of servers," *Math. Problems Eng.*, vol. 2014, Nov. 2014, Art. no. 813150, doi: 10.1155/2014/813150.

[75] N. Tadayon and S. Aïssa, "Multi-channel cognitive radio networks: Modeling, analysis and synthesis," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2065–2074, Nov. 2014.

[76] A. Azarfar, J.-F. Frigon, and B. Sansò, "Queueing model for heterogeneous opportunistic spectrum access," *IET Commun.*, vol. 9, no. 6, pp. 819–827, Apr. 2015.

[77] B. Almási, T. Bérczes, A. Kuki, J. Sztrik, and J. Wang, "Performance modeling of finite-source cognitive radio networks," *Acta Cybern.*, vol. 22, no. 3, pp. 617–631, Jun. 2016.

[78] V. Kumar, S. Minz, and V. Kumar, "Performance analysis of cognitive radio networks under spectrum sharing using queuing approach," *Comput. Elect. Eng.*, vol. 52, pp. 390–398, May 2016.

[79] B. S. Awoyemi, B. T. Maharaj, and A. S. Alfa, "Resource allocation in heterogeneous buffered cognitive radio networks," *Wireless Commun. Mobile Comput.*, vol. 2017, Jul. 2017, Art. no. 7385627, doi: 10.1155/2017/7385627.

[80] B. Canberk, I. F. Akyildiz, and S. Oktug, "A QoS-aware framework for available spectrum characterization and decision in cognitive radio networks," in *Proc. 21st Annu. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Istanbul, Turkey, Sep. 2010, pp. 1533–1538.

[81] B. Jashni, A. A. Tadaion, and F. Ashtiani, "Dynamic link/frequency selection in multi-hop cognitive radio networks for delay sensitive applications," in *Proc. 17th Int. Conf. Telecommun.*, Doha, Qatar, Apr. 2010, pp. 128–132.

[82] Y. Zhao, M. Song, and C. Xin, "Delay analysis for cognitive radio networks supporting heterogeneous traffic," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, Salt Lake City, UT, USA, Jun. 2011, pp. 215–223.

[83] S. A. AlQahtani, "Performance analysis of cognitive-based radio resource allocation in multi-channel LTE-A networks with M2M/H2H coexistence," *IET Commun.*, vol. 11, no. 5, pp. 655–663, Mar. 2017.

[84] K. Wu, W. Wang, H. Luo, G. Yu, and Z. Zhang, "Optimal resource allocation for cognitive radio networks with imperfect spectrum sensing," in *Proc. IEEE 71st Veh. Technol. Conf.*, Taipei, Taiwan, May 2010, pp. 1–4.

[85] C. T. Do, N. H. Tran, and C. S. Hong, "Throughput maximization for the secondary user over multi-channel cognitive radio networks," in *Proc. Int. Conf. Inf. Netw.*, Bali, Indonesia, Feb. 2012, pp. 65–69.

[86] M. Dashti and P. Azmi, "Joint power and rate allocation in CDMA-based underlay cognitive radio networks for a mixture of streaming and elastic traffic," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, p. 262, Dec. 2012, doi: 10.1186/1687-1499-2012-262.

[87] T. Jiang, H. Wang, and A. V. Vasilakos, "QoE-driven channel allocation schemes for multimedia transmission of priority-based secondary users over cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 7, pp. 1215–1224, Aug. 2012.

[88] M. R. M. Falcão, A. M. Balieiro, and K. L. Dias, "A flexible-bandwidth model with channel reservation and channel aggregation for three-layered cognitive radio networks," *Comput. Netw.*, vol. 135, pp. 213–225, Apr. 2018.

[89] N. Suganthi and S. Meenakshi, "An efficient scheduling algorithm using queuing system to minimize starvation of non-real-time secondary users in cognitive radio network," *Cluster Comput.*, pp. 1–11, Mar. 2018, doi: 10.1007/s10586-017-1595-8.

[90] L. Xu, A. Nallanathan, X. Pan, J. Yang, and W. Liao, "Security-aware resource allocation with delay constraint for NOMA-based cognitive radio network," , vol. 13, no. 2, pp. 366–376, Feb. 2018.

