

# Queue management for two-user cognitive radio with delay-constrained primary user

Kamal Adli Mehr<sup>a,1</sup>, Javad Musevi Niya<sup>a</sup>, Nail Akar<sup>b,2,\*</sup>

<sup>a</sup> WiLab, Department of Electrical and Computer Engineering, University of Tabriz, Iran

<sup>b</sup> Electrical and Electronics Engineering Department, Bilkent University, Turkey

## ARTICLE INFO

### Article history:

Received 27 September 2017

Revised 11 April 2018

Accepted 30 May 2018

Available online 31 May 2018

### Keywords:

Queue management

Multi-regime Markov fluid queue

Overlay cognitive radio

Physical layer security

## ABSTRACT

In this paper, two novel Queue Management Policies (QMP) are proposed for Quality of Service (QoS) enhancement of a two-user Cognitive Radio Network (CRN) comprising a Primary User (PU) and Secondary User (SU), the latter having non-causal information on PU's messages (or packets). Specifically, we aim to maximize the throughput of the SU while satisfying the delay criterion of the Primary User (PU). The first proposed QMP is a hybrid interweave/overlay scheme where all the SU's resources are devoted to the transmission of PU's packets. The second proposed QMP adaptively uses all or some of the SU's resources towards the transmission of the PU's packet, this decision being based on the packet's delay experienced in the PU queue. For this adaptive QMP, a novel multi-regime Markov fluid queue model is proposed via which closed-form expressions are derived and validated for the exact delay distribution for Poisson PU traffic and exponentially distributed packet lengths. Using this analytical tool, we optimally tune the parameters of the adaptive QMP and we show through numerical examples that it consistently outperforms the hybrid interweave/overlay model as well as two other conventional schemes in terms of SU throughput. We also show that the performance improvement attainable by the proposed QMP depends on the intensity of PU traffic as well as the channel conditions. A heuristic suboptimal parameter tuning scheme is also proposed with lesser computational complexity.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Demand for access to wireless spectrum has been growing dramatically. However, conservative spectrum allocation policies have given rise to underutilized licensed channels [1]. CRNs have emerged as a promising solution to cope with spectrum inefficiency in these channels [2]. Key to CRNs is the incorporation of Secondary Users (SU) of lower priority which should transmit their messages in a way that the Primary Users (PU) of the licensed channel would not be adversely affected. There are three distinct deployment paradigms for cognitive radio, namely underlay, interweave, and overlay paradigms [3]. These paradigms are described briefly in Table 1. In the underlay paradigm, the cognitive transmission is allowed if the interference generated by the cognitive user on the primary receiver is below a predefined threshold. On

the other hand, in the interweave paradigm, the SU utilizes temporary spectrum holes in space, time, and/or frequency opportunistically to transmit its messages, i.e., SU transmits its messages whenever there is no ongoing primary transmission on the channel. In contrast, in the overlay paradigm, the secondary transmitter completely cooperates with the primary transmitter to deliver the PU's messages to the intended receivers. In exchange, the PU permits the cooperative SU to utilize the licensed channel of PU [3]. In this paradigm, the primary message is known at the secondary transmitter causally or non-causally. The overlay paradigm is suitable for modeling primary networks that sublease their spectrum for monetary or security purposes for which a level of cooperation between the primary and secondary networks is not unusual [4]. For instance, a cognitive radio system where the cognitive transmitter has acquired the message of the primary user in the previous phases, or, a cognitive radio system, where both primary and secondary transmitters belong to the same authority (such as cloud empowered cellular network [5]) are examples of overlay cognitive radio systems.

In this paper, a novel framework is introduced for managing the queues in a two-user overlay cognitive radio scenario with the SU having non-causal information about PU's messages and the PU

\* Corresponding author.

E-mail addresses: [k.adlimehr@tabrizu.ac.ir](mailto:k.adlimehr@tabrizu.ac.ir) (K.A. Mehr), [niya@tabrizu.ac.ir](mailto:niya@tabrizu.ac.ir) (J.M. Niya), [akar@ee.bilkent.edu.tr](mailto:akar@ee.bilkent.edu.tr) (N. Akar).

<sup>1</sup> The research of K. A. Mehr was carried out when he was visiting Bilkent University, Turkey.

<sup>2</sup> His research is supported in part by Tübitak project 115E360.

**Table 1**  
Comparison of underlay, interweave, and overlay cognitive radio paradigms.

	Network-side information known at the SU	(Simultaneous) transmission condition	SU's transmit power is limited by
<b>Underlay</b>	Interference generated to the primary receiver	When the interference generated to the primary receiver is below a predefined threshold	Limited by the interference caused to the primary transmitter
<b>Interweave</b>	Spectrum holes in space, time, and/or frequency	When there is detection of loss of PU activity	Limited by the range of PU activity that the SU can detect
<b>Overlay</b>	Channel gains, encoding techniques, and possibly PU's messages	When the interference to the PU can be offset by using part of SU's power to relay PU's messages	SU can transmit at any power level

is subject to statistical delay constraints. Furthermore, two novel QMPs are proposed for the packet queue held at the PU. Two-user overlay scenarios (also known as Cognitive Interference Channel (CIFIC)) have been studied by information-theoretic methods in [3,6,7]. A recent survey on the extension of CIFIC to the case of multiple users is presented in [8]. In this paper, we focus on the two-user CIFIC scenario similar to some recent studies on cognitive radio with QoS constraints [9–11]. Building upon the existing information-theoretic results, we propose queue management policies so as to provide as many transmission opportunities as possible for the SU (providing maximum benefit for the cooperative cognitive user) while satisfying the PU's delay constraint given in terms of the probability that a packet's queuing delay exceeding a given threshold being less than a certain value. Nonetheless, the proposed model can also directly be used to model the case of multiple PU-SU pairs where each pair receives a static time-frequency resource allocation. Moreover, the proposed model can be viewed as a preliminary step towards gaining insight in delay-constrained overlay cognitive radio and also towards modeling of dynamical sharing of resources across PU-SU pairs with distributed MAC (Medium Access Control) protocols which is left for future research.

The first proposed QMP is inspired by a combination of conventional interweave and overlay paradigms and in this QMP, the PU leverages all of SU's resources unless the PU's queue is empty. The SU is only allowed to transmit its own packets whenever the PU's queue is empty. The second proposed QMP adaptively uses all or some of the SU's resources towards the transmission of the PU's packet through a quantity called the power allocation factor. This decision depends on the packet's delay experienced in the PU queue exceeding a certain threshold, i.e., delay-aware QMP. Conventional queuing models fall short of modeling this delay-aware QMP. The main contribution of this paper lies in the novel queuing model of the delay-aware QMP assuming that the packet arrival process at the PU queue is Poisson while the SU's queue is assumed to be saturated, i.e., SU always has a packet to transmit. Moreover, PU packet lengths are assumed to be exponentially distributed. Key to the proposed queuing model is the tool of Multi-Regime Markov Fluid Queues (MRMFQ) which is used to obtain the exact steady-state delay distribution for the proposed delay-aware QMP. We refer to [12] and the references therein for description and steady-state solutions of MRMFQs.

In the numerical experimentation of the proposed QMPs, we focus on a fairly general secure overlay cognitive radio scenario due to the fact that physical layer security is gaining more attention in the context of cognitive radio [13]. The SU performance encountered with both QMPs are thoroughly investigated in this scenario and compared with the conventional interweave and overlay paradigms. For this purpose, we first derive an achievable secure rate region for the Additive White Gaussian Noise (AWGN) channel model. As the next step, based on the analytical expressions derived using MRMFQs, the optimal power allocation factor and delay threshold values are numerically obtained to maximize the throughput of the SU while meeting the PU's delay constraint.