[91] W. Wang, K. Wu, H. Luo, G. Yu, and Z. Zhang, "Sensing error aware delay-optimal channel allocation scheme for cognitive radio networks," *Telecommun. Syst.*, vol. 52, no. 4, pp. 1895–1904, Apr. 2013.

[92] L. Zhang, T. Song, X. Bao, M. Wu, D. Sun, and J. Hu, "Optimal spectrum access strategy for multi-channel cognitive radio networks with Nakagami fading and finite-size buffer," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 1428–1433.

[93] Z. Ni, H. Shan, W. Shen, J. Wang, A. Huang, and X. Wang, "Dynamic channel allocation-based call admission control in cognitive radio networks," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Hangzhou, China, Oct. 2013, pp. 1–6.

[94] L. Chouhan and A. Trivedi, "Priority based MAC scheme for cognitive radio network: A queuing theory modelling," in *Proc. 9th Int. Conf. Wireless Opt. Commun. Netw. (WOCN)*, Indore, India, Sep. 2012, pp. 1–5.

[95] S. Feizi-Khankandi, F. Ashtiani, S. H. Hassani, and P. Tootoonchi, "Modeling the behavior of contending opportunistic cognitive radios," in *Proc. IEEE Int. Conf. Signal Process. Commun.*, Dubai, United Arab Emirates, Nov. 2007, pp. 720–723.

[96] S. Jiang, L.-H. Chao, and H.-L. Chao, "A decentralized MAC protocol for cognitive radio networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Shanghai, China, Apr. 2011, pp. 24–29.

[97] S. N. Shankar, "Squeezing the most out of cognitive radio: A joint MAC/PHY perspective," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 4, Apr. 2007, pp. 1361–1364.

[98] D. B. Zhu, H.-M. Wang, and Y.-N. Xu, "Performance analysis of CSMA in an unslotted cognitive radio network under non-saturation condition," in *Proc. 2nd Int. Conf. Instrum., Meas., Comput., Commun. Control*, Harbin, China, Dec. 2012, pp. 1122–1126.

[99] X. Li, J. Wang, H. Li, and S. Li, "Delay performance analysis and access strategy design for a multichannel cognitive radio network," *Chin. Sci. Bull.*, vol. 57, nos. 28–29, pp. 3705–3712, Oct. 2012.

[100] C. Jiang, Y. Chen, K. J. R. Liu, and Y. Ren, "Renewal-theoretical dynamic spectrum access in cognitive radio network with unknown primary behavior," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 406–416, Mar. 2013.

[101] Z. Zeng and G. Wang, "Dynamic spectrum access with prioritized secondary users in cognitive radio networks," *J. Inf. Comput. Sci.*, vol. 10, no. 18, pp. 5811–5820, Dec. 2013.

[102] L. Zhang, T. Song, J. Hu, and X. Bao, "Analysis of spectrum access strategy with multiple cross-layer considerations in cognitive radio networks," *Wireless Pers. Commun.*, vol. 87, no. 4, pp. 1383–1400, Apr. 2016.

[103] T. E. Fahim, A. Y. Zakariya, and S. I. Rabia, "A novel hybrid priority discipline for multi-class secondary users in cognitive radio networks," *Simul. Model. Pract. Theory*, vol. 84, pp. 69–82, May 2018.

[104] S. Tang and B. L. Mark, "Modeling and analysis of opportunistic spectrum sharing with unreliable spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1934–1943, Apr. 2009.

[105] Y. Wu *et al.*, "A learning-based QoE-driven spectrum handoff scheme for multimedia transmissions over cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2134–2148, Nov. 2014.

[106] Y. Wu, F. Hu, Y. Zhu, and S. Kumar, "Optimal spectrum handoff control for CRN based on hybrid priority queuing and multi-teacher apprentice learning," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2630–2642, Mar. 2017.

[107] A. M. Koushik, F. Hu, and S. Kumar, "Multi-class 'channel+beam' handoff in cognitive radio networks with multi-beam smart antennas," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington DC, USA, Dec. 2016, pp. 1–6.

[108] L. Zhang, T. Song, M. Wu, X. Bao, J. Guo, and J. Hu, "Traffic-adaptive proactive spectrum handoff strategy for graded secondary users in cognitive radio networks," *Chin. J. Electron.*, vol. 24, no. 4, pp. 844–851, Oct. 2015.

[109] A. F. Tayel, S. I. Rabia, and Y. Abouelseoud, "An optimized hybrid approach for spectrum handoff in cognitive radio networks with non-identical channels," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4487–4496, Nov. 2016.

[110] S. T. Talat, C.-W. Wang, and L.-C. Wang, "Spectrum decision for cognitive radio networks with various-bandwidth channels," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Doha, Qatar, Apr. 2016, pp. 1–6.

[111] L.-C. Wang, C.-W. Wang, and C.-J. Chang, "Modeling and analysis for spectrum handoffs in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 9, pp. 1499–1513, Sep. 2012.

[112] L.-C. Wang, C.-W. Wang, and C.-J. Chang, "Optimal target channel sequence design for multiple spectrum handoffs in cognitive radio networks," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2444–2455, Sep. 2012.

[113] S. Zahed, I. Awan, and A. Cullen, "Analytical modeling for spectrum handoff decision in cognitive radio networks," *Simul. Model. Pract. Theory*, vol. 38, pp. 98–114, Nov. 2013.

[114] A. M. Koushik, F. Hu, and S. Kumar, "Intelligent spectrum management based on transfer actor-critic learning for rateless transmissions in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 5, pp. 1204–1215, May 2018.

[115] X. Yang, X. Tan, L. Ye, and L. Ma, "Spectrum handoffs based on preemptive repeat priority queue in cognitive radio networks," *Sensors*, vol. 16, no. 7, p. 1127, Jul. 2016, doi: 10.3390/s16071127.

[116] L. Sibomana, H.-J. Zepernick, H. Tran, and C. Kabiri, "A framework for packet delay analysis of point-to-multipoint underlay cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 9, pp. 2408–2421, Sep. 2017.

[117] S. Agrawal, V. Rana, and A. K. Jagannatham, "Queuing analysis for multiple-antenna cognitive radio wireless networks with beamforming," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 334–338, Mar. 2017.

[118] A. K. Farraj, S. L. Miller, and K. A. Qaraqe, "Queue performance measures for cognitive radios in spectrum sharing systems," in *Proc. IEEE GLOBECOM Workshops*, Houston, TX, USA, Dec. 2011, pp. 997–1001.

[119] H. M. Tsimba, B. T. Maharaj, and A. S. Alfa, "Increased spectrum utilisation in a cognitive radio network: An M/M/1-PS queue approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, Mar. 2017, pp. 1–6.

[120] L. Chen, L. Huang, H. Xu, and Y. Du, "Queueing analysis for preemptive transmission in underlay CRNs," *Int. J. Commun. Syst.*, vol. 29, no. 6, pp. 1138–1155, Apr. 2016.

[121] Z. Chang, T. Ristaniemi, and Z. Han, "Queueing game for spectrum access in cognitive radio networks," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 2017–2020, Nov. 2015.

[122] L. Guijarro, V. Pla, and B. Tuffin, "Entry game under opportunistic access in cognitive radio networks: A priority queue model," in *Proc. IFIP Wireless Days (WD)*, Valencia, Spain, Nov. 2013, pp. 1–6.

[123] G. Wang, Z. Zeng, and M. Liu, "Equilibrium threshold strategy in observable queueing systems in cognitive radio networks," in *Proc. Int. Conf. Meas., Inf. Control (ICMIC)*, vol. 1, Aug. 2013, pp. 328–332.

[124] J. Wang, Y. Zhang, and W. W. Li, "Strategic joining and optimal pricing in the cognitive radio system with delay-sensitive secondary users," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 3, pp. 298–312, Sep. 2017.

[125] M. A. Safwat, "Equilibrium analysis for spectrum access in cognitive radio networks with traffic prioritization," in *Proc. Int. Conf. Comput. Commun. Eng. (ICCCE)*, Kuala Lumpur, Malaysia, Jul. 2016, pp. 355–360.