The remainder of the paper is organized as follows: Section 2 provides a brief literature overview. Section 3 presents the system model. Analytical models of the QMPs are presented in Section 4. In Section 5, parameter optimization of the proposed delay-aware QMP is addressed. Section 6 is devoted to numerical (both simulation and analysis) results. Finally, we conclude.

## 2. Related work

The interaction of cognitive radio with the higher layers of the network hierarchy has attracted significant attention of researchers. For the overlay paradigm, the SU devotes a portion of

its power to keep the PU's rate unchanged, while dedicating the rest of its power to deliver its own messages, without considering QoS constraints [3]. However, most of the wireless services are QoS-constrained and the considered policy in [3] may not be an appropriate choice for serving delay-sensitive PU applications that have been gaining popularity. In this scenario, the PU carrying delay sensitive information (like TV broadcasting for cognitive radios in TV white space) may be willing to permit the SU to utilize its licensed spectrum as long as its QoS requirements, e.g., delay requirements, are not compromised. Furthermore, the cognition capabilities of SUs can be deployed to improve the user experience of PUs. QoS-constrained SU in an underlay cognitive radio relay channel is studied in [4] and the so called effective capacity [14] is derived. This scenario is extended to an energy-constrained SU in [9] where a proper band from a pool of available spectrum bands is to be chosen. The reference [10] considers optimal power allocation for a delay-QoS aware secure underlay cognitive network which consists of one PU-SU pair and an eavesdropper. In [11], queuing analysis of a cooperative cognitive radio with a QoS-constrained PU and a cooperative SU is considered. Optimum admission control parameters of the primary packets at the SU is derived to maximize the SU's throughput while meeting PU's QoS conditions. Despite the rich literature in regards with the underlay cognitive radio paradigm with QoS-constrained SU, the overlay cognitive radio and QoS constraints of the PU did not get enough attention, which motivates us to conduct this research.

To become mainstream, another major challenge of cognitive radios is security. Recently, physical layer approaches are gaining more attention to provide security mechanisms for cognitive radio [13]. Due to node limitations, cooperative methods are deployed to improve secrecy performance. The cooperation between the PU and SU in cognitive radio is typically studied in the context of the overlay paradigm. On the other hand, performance of QoS constrained wireless communication systems with Physical Layer Security (PLS) is investigated in a few studies. In [15], a cross-layer resource allocation problem is investigated in a wireless single-hop uplink scenario by taking the information-theoretic secrecy as a QoS metric. A similar problem is studied in [16] to control the uplink of a cellular network by deploying hybrid ARQ. They maximize a network utility function while keeping the queues stable and meeting secrecy constraint in a block-fading channel. In [17], a secret key queue is deployed to secure the private information in a single-hop wireless communication system over a block fading channel while meeting the delay constraints. Similarly, fair rate allocation is studied in [18] for a secure broadcast channel where unintended receivers are considered as internal eavesdroppers. Instead of devoting all available resources to the user with the best channel condition, secure transmission of less favorable users are facilitated simultaneously by using secret keys, for the sake of fairness. Furthermore, a dynamic network control mechanism is proposed in [19] for a multi-hop wireless network while taking privacy and delay into account. They investigate the optimal rate allocation to preserve the confidentiality of the transmitted data while meeting delay constraints.

### 3. System model

In this paper, we focus on the two-user cognitive radio scenario depicted in Fig. 1 where  $g_{ij}$  refers to channel coefficients. In order to apply the proposed QMPs to the overlay cognitive radio system, the achievable rate region of the setup should be determined. The achievable rate region for the overlay cognitive interference channel without secrecy is derived in [6] and [7]. As an illustrative example, the achievable rate region for a sample scenario with  $g_{11} = 1$ ,  $g_{12} = 0.5$ ,  $g_{21} = 0.6$ , and  $g_{22} = 1$ , unity noise variances, and  $P_1 = P_2 = 1$  mW is depicted in Fig. 2, which is

derived based on the cognitive interference channel model presented in [7]. The set of achievable rates is denoted with  $\mathbb{A}$ , which is essentially the set of points located below the curve in Fig. 2. Moreover, the set of all points on the curve in Fig. 2 is denoted by  $\mathbb{C}$ . In this paper, we envision a system with one PU-SU pair only (optionally along with an eavesdropper). However, the model also covers the case with multiple PU-SU pairs where each pair receives a static time-frequency resource allocation. The modeling of dynamic sharing of resources across PU-SU pairs with distributed MAC (Medium Access Control) is left for future research.

The PU permits the SU to utilize its licensed band as long as the delay constraints are not violated. A QMP controls the interaction between the primary and the secondary transmitters by means of deciding which achievable rate pair to use (and also when) for the transmission of PU and SU packets. Clearly, in order to achieve optimum performance, a QMP will try to select the rate pairs on  $\mathbb{C}$ . Any arbitrary rate pair on  $\mathbb{C}$  is denoted by  $(r_{p,c}, r_{s,c})$ . The maximum achievable rate of the PU (SU) when all of the resources of the SU are devoted to PU's (SU's) messages is denoted by  $r_{p,m}$  ( $r_{s,m}$ ), which is located at the intersection of  $\mathbb{C}$  and the  $r_1$  ( $r_2$ ) axis. The achievable rate of PU in the absence of the SU is denoted by  $r_{p,w}$ . Moreover, the corresponding SU rate of  $r_{p,w}$  on  $\mathbb{C}$  is denoted by  $r_{s,w}$ . That is, PU and SU can achieve  $r_{p,w}$  and  $r_{s,w}$  rates, respectively, if they operate simultaneously.

Now, the conventional interweave and overlay models are described:

- *Conventional Interweave Model (CIM)*: The cognitive user does not cooperate with the PU. It transmits its messages during the inactivity times of PU with the maximum rate  $r_{s,m}$  while trying to keep the primary transmissions unharmed [3]. Hence, the system alternates between the rate pairs  $(0, r_{s,m})$  and  $(r_{p,w}, 0)$ , depending on whether the PU's queue is empty or not, respectively.
- *Conventional Overlay Model (COM)*: The cognitive user, who has non-causal knowledge of the PU's message, cooperates with the PU. By devoting a portion of its resources to the primary message, the cognitive transmitter keeps the rate of the PU unchanged at  $r_{p,w}$ . Simultaneously, the SU transmits its own messages by the remaining of its resources, which results in the rate of  $r_{s,w}$  for the SU [3]. Nevertheless, SU devotes all of its available resources to transmit its own messages, whenever the PU's queue is empty. Thus, the system alternates between the rate pairs  $(0, r_{s,m})$  and  $(r_{p,w}, r_{s,w})$ , depending on whether the PU's queue is empty or not, respectively.

In this paper, we propose two novel QMPs for the overlay CRN with the delay-constrained PU which are described in the following.