[126] R. Nissel and M. Rupp, "Dynamic spectrum allocation in cognitive radio: Throughput calculations," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, Jun. 2016, pp. 1–5.

[127] J. Elias, F. Martignon, L. Chen, and E. Altman, "Joint operator pricing and network selection game in cognitive radio networks: Equilibrium, system dynamics and price of anarchy," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4576–4589, Nov. 2013.

[128] N. H. Tran, L. B. Le, S. Ren, Z. Han, and C. S. Hong, "Joint pricing and load balancing for cognitive spectrum access: Non-cooperation versus cooperation," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 5, pp. 972–985, May 2015.

[129] M. Fahimi and A. Ghasemi, "Joint spectrum load balancing and handoff management in cognitive radio networks: A non-cooperative game approach," *Wireless Netw.*, vol. 22, no. 4, pp. 1161–1180, May 2016.

[130] J. Wang and W. W. Li, "Noncooperative and cooperative joining strategies in cognitive radio networks with random access," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5624–5636, Jul. 2016.

[131] J. Zou and L. Huang, "Joint pricing and service selection for service-differentiated duopoly in cognitive radio networks," *IET Commun.*, vol. 11, no. 13, pp. 2074–2081, Oct. 2017.

[132] K. Jagannathan, I. Menache, E. Modiano, and G. Zussman, "Non-cooperative spectrum access—The dedicated vs. free spectrum choice," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2251–2261, Dec. 2012.

[133] K. Kotobi and S. G. Bilén, "Spectrum sharing via hybrid cognitive players evaluated by an M/D/1 queuing model," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, p. 85, Dec. 2017, doi: 10.1186/s13638-017-0871-x.

[134] R. Hassin and R. I. Snitkovsky, "Strategic customer behavior in a queueing system with a loss subsystem," *Queueing Syst.*, vol. 86, nos. 3–4, pp. 361–387, Aug. 2017.

[135] C. Zhang, X. Wang, and J. Li, "Cooperative cognitive radio with priority queueing analysis," in *Proc. IEEE Int. Conf. Commun.*, Dresden, Germany, Jun. 2009, pp. 1–5.

[136] H.-B. Chang, S.-M. Cheng, S.-Y. Lien, and K.-C. Chen, "Statistical delay control of opportunistic links in cognitive radio networks," in *Proc. 21st Annu. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Istanbul, Turkey, Sep. 2010, pp. 2250–2254.

[137] B. Shahrasbi and N. Rahnavard, "Rateless-coding-based cooperative cognitive radio networks: Design and analysis," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, Salt Lake, UT, USA, Jun. 2011, pp. 224–232.

[138] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.

[139] H. Tran, H.-J. Zepernick, H. Phan, and L. Sibomana, "Performance analysis of a cognitive radio network with a buffered relay," *IEEE Trans. Veh. Technol.*, vol. 64, no. 2, pp. 566–579, Feb. 2015.

[140] A. Balieiro, M. Falcão, and K. L. Dias, "Secondary virtual network mapping onto cognitive radio substrate: A collision probability analysis," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 600–603, Mar. 2017.

[141] A. P. T. R. da Silva and J. M. C. Brito, "Influence of the channel model in the optimum switching points in a cognitive radio network combined with adaptive modulation," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, May 2016, pp. 1–6.

[142] N. Adem and B. Hamdaoui, "Delay performance modeling and analysis in clustered cognitive radio networks," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, Dec. 2014, pp. 193–198.

[143] B. Zaman, Z. H. Abbas, and F. Y. Li, "Spectrum occupancy and residual service analysis in CRNs using a multi-server queueing model," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, Glasgow, U.K., May 2015, pp. 1–5.

[144] S. Jin, S. Ge, and W. Yue, "Performance evaluation for an opportunistic spectrum access mechanism with impatience behavior and imperfect sensing results," in *Proc. 7th Int. Conf. Ubiquitous Future Netw.*, Sapporo, Japan, Jul. 2015, pp. 959–962.

[145] J. Jeon, M. Codreanu, M. Latva-aho, and A. Ephremides, "The stability property of cognitive radio systems with imperfect sensing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 3, pp. 628–640, Mar. 2014.