- *Hybrid Overlay-Interweave Scheme (HOIS)*: This scheme is a combination of CIM and COM. Whenever the PU has a packet to transmit, all SU's resources are devoted to the transmission of the packet. When the PU's queue is empty, the SU transmits its own packets at the maximum rate. Thus, the system with HOIS alternates between  $(0, r_{s,m})$  and  $(r_{p,m}, 0)$  depending on whether the PU's queue is empty or not, respectively.
- *Delay-Aware Adaptive Rate Control (DAARC)*: The queue delay experienced by each packet in the primary transmitter is monitored. When a packet gets to be transmitted, its queuing delay is checked. If this delay is below a certain threshold (denoted by  $T^{(1)}$ ), i.e., in regime 1, the SU simultaneously transmits its own messages and the PU's packet at rates  $r_{s,c}$  and  $r_{p,c}$ , respectively. If the queue delay is above the threshold  $T^{(1)}$ , i.e., in regime 2, all of SU's resources are devoted to transmit only PU's packets at rate  $r_{p,m}$  while the SU's rate is zero. Whenever the primary transmitter's queue becomes empty, the SU achieves

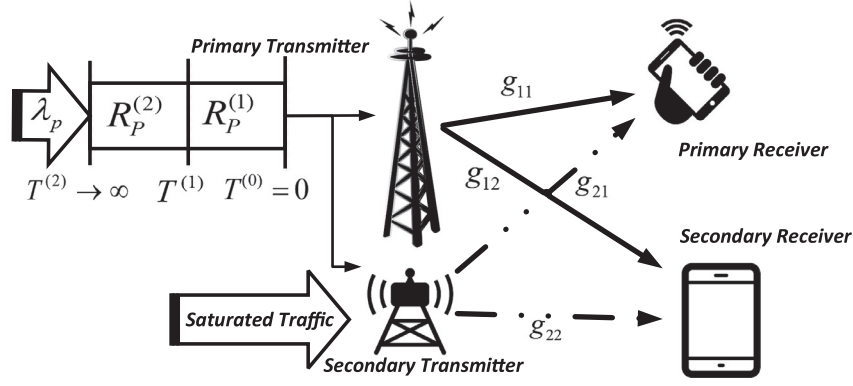
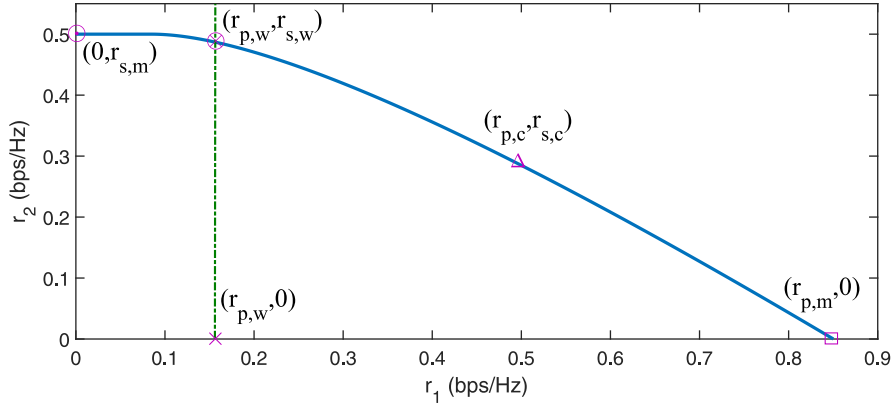


Fig. 1. Overlay cognitive radio system model.

Fig. 2. Achievable rate region of an overlay cognitive radio system for channel coefficients set as  $g_{11} = 1$ ,  $g_{12} = 0.5$ ,  $g_{21} = 0.6$ , and  $g_{22} = 1$ , unit noise variances, and power budget of the PU and SU set to  $P_1 = P_2 = 1$  mW.

the rate  $r_{s,m}$  by deploying all its resources. With the appropriate choice of the power allocation parameter  $\zeta$  (or equivalently the rate pair  $(r_{p,c}, r_{s,c})$ ) along with the choice of the threshold  $T^{(1)}$ , the performance of DAARC can be controlled. Note that when  $T^{(1)} = 0$ , all PU packets are served at a rate  $r_{p,m}$ , i.e., the system reduces to the HOIS scheme.

The PU packets are assumed to arrive at the primary transmitter according to a Poisson process with rate  $\lambda_p$  packets per milliseconds (ms), and, the length of each packet is exponentially distributed with mean  $L$ . It is assumed that the channel bandwidth is  $B$  kHz. Thus, the service time of a packet to be served at rate  $r_{i,j}$  for  $i \in \{p, s\}$  and  $j \in \{c, m, w\}$  throughout its transmission will be exponentially distributed with mean  $\frac{1}{R_{i,j}} = \frac{L}{Br_{i,j}}$  ms. Furthermore, it is assumed that the SU's traffic is saturated, i.e., it always has packets to transmit. The motivation behind the saturated traffic assumption in the SU is two-fold. First, the maximum achievable throughput for the SU occurs when SU has always traffic to send and in this way the throughput capacity of the SU can be obtained. Second, the saturated traffic decouples the primary queue from the secondary queue, which facilitates the analytical solution of the system and makes it possible to obtain closed-form expressions for the system performance metrics.

#### 4. Stochastic model for the QMPs

In CIM, COM, and HOIS, the packets at the primary transmitter are served at a fixed rate, which is equal to  $r_{p,w}$  for CIM and COM, and  $r_{p,m}$  for HOIS. Assuming a Poisson packet arrival process for the PU queue with rate  $\lambda_p$ , the queue of the PU is modeled by the well-known  $M/M/1$  queue model [20]. Since the SU transmits

only when the primary queue is empty for CIM (with rate  $r_{s,w}$ ) and HOIS (with rate  $r_{s,m}$ ), the throughput of the SU, denoted by  $S_2$ , can be expressed as follows in bits per second (bps):

$$S_2^{CIM} = (1 - \lambda_p/R_{p,w})Br_{s,w}, \quad S_2^{HOIS} = (1 - \lambda_p/R_{p,m})Br_{s,m}. \quad (1)$$

On the other hand, the SU transmits with rate  $r_{s,m}$  ( $r_{s,w}$ ) when the primary queue is empty (not empty) for COM. Thus, the SU's throughput can be expressed as follows:

$$S_2^{COM} = (1 - \lambda_p/R_{p,w})Br_{s,m} + (\lambda_p/R_{p,w})Br_{s,w}. \quad (2)$$

DAARC deploys a delay-aware service mechanism to manage the primary queue, i.e., the transmission rate of both users depends on the delay experienced by the Head of the Line (HoL) packet in the primary queue. Delay-aware mechanisms including DAARC can not be modeled with conventional queue models like  $M/M/1$ . Hence, we deploy multi-regime Markov fluid queues for modeling DAARC and deriving the steady-state distribution of the queuing delay. The following thresholds are defined:  $0 = T^{(0)} < T^{(1)} < T^{(2)} \rightarrow \infty$ . In DAARC, a service rate decision is to be made based on the delay already experienced by the packet (queue delay denoted by  $D_Q(t)$ ). Whenever  $T^{(0)} \leq D_Q(t) < T^{(1)}$ , the system is in regime 1 and the packet is to be served with rate  $R_p^{(1)} = R_{p,c}$ , and whenever  $T^{(1)} \leq D_Q(t) < T^{(2)}$ , the system is in regime 2 and the packet is to be served with rate  $R_p^{(2)} = R_{p,m}$  by devoting all of the resources of the SU to the PU's message; see also Fig. 1. Intuitively, it should hold that  $R_p^{(2)} > R_p^{(1)}$  for meeting the delay constraints.

In order to model the queuing system of the PU as a Markov fluid queue, three auxiliary random processes are defined. First, we introduce the sojourn time process  $U(t)$  which is the overall time spent in the system including service time for the packet to be served by the server. If there are no packets being served at



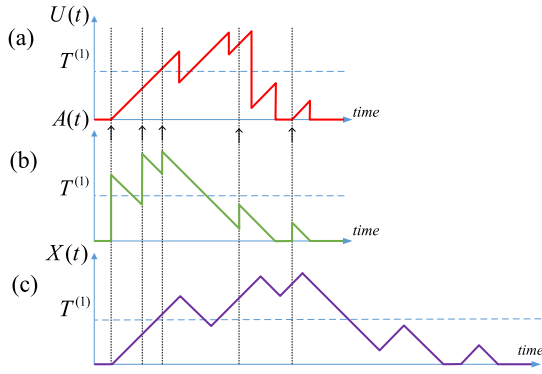


Fig. 3. Sample paths of the following processes: (a)  $U(t)$ , (b)  $A(t)$ , and (c)  $X(t)$ .

time  $t$ , then  $U(t) = 0$ . Moreover, let the unfinished work process  $A(t)$  denote the amount of time needed to serve all waiting packets including the one in the service at time  $t$ . Clearly, a packet that had arrived at time  $t$  with  $T^{(0)} \leq A(t) < T^{(1)}$  ( $T^{(1)} \leq A(t) < T^{(2)}$ ) will eventually be served at rate  $R_p^{(1)}$  ( $R_p^{(2)}$ ). The sample paths for the two processes  $U(t)$  and  $A(t)$  are given in Fig. 3(a) and Fig. 3(b), respectively, for an example scenario. The packet arrival instants are indicated by the small arrows. Due to abrupt jumps, neither the sojourn time process, nor the unfinished work process are suitable to be modeled as a fluid queue [21]. Therefore, the random process  $X(t)$  is introduced, by replacing the abrupt jumps in the sojourn time process with linear decrements corresponding to a drift of minus one (see Fig. 3(c)). Moreover, it is clear from sample path arguments that the steady-state distribution of the process  $U(t)$  ( $A(t)$ ) can be derived from that of  $X(t)$  by censoring out the states corresponding to negative (positive) drifts.

We first focus on the MRMFQ model for  $X(t)$  for which we define two service states 1 and 2 during which the packets are served with rate  $R_p^{(2)}$  and  $R_p^{(1)}$ , respectively, and  $X(t)$  increases with drift equal to 1. When the service of the current job completes in states 1 and 2, the system transits into state 3. During state 3,  $X(t)$  experiences a decrease with slope one for an exponentially distributed amount of time with mean  $1/\lambda_p$ , so the delay of the new HoL packet is reduced by an amount corresponding to its inter-arrival time. If  $T^{(0)} \leq X(t) < T^{(1)}$  ( $T^{(1)} \leq X(t) < T^{(2)}$ ), the system transits from state 3 to state 2 (state 1). Also, it is possible that  $X(t)$  may hit zero while in state 3. In this case, upon the arrival of a new packet to the system, this packet is going to be served at a rate of  $R_p^{(1)}$ . Therefore, only transition from the boundary  $X(t) = 0$  occurs out of state 3 to state 2 with rate  $\lambda_p$ . In summary, the background modulating Markov system denoted by  $Z(t)$  has three different states, namely, 1, 2, and 3. Therefore, the sample path followed by  $X(t)$  can well be modeled as the modulated process of an MRMFQ, namely the process  $(X(t), Z(t))$ , with two regimes and three states. This MRMFQ is characterized by two infinitesimal generator matrices for regimes 1 and 2, denoted by  $Q^{(i)}$  for  $i = 1, 2$ , two infinitesimal generator matrices for the finite boundaries, denoted by  $\tilde{Q}^{(i)}$  for  $i = 0, 1$ , and the corresponding drift matrices which are denoted by  $R^{(i)}$  for  $i = 1, 2$ , and  $\tilde{R}^{(i)}$  for  $i = 0, 1$  for the regimes and boundaries, respectively. See [12] for an elaborate description of MRMFQs. The state transition diagrams of  $Z(t)$  are depicted in Fig. 4 for both regimes and boundary  $X(t) = 0$ .

On the basis of the above description, the infinitesimal generator matrices for the two regimes can be obtained as follows:

$$Q^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -R_p^{(1)} & R_p^{(1)} \\ 0 & \lambda_p & -\lambda_p \end{bmatrix}, \quad Q^{(2)} = \begin{bmatrix} -R_p^{(2)} & 0 & R_p^{(2)} \\ 0 & -R_p^{(1)} & R_p^{(1)} \\ \lambda_p & 0 & -\lambda_p \end{bmatrix}. \quad (3)$$

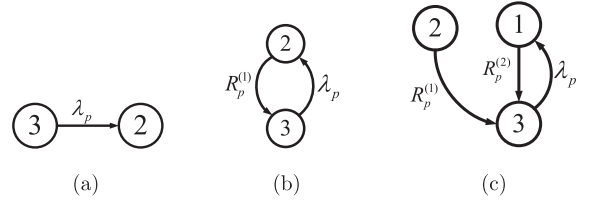


Fig. 4. State transition diagrams of  $Z(t)$  for (a) boundary  $X(t) = 0$ , (b) first regime, and (c) second regime.

Note that since  $X(t)$  increases in the service state 2, there may be transitions from state 2 to state 3 in regime 2. Furthermore, the transition matrix at threshold  $T^{(1)}$  is equal to  $Q^{(2)}$ . Also,  $\tilde{Q}^{(0)}$  is similar to  $Q^{(1)}$  except that there is no transition from state 2 to state 3 at the boundary  $X(t) = 0$ . Therefore, it holds that

$$\tilde{Q}^{(0)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \lambda_p & -\lambda_p \end{bmatrix}. \quad (4)$$

Furthermore, the drift matrices for each regime and boundary are written as  $R^{(1)} = \tilde{R}^{(1)} = R^{(2)} = \text{diag}[1, 1, -1]$  and  $\tilde{R}^{(0)} = \text{diag}[1, 1, 0]$  where **diag** denotes a diagonal matrix. The following theorem provides a closed-form expression for the steady-state distribution of queuing delay for the DAARC system of interest.

**Theorem 1.** The steady-state cumulative distribution function (CDF) of the DAARC queuing delay, denoted by  $F(x)$ , is expressed as follows:

$$F(x) = \begin{cases} c_3^{(0)}, & x = 0, \\ \frac{a_1^{(1)}}{R_p^{(1)} - \lambda_p} \left( 1 - e^{-(R_p^{(1)} - \lambda_p)x} \right), & 0 < x \leq T^{(1)}, \\ \frac{a_1^{(2)}}{R_p^{(1)}} \left( e^{-R_p^{(1)}x} - e^{-R_p^{(1)}T^{(1)}} \right) - \frac{a_1^{(1)}}{R_p^{(1)} - \lambda_p} \left( 1 - e^{-(R_p^{(1)} - \lambda_p)T^{(1)}} \right) + \frac{a_2^{(2)}}{R_p^{(2)} - \lambda_p} \left( e^{-(R_p^{(2)} - \lambda_p)x} - e^{-(R_p^{(2)} - \lambda_p)T^{(1)}} \right), & x > T^{(1)}, \end{cases} \quad (5)$$

where

$$c_3^{(0)} = \left[ \frac{R_p^{(1)} + \lambda_p}{R_p^{(1)} - \lambda_p} + \frac{2\lambda_p(R_p^{(2)} - R_p^{(1)})}{R_p^{(1)}(R_p^{(2)} - R_p^{(1)} - \lambda_p)} e^{\lambda_p T^{(1)}} - \frac{2\lambda_p(R_p^{(2)} - R_p^{(1)})(R_p^{(2)} - \lambda_p)}{(R_p^{(1)} - \lambda_p)(R_p^{(2)} - \lambda_p)(R_p^{(2)} - R_p^{(1)} - \lambda_p)} e^{-(R_p^{(1)} - \lambda_p)T^{(1)}} \right]^{-1}, \quad (6)$$

$$a_1^{(1)} = c_3^{(0)}\lambda_p, \quad a_1^{(2)} = \frac{R_p^{(2)} - R_p^{(1)} - \lambda_p}{R_p^{(2)} - R_p^{(1)}} c_3^{(0)}\lambda_p e^{\lambda_p T^{(1)}}, \quad (7)$$

$$a_2^{(2)} = \frac{-\lambda_p}{R_p^{(2)} - R_p^{(1)} - \lambda_p} c_3^{(0)}\lambda_p e^{(R_p^{(2)} - R_p^{(1)})T^{(1)}},$$

$$\Delta = c_3^{(0)} + \frac{a_1^{(1)}}{R_p^{(1)} - \lambda_p} \left( 1 - e^{-(R_p^{(1)} - \lambda_p)T^{(1)}} \right) + \frac{a_1^{(2)}}{R_p^{(1)}} e^{-R_p^{(1)}T^{(1)}} + \frac{a_2^{(2)}}{R_p^{(2)} - \lambda_p} e^{-(R_p^{(2)} - \lambda_p)T^{(1)}}.$$