[146] M. Luís, R. Oliveira, R. Dinis, and L. Bernardo, "Characterization of the opportunistic service time in cognitive radio networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 2, no. 3, pp. 288–300, Sep. 2016.

[147] I. A. A. Maksoud, S. I. Rabia, and M. A. Algundi, "A discrete-time multi-server queueing model for opportunistic spectrum access systems," *Perform. Eval.*, vol. 109, pp. 1–7, Mar. 2017.

[148] B. Zaman, Z. H. Abbas, and F. Y. Li, "A lightweight spectrum occupancy and service time model for centralized cognitive radio networks," *Wireless Pers. Commun.*, vol. 92, no. 4, pp. 1675–1694, Feb. 2017.

[149] Y. Zhao, S. Jin, and W. Yue, "Performance analysis of cognitive radio networks for secondary users with slotted central control," *Telecommun. Syst.*, vol. 66, no. 4, pp. 689–699, Dec. 2017.

[150] Y. Zhao and W. Yue, "Performance analysis and optimization of cognitive radio networks with retransmission control," *Optim. Lett.*, vol. 12, no. 6, pp. 1281–1300, 2017, doi: 10.1007/s11590-017-1119-8.

[151] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Stable throughput of cognitive radios with and without relaying capability," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2351–2360, Dec. 2007.

[152] A. A. El-Sherif and A. Mohamed, "Decentralized throughput maximization in cognitive radio wireless mesh networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 9, pp. 1967–1980, Sep. 2014.

[153] H. Cao, H. Tian, J. Cai, A. S. Alfa, and S. Huang, "Dynamic load-balancing spectrum decision for heterogeneous services provisioning in multi-channel cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5911–5924, Sep. 2017.

[154] A. Bayat, M. Mahdavi, and S. Aïssa, "Performance evaluation of cognitive radio VoIP users in fading environment," *Wireless Pers. Commun.*, vol. 97, no. 3, pp. 3951–3977, Dec. 2017.

[155] S. Jin and Y. Sun, "System model and performance estimation of dynamic spectrum allocation strategy with multi-channel and imperfect sensing," *Int. J. Comput. Math.*, vol. 94, no. 9, pp. 1727–1737, 2017.

[156] E. F. Khairullah and M. Chatterjee, "Queue based scheduling in single and multi channel dynamic spectrum access networks," *Pervasive Mobile Comput.*, vol. 46, pp. 73–95, Jun. 2018.

[157] A. Azarfar, J. F. Frigon, and B. Sansò, "Delay analysis of multichannel opportunistic spectrum access MAC protocols," *IEEE Trans. Mobile Comput.*, vol. 15, no. 1, pp. 92–106, Jan. 2016.

[158] A. M. Arafa, K. G. Seddik, A. K. Sultan, T. ElBatt, and A. A. El-Sherif, "A feedback-soft sensing-based access scheme for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3226–3237, Jul. 2013.

[159] M. R. Amini, M. Mahdavi, and M. J. Omidi, "Maximizing dynamic access energy efficiency in multiuser CRNs with primary user return," *IEEE Systems J.*, to be published.

[160] C.-M. Lee, J.-S. Lin, Y.-P. Hsu, and K.-T. Feng, "Design and analysis of optimal channel-hopping sequence for cognitive radio networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Sydney, NSW, Australia, Apr. 2010, pp. 1–6.

[161] M. E. Bayrakdar and A. Çalhan, "Improving spectrum handoff utilization for prioritized cognitive radio users by exploiting channel bonding with starvation mitigation," *Int. J. Electron. Commun.*, vol. 71, pp. 181–191, Jan. 2017.

[162] S. Kompella, G. Nguyen, C. Kam, J. E. Wieselthier, and A. Ephremides, "Cooperation in cognitive underlay networks: Stable throughput tradeoffs," *IEEE/ACM Trans. Netw.*, vol. 22, no. 6, pp. 1756–1768, Dec. 2014.

[163] D. Zhang, Z. Chen, M. K. Awad, N. Zhang, H. Zhou, and X. S. Shen, "Utility-optimal resource management and allocation algorithm for energy harvesting cognitive radio sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3552–3565, Dec. 2016.