**Proof.** The detailed proof is provided in Appendix A.

Eq. (5) provides a closed-form expression for the queue delay CDF of PU's packets for the DAARC QMP and this expression constitutes the main contribution of this paper. Based on this expression, the SU's throughput is calculated as follows. If the PU's queue

is empty, the SU transmits at rate  $Br_{s,m}$ . On the other hand, if the PU's queue is not empty, the secondary transmits at rate  $Br_{s,w}$  or zero, whenever the delay experienced by the HoL packet of PU is below or above  $T^{(1)}$ , respectively. Therefore,

$$S_2^{DAARC} = Br_{s,m}F(0) + Br_{s,w}(F(T^{(1)}) - F(0)). \quad (8)$$

## 5. System optimization for DAARC

Upon deriving the closed-form analytical expressions for the performance metrics of DAARC, we focus our attention to the optimum selection of the two DAARC parameters, namely the parameter  $\zeta$  which controls the primary and secondary rate pairs as illustrated in Fig. 2, and the delay threshold  $T^{(1)}$ . Our proposed QMPs are applicable to any overlay cognitive radio system. However, motivated by the increasing application of PLS in cognitive radio, a more general secure overlay cognitive radio system is considered in this paper as an example to apply the proposed QMPs. This model is inspired by the secure overlay cognitive interference channel model presented in [7] with an additional eavesdropper to take into account the security of the PU's messages as well. This secure overlay cognitive radio model is formed by introducing an external eavesdropper to the system model depicted in Fig. 1.

In this model, the delay-constrained PU intends to transmit its packets securely to both primary and secondary receivers in a timely manner in cooperation with the secondary transmitter, while keeping it secret from the external eavesdropper. On the other hand, the SU intends to deliver its messages to the secondary receiver while keeping them secret from the primary receiver and the eavesdropper. It is assumed that the secondary transmitter has non-causal knowledge about the PU's message.

In order to apply the proposed QMPs, the secure achievable rate region of the system is required, which is derived as follows. Let  $P_1$  ( $P_2$ ) be transmission power of the PU (SU) in milliwatts (mW). The SU devotes a portion of its power, i.e.,  $\zeta P_2$ , to help the PU deliver its messages whereas the remaining power, i.e.,  $(1 - \zeta)P_2$ , is used for transmitting SU's packets. Let the achievable rates of the PU and SU, be denoted by  $r_1$  and  $r_2$ , respectively, in units of bits per second per Hertz (bps/Hz). Also let

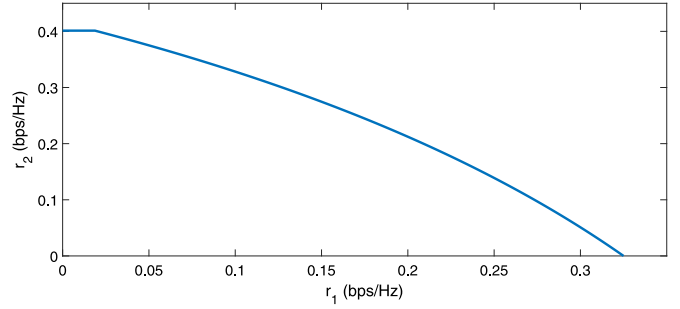
$$r_{1j} = C\left(\frac{|g_{1j}|^2 P_1 + |g_{2j}|^2 \zeta P_2}{|g_{2j}|^2 (1 - \zeta) P_2 + \sigma_j^2}\right), \quad r_{2j} = C\left(\frac{|g_{2j}|^2 (1 - \zeta) P_2}{\sigma_j^2}\right),$$

for  $j = 1, 2, 3$ , (9)

where  $C(x) = \frac{1}{2} \log(1 + x)$ ,  $g_{13}$  and  $g_{23}$  are respectively the channel coefficients between the eavesdropper and primary and secondary transmitters, and  $\sigma_j^2$  denotes the variance of the receiver noise at receiver  $j$ . Based on Bloch et al. [22], the set of achievable rates is derived in terms of the power allocation factor  $\zeta$ :

$$r_1 \leq \min\{r_{11}, r_{12}\} - r_{13}, \quad r_2 \leq r_{22} - \max\{r_{21}, r_{23}\}. \quad (10)$$

Note that in this scenario, the eavesdropper does not acquire neither PU's nor SU's messages. However, in order to consider the worst case scenario, it is assumed that the eavesdropper knows the PU's message and can cancel its interference out, while it tries to decode the SU's message. Note that the rates provided in (10) are the secure rates of the cognitive interference channel. The achievable rate for the same channel without security constraints can be derived by ignoring the presence of the illegitimate receivers, i.e., the rate of the PU is the achievable rate of the worst legitimate receiver and the rate of SU is the achievable rate of the secondary receiver. As an illustrative example, the achievable secrecy rate region which is derived by varying the value of  $\zeta$  is depicted in Fig. 5 for a sample scenario with  $g_{11} = 1$ ,  $g_{12} = 0.5$ ,  $g_{13} = 0.25$ ,  $g_{21} = 0.6$ ,  $g_{22} = 1$ , and  $g_{23} = 0.25$ , unity noise variances,  $P_1 = 1$  mW, and  $P_2 = 2$  mW.



**Fig. 5.** Achievable secrecy rate region of secure overlay cognitive radio system for channel coefficients set as  $g_{11} = 1$ ,  $g_{12} = 0.5$ ,  $g_{13} = 0.25$ ,  $g_{21} = 0.6$ ,  $g_{22} = 1$ , and  $g_{23} = 0.25$ , unit noise variances, power budget of the PU set to  $P_1 = 1$  mW, and power budget of the SU set to  $P_2 = 2$  mW.

In the cognitive radio network, the SU desires to achieve the highest throughput possible while meeting the PU delay constraint. In this section, two different versions of DAARC are considered depending on the parameter selection method.

In the first version named  $DAARC^\dagger$ , simultaneous optimization of the power allocation factor  $\zeta$  and the threshold  $T^{(1)}$  is considered. Thus, the optimization problem is the maximization of the SU's throughput while satisfying the PU's delay constraint with appropriate choices of  $\zeta$  and  $T^{(1)}$ . The closed-form expressions of the SU's throughput and PU packets' delay distribution facilitates numerical calculation of the optimum parameters by deploying standard NLP (Non-Linear Programming) methods. First, the intersection point of  $r_{11}$  with  $r_{12}$ , and  $r_{21}$  with  $r_{23}$  is determined for a given channel condition. Hence, the *min* and *max* functions in the rate region can be eliminated by dividing the region into at most 3 sub-regions. Then, the optimum point is calculated for each sub-region by using NLP. Finally, the global optimum point is obtained by comparing the optimum points in each sub-region; see Algorithm 1 for the pseudo-code for  $DAARC^\dagger$ .

---

### Algorithm 1: Pseudo-code of $DAARC^\dagger$ .

---

**Input:** Channel gains and transmission powers

**Output:**  $T^{(1)\dagger}$ ,  $\zeta^\dagger$

- 1 Find  $\zeta^{(1)}$  such that  $r_{11} = r_{12}$  and find  $\zeta^{(2)}$  such that  $r_{21} = r_{23}$ ,
- 2 Divide the support set of  $\zeta$  into 3 sub-regions:
  - a)  $S_1 = [0, \min\{\zeta^{(1)}, \zeta^{(2)}\}]$ ,
  - b)  $S_2 = [\min\{\zeta^{(1)}, \zeta^{(2)}\}, \max\{\zeta^{(1)}, \zeta^{(2)}\}]$ ,
  - c)  $S_3 = [\max\{\zeta^{(1)}, \zeta^{(2)}\}, 1]$ ,
- 3 Perform NLP in each sub-region in order to solve the following problem for  $i \in \{1, 2, 3\}$ :

$$[T_i^{(1)}, \zeta_i] = \arg \max_{T^{(1)}, \zeta} S_2^{DAARC}$$

s.t.  $\Pr\{D_q > D_T\} < \varepsilon, \zeta \in S_i$

Choose the  $[T_i^{(1)}, \zeta_i]$  pair among the three candidates that provide the maximum SU throughput as  $[T^{(1)\dagger}, \zeta^\dagger]$ .

---

Two-dimensional NLP used for  $DAARC^\dagger$  is computationally costly. Hence, a heuristic sub-optimal DAARC version is proposed, which is named as  $DAARC^\ddagger$ , to reduce the complexity of the optimization problem. This heuristic method stems from the observation that the optimal value of  $\zeta$  does not vary much beyond a certain value of  $T^{(1)}$ . In  $DAARC^\ddagger$ , the PU's delay constraint is neglected in the first step, and the optimal  $\zeta$  is calculated by deploying a one-dimensional NLP, using a sufficiently large value for  $T^{(1)}$ . Then, given the optimal  $\zeta$  and corresponding  $r_p^{(1)}$  which is

denoted by  $r_p^{(1)\ddagger}$ , the optimal value of  $T^{(1)}$  is calculated. For this purpose, we first derive the maximum value of  $T^{(1)}$ , denoted by  $\bar{T}^{(1)\ddagger}$ , that satisfies the delay constraint  $\Pr\{D_q > D_T\} = \varepsilon$ . Since the SU's throughput is an increasing function of  $T^{(1)}$  except at the origin, the optimum choice for  $T^{(1)}$  denoted by  $T^{(1)\ddagger}$  will either be  $\bar{T}^{(1)\ddagger}$  or zero, depending on whichever yields higher SU throughput; the pseudo-code of this second version is given in Algorithm 2.

---

**Algorithm 2:** Pseudo-code of  $DAARC^\ddagger$ .

---

**Input:** Channel gains and transmission powers

**Output:**  $T^{(1)\ddagger}$ ,  $\zeta^\ddagger$

- 1 Same as the steps 1–3 of Alg. 1,
- 2 Set  $T^{(1)}$  to a sufficiently large value,
- 3 Perform one-dimensional NLP just for  $\zeta$  for each sub-region in order to solve the following optimization problem for  $i \in \{1, 2, 3\}$ :

$$\zeta_i = \arg \max_{\zeta} S_2^{DAARC}$$

s.t.  $\zeta \in S_i$

Choose the  $\zeta_i$  among the three candidates that provides the maximum SU throughput as  $\zeta^\ddagger$ ,

- 4 Obtain  $r_p^{(1)\ddagger}$  corresponding to  $\zeta^\ddagger$ ,
- 5 Find the value of  $T^{(1)}$  given  $r_p^{(1)\ddagger}$  that satisfies the delay constraint with equality, i.e.,

$$\bar{T}^{(1)\ddagger} : \Pr\{D_q > D_T\} = \varepsilon$$

Choose  $T^{(1)\ddagger}$  as either 0 or  $\bar{T}^{(1)\ddagger}$  depending on whichever yields higher SU throughput:

$$T^{(1)\ddagger} = \arg \max_{T^{(1)} \in \{0, \bar{T}^{(1)\ddagger}\}} S_2^{DAARC}$$


---

## 6. Simulations and numerical results

We assume that the channel bandwidth is  $B = 1$  MHz, and  $L = 1$  kbits. The time unit is set to ms. Therefore, the packet service times are exponentially distributed with mean  $1/R_{i,j} = 1/r_{i,j}$  for  $i \in \{p, s\}$  and  $j \in \{c, m, w\}$ . The numerical results are obtained by MATLAB 2014a running on a machine with Intel core-i5 4300U CPU and 8 GB of RAM.

In order to validate the analytical method developed for DAARC, the analytical results are compared against simulations for the particular setting of  $R_p^{(1)} = r_{p,w} = 0.1327$  and  $R_p^{(2)} = r_{p,m} = 1.2266$ . In particular, the primary user transmits at rate  $r_{p,w}$  when it is operating in regime 1 and it transmits with the maximum rate  $r_{p,m}$  when it switches to regime 2. In Fig. 6, the CDF of the queuing delay in the PU queue obtained by both the analytical model as well as simulations are plotted for four different values of  $\lambda_p$ . Analytical results are in total agreement with the simulation results, which validates the analytic model. Next, we consider the performance of the four QMPs in the following four separate scenarios, associated with different channel conditions. In two of these scenarios, the channel coefficients  $g_{12}$  and  $g_{13}$  are varied such that the rate  $r_{p,w}$  becomes poor and good, respectively:

- In the *poor PU* scenario,  $g_{12} = 0.5$ ,  $g_{13} = 0.45$ ,  $g_{21} = 0.7$ .
- In the *good PU* scenario,  $g_{12} = 0.8$ ,  $g_{13} = 0.25$ ,  $g_{21} = 0.7$ .

Meanwhile, in the third and fourth scenarios, the channel coefficient  $g_{21}$  is altered to change the impact of secondary transmitter on the PU's rate:

- In the *poor SU* scenario,  $g_{12} = 0.5$ ,  $g_{13} = 0.25$ ,  $g_{21} = 0.3$ .
- In the *good SU* scenario,  $g_{12} = 0.5$ ,  $g_{13} = 0.25$ ,  $g_{21} = 0.7$ .

Note that the remaining channel coefficients are set to  $g_{11} = 1$ ,  $g_{22} = 1$ , and  $g_{23} = 0.25$  in all the four scenarios.

In the first numerical example, the four QMPs CIM, COM, HOIS, and the particular policy  $DAARC^\ddagger$  systems are studied for  $P_i = 1$  mW for  $i = 1, 2$ . The delay constraint parameters are set to  $D_T = 80, 320$  ms and  $\varepsilon = 10^{-2}$ . The NLP step in  $DAARC^\ddagger$  is conducted using the `fmincon` function in MATLAB. The throughput of the SU, while the primary's delay constraint is satisfied, is considered as the performance metric in all comparisons. This metric not only provides a tool to compare the performance of the SU in different QMPs, but also, implicitly determines the maximum admissible arrival rate for the PU traffic, which is the maximum amount of arrival rate that the PU can serve while satisfying the delay criterion.