[164] R. Amer, A. A. El-Sherif, H. Ebrahim, and A. Mokhtar, "Stability analysis for multi-user cooperative cognitive radio network with energy harvesting," in *Proc. 2nd IEEE Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Oct. 2016, pp. 2369–2375.

[165] M. A. Abd-Elmagid, T. ElBatt, and K. G. Seddik, "On the role of finite queues in cooperative cognitive radio networks with energy harvesting," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Silicon Valley, CA, USA, Jan. 2017, pp. 137–143.

[166] Y. Lu, W. Wang, Z. Zhang, A. Huang, and V. K. N. Lau, "Random access for a cognitive radio transmitter with RF energy harvesting," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, Dec. 2014, pp. 923–928.

[167] D. Niyato, P. Wang, and D. I. Kim, "Performance analysis of cognitive radio networks with opportunistic RF energy harvesting," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, Dec. 2014, pp. 1096–1101.

[168] S. K. Nobar, K. A. Mehr, J. M. Niya, and B. M. Tazehkand, "Cognitive radio sensor network with green power beacon," *IEEE Sensors J.*, vol. 17, no. 5, pp. 1549–1561, Mar. 2017.

[169] N. Pappas, J. Jeon, A. Ephremides, and A. Traganitis, "Optimal utilization of a cognitive shared channel with a rechargeable primary source node," *J. Commun. Netw.*, vol. 14, no. 2, pp. 162–168, Apr. 2012.

[170] S. Jin, X. Ma, and W. Yue, "Energy-saving strategy for green cognitive radio networks with an LTE-advanced structure," *J. Commun. Netw.*, vol. 18, no. 4, pp. 610–618, Aug. 2016.

[171] H. Li and Z. Han, "Socially optimal queuing control in cognitive radio networks subject to service interruptions: To queue or not to queue?" *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1656–1666, May 2011.

[172] S. Jin, W. Yue, and S. Ge, "Equilibrium analysis of an opportunistic spectrum access mechanism with imperfect sensing results," *J. Ind. Manage. Optim.*, vol. 13, no. 3, pp. 1255–1271, Jul. 2017.

[173] A. Homayounzadeh and M. Mahdavi, "Performance analysis of cooperative cognitive radio networks with imperfect sensing," in *Proc. Int. Conf. Commun., Signal Process., Appl. (ICCSPA)*, Feb. 2015, pp. 1–6.

[174] J. Feng, W. Liang, S. X. Ng, and L. Hanzo, "Distributed reciprocal-selection-based 'win-win' cooperative medium access and its stability analysis," *IEEE Access*, vol. 4, pp. 7703–7715, 2016.

[175] A. El Shafie, T. Khattab, and A. S. Salem, "Relay-assisted primary and secondary transmissions in cognitive radio networks," *IEEE Access*, vol. 4, pp. 6386–6400, 2016.

[176] A. M. Elmahdy, A. El-Keyi, T. ElBatt, and K. G. Seddik, "Optimizing cooperative cognitive radio networks performance with primary QoS provisioning," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1451–1463, Apr. 2017.

[177] D. Kiwan, A. El Sherif, and T. ElBatt, "Stability analysis of a cognitive radio system with a dedicated relay," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Silicon Valley, CA, USA, Jan. 2017, pp. 750–756.

[178] Y. Huang, Q. Wu, J. Wang, and Y. Cheng, "Protocol design and performance analysis for cognitive cooperative networks with multiple antennas," *EURASIP J. Wireless Commun. Netw.*, vol. 2013, p. 70, Dec. 2013, doi: 10.1186/1687-1499-2013-70.

[179] A. A. El-Sherif, A. K. Sadek, and K. Liu, "Opportunistic multiple access for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 704–715, Apr. 2011.

[180] J. Liu, W. Chen, Z. Cao, and Y. J. A. Zhang, "Cooperative beamforming for cognitive radio networks: A cross-layer design," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1420–1431, May 2012.

[181] X. Bao, P. Martins, T. Song, and L. Shen, "Stable throughput and delay performance in cognitive cooperative systems," *IET Commun.*, vol. 5, no. 2, pp. 190–198, Jan. 2011.