In Fig. 7, SU throughput of  $DAARC^\ddagger$  is compared to that of CIM, COM and HOIS for the four channel scenarios. The policies CIM and COM do not cope well with the delay requirements of the PU for relatively high values of  $\lambda_p$ . Therefore, parts of their throughput curves are not depicted in Fig. 7 corresponding to situations when the PU's delay constraint is violated, i.e., the maximum admissible rate for the PU traffic in these QMPs is relatively low. Although COM outperforms HOIS for low values of  $\lambda_p$ , it can not satisfy the delay requirement of the PU for large values of  $\lambda_p$ . On the other hand,  $DAARC^\ddagger$  not only provides superior SU throughput performance, but also, it improves the maximum admissible rate of PU traffic. The SU achieves higher throughput by manipulating the delay tolerance of PU packets to its own favor in  $DAARC^\ddagger$ . As shown in Fig. 7, the achievable throughput of the SU increases in  $DAARC^\ddagger$  as the PU delay constraint is relaxed. Furthermore, the performance of  $DAARC^\ddagger$  improves with respect to HOIS as the PU channel conditions improve since the system operates more frequently in regime 1 and PU can handle its packets on its own. Note that the performance of  $DAARC^\ddagger$  reduces to HOIS as the conditions, i.e., channel gains or arrival rate, turn out to be unfavorable. In these situations, the system benefits more by transmitting each message one at a time instead of simultaneous transmission of primary and secondary messages. This situation happens when  $T^{(1)}$  is chosen as zero. This scenario is highlighted in the poor PU channel condition (Fig. 7(a)) where  $DAARC^\ddagger$  performs identical to HOIS independent of the delay tolerance of primary packets. In contrast,  $DAARC^\ddagger$  performs very similarly to COM for some specific channel gains and/or arrival rates for example when Algorithm 1 selects  $(r_p, w, r_s, w)$  as the rate pair in the first regime and  $T^{(1)} \rightarrow \infty$ . In these conditions, DAARC prefers to transmit the messages of PU and SU simultaneously.

In summary, the proposed  $DAARC^\ddagger$  method provides a robust mechanism for the PU to handle its delay sensitive traffic in a resource limited environment by leveraging SU's resources. The SU, which does not possess licensed bands, helps the PU to deliver its messages in a timely manner, in exchange for its own transmission privileges. It is shown that the SU achieves higher throughput compared to conventional methods as CIM and COM. Furthermore, as the PU delay constraint is relaxed, the SU throughput improves.

In the next step, we focus our attention to the performance analysis of  $DAARC^\ddagger$ , and comparing the results with  $DAARC^\ddagger$  which is the best performing method in the first set of numerical examples. Again, the NLP steps in  $DAARC^\ddagger$  are conducted using the `fmincon` function in MATLAB.

The two different DAARC versions are compared in Fig. 8 for  $D_T = 320$  ms and  $P_1 = P_2 = 1$  mW. Furthermore, in the second step of Algorithm 2, the sufficiently large number value is chosen as 600 ms in this example. It is illustrated that the proposed heuristic method  $DAARC^\ddagger$  performs very close to  $DAARC^\ddagger$  while requiring less computational power. In particular, the average run time of  $DAARC^\ddagger$  and  $DAARC^\ddagger$ , which are the average values of over 50 runs, are 79.5743 and 6.5552 s, respectively, which translates into more than 90% reduction in computation time with  $DAARC^\ddagger$ .

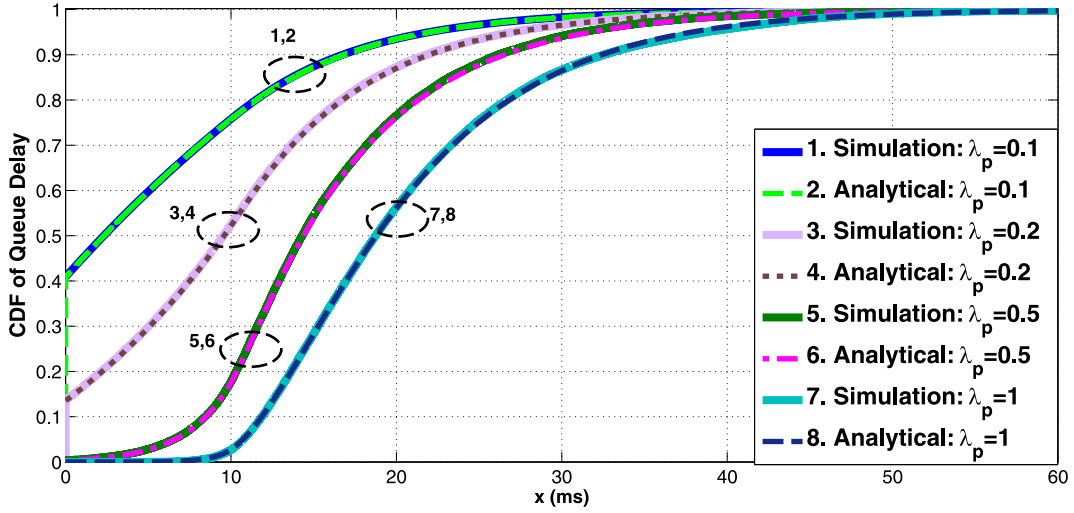


Fig. 6. CDF of the delay obtained by both simulations and the proposed analytical method when  $R_p^{(1)} = 0.1327$  and  $R_p^{(2)} = 1.2266$ .

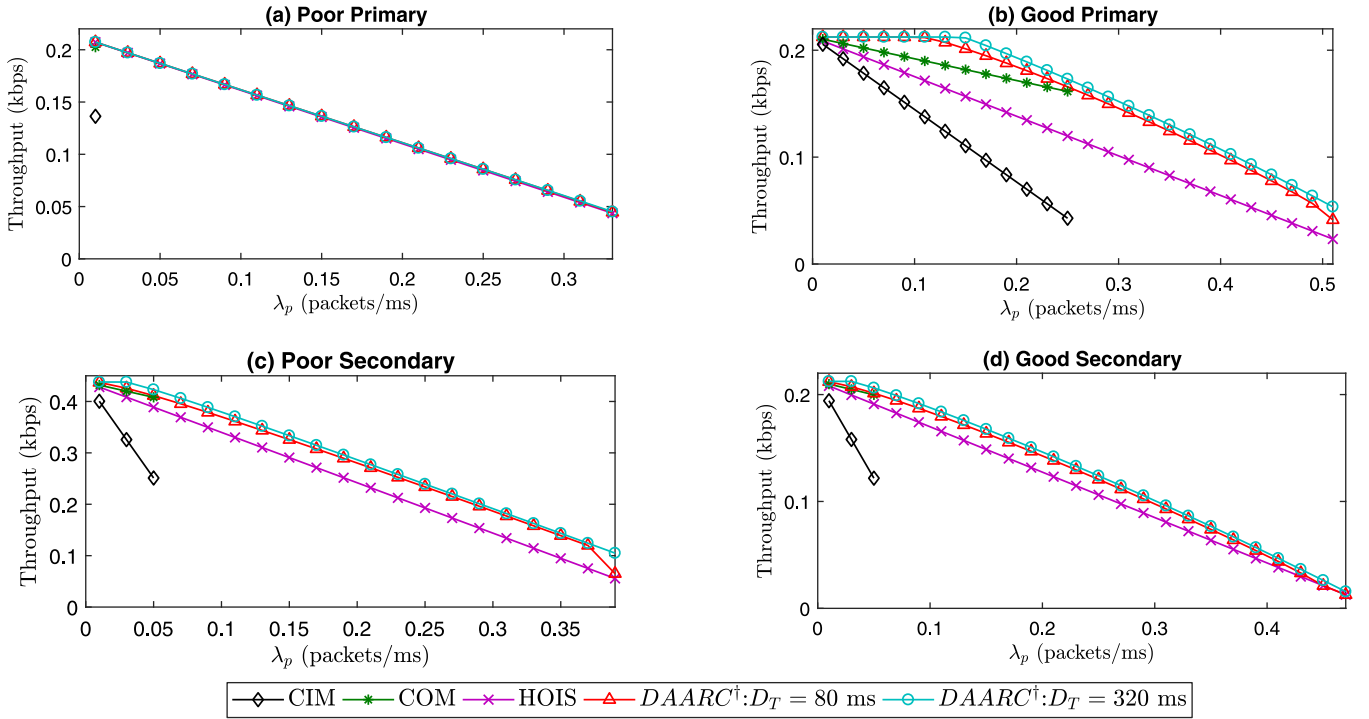


Fig. 7. SU throughput with COM, CIM, HOIS, and DAARC† for four different channel scenarios.