[182] R. Urgaonkar and M. J. Neely, "Opportunistic cooperation in cognitive femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 607–616, Apr. 2012.

[183] D. Hamza and S. Aïssa, "Enhanced primary and secondary performance through cognitive relaying and leveraging primary feedback," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2236–2247, Jun. 2014.

[184] M. Ashour, A. A. El-Sherif, T. ElBatt, and A. Mohamed, "Cognitive radio networks with probabilistic relaying: Stable throughput and delay tradeoffs," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4002–4014, Nov. 2015.

[185] K. Kulkarni and A. Banerjee, "On optimal spectrum access of cognitive relay with finite packet buffer," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7584–7588, Aug. 2017.

[186] D. Xu, Y. Gao, C. Yang, Y. Yao, and B. Xia, "Performance analysis of opportunistic cooperation schemes in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3658–3662, Apr. 2018.

[187] W. M. Jang and W. S. Chang, "Spectral occupancy rate of cognitive radio networks in queueing with adaptive modulation and coding over multiple-input and multiple-output transmit antenna selection/maximal ratio combining in Nakagami-*m* fading," *IET Netw.*, vol. 3, no. 2, pp. 160–167, Jun. 2014.

[188] P. Rahimzadeh and F. Ashtiani, "Analytical evaluation of saturation throughput of a cognitive WLAN overlaid on a time-scheduled OFDMA network," *IEEE Trans. Mobile Comput.*, vol. 16, no. 3, pp. 634–647, Mar. 2017.

[189] A. S. Alfa, V. Pla, J. Martinez-Bauset, and V. Casares-Giner, "Discrete time analysis of cognitive radio networks with imperfect sensing and saturated source of secondary users," *Comput. Commun.*, vol. 79, pp. 53–65, Apr. 2016.

[190] E. Biglieri, A. J. Goldsmith, L. J. Greenstein, N. B. Mandayam, and H. V. Poor, *Principles of Cognitive Radio*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

FILIP PALUNČIĆ was born in Belgrade, Serbia. He received the M.Ing. and D.Ing. degrees from the University of Johannesburg, South Africa, in 2008 and 2012, respectively. He spent four years in industry as a Research and Development Engineer with IDX, a company specializing in industrial communications. From 2016 to 2017, he was a Post-Doctoral Research Fellow in broadband wireless multimedia communications with the Sentech Group, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, where he is currently a Faculty Member with the Department of Electrical, Electronic and Computer Engineering. His research interests include coding techniques (in particular error control coding and constrained coding), information theory, cognitive radio networks, and wireless communications.

ATTAHIRU S. ALFA is currently a Professor Emeritus with the Department of Electrical and Computer Engineering, University of Manitoba, and also the UP/CSIR Co-Hosted SARChI Chair Professor with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria. He is also involved in the application of queuing theory to other areas, such as transportation systems, manufacturing systems, and healthcare systems. He has authored two books *Queueing Theory for Telecommunications: Discrete Time Modelling of a Single Node System* (Springer, 2010) and *Applied Discrete-Time Queues* (Springer, 2015). His research covers, but not limited to, the following areas, such as queuing theory and applications, optimization, performance analysis and resource allocation in telecommunication systems, modeling of communication networks, analysis of cognitive radio networks, modeling and analysis of wireless sensor networks, developing efficient decoding algorithms for LDPC codes, channel modeling, traffic estimation for the Internet, and cross layer analysis.

B. T. (SUNIL) MAHARAJ received the Ph.D. degree in engineering in the areas of wireless communications from the University of Pretoria. He is currently a Full Professor and currently holds the research position of Sentech Chair in broadband wireless multimedia communications with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria. His research interests are in OFDM-MIMO, massive MIMO systems, cognitive radio resource allocation, and 5G cognitive radio sensor networks.

HILARY M. TSIMBA received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Pretoria, South Africa, in 2014 and 2017, respectively. He studied in broadband wireless multimedia communications with the Sentech Group, University of Pretoria. He was a Research Intern with the Postal and Telecommunications Regulatory Authority of Zimbabwe. His research interests include cognitive radio networks, queuing theory, and matrix-geometric analytic methods.

• • •