The secondary throughput and their relative gains with respect to HOIS (denoted by  $G^k$  for  $k \in \{DAARC^\dagger, DAARC^\ddagger\}$ ) for both versions of DAARC are given in Tables 2–4, for good primary, poor secondary and good secondary scenarios, respectively. As expected, DAARC† appears to be superior to all the other methods. The gain of DAARC† depends on the channel conditions and the arrival rate of the PU traffic. We also observe from Tables 2–4 that the gain of DAARC† grows consistently as the arrival rate increases and this gain reaches over 300% for relatively high values of  $\lambda_p$  in the poor secondary scenario. We also observe that DAARC† performance is very close to DAARC‡ for most of the cases, which makes it an attractive choice considering its significantly lower computational complexity.

Note that for poor PU channel conditions, both algorithms perform identically since the optimization is conducted for the parameters of the first regime which is practically not used due to very

Table 2

The SU throughput (in units of kbps) obtained by HOIS and the two versions of DAARC with  $D_T = 320$  ms in the good primary scenario. The percentage throughput gain of the two DAARC versions with respect to HOIS are also provided.

$\lambda_p$	$S_2^{HOIS}$	$S_2^{DAARC^\dagger}$	$S_2^{DAARC^\ddagger}$	$G^{DAARC^\dagger}(\%)$	$G^{DAARC^\ddagger}(\%)$
0.09	0.1740	0.2123	0.2123	22.0115	22.0115
0.13	0.1642	0.2123	0.2123	29.2935	29.2935
0.19	0.1419	0.1968	0.1968	38.6892	38.6892
0.29	0.1049	0.1564	0.1564	49.0944	49.0944
0.31	0.0975	0.1478	0.1478	51.6053	51.6053
0.39	0.0678	0.1120	0.1120	65.0943	65.0943
0.49	0.0308	0.0638	0.0637	106.9178	106.7230

low performing PU. That is, in poor PU channel conditions, both algorithms prefer to devote all available resources to empty the



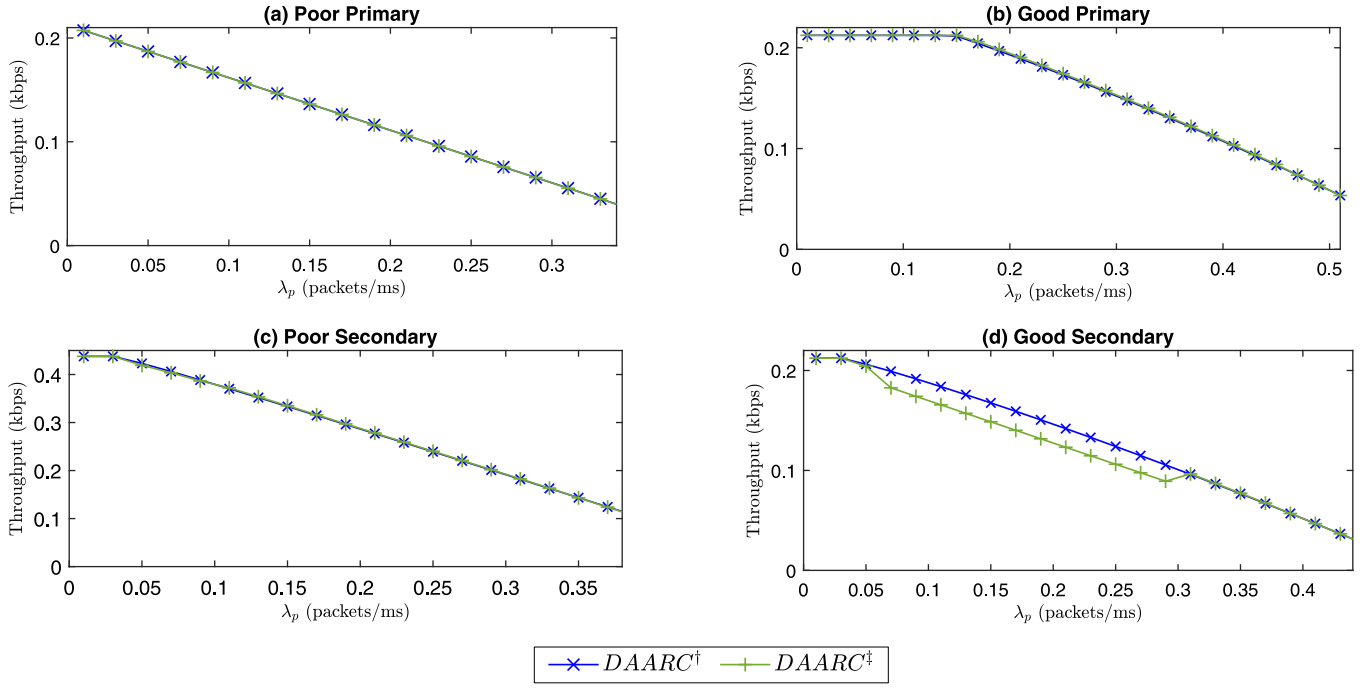


Fig. 8. Comparison of the two different versions of DAARC QMP for  $D_T = 320$  ms.

Table 3

The SU throughput (in units of kbps) obtained by HOIS and the two versions of DAARC with  $D_T = 320$  ms in the poor secondary scenario. The percentage throughput gain of the two DAARC versions with respect to HOIS are also provided.

$\lambda_p$	$S_2^{HOIS}$	$S_2^{DAARC^+}$	$S_2^{DAARC^+}$	$G^{DAARC^+}(\%)$	$G^{DAARC^+}(\%)$
0.01	0.4280	0.4378	0.4378	2.2897	2.2897
0.05	0.3888	0.4231	0.4191	8.8220	7.7932
0.09	0.3496	0.3885	0.3867	11.1270	10.6121
0.11	0.3300	0.3704	0.3704	12.2424	12.2424
0.13	0.3104	0.3521	0.3521	13.4343	13.4343
0.17	0.2712	0.3148	0.3148	16.0767	16.0767
0.21	0.2320	0.2771	0.2771	19.4397	19.4397
0.29	0.1536	0.2011	0.2011	30.9245	30.9245
0.39	0.0557	0.1050	0.1044	88.6793	87.6011
0.43	0.0165	0.0664	0.0582	303.3435	253.8602

Table 4

The SU throughput (in units of kbps) obtained by HOIS and the two versions of DAARC with  $D_T = 320$  ms in the good secondary scenario. The percentage throughput gain of the two DAARC versions with respect to HOIS are also provided.

$\lambda_p$	$S_2^{HOIS}$	$S_2^{DAARC^+}$	$S_2^{DAARC^+}$	$G^{DAARC^+}(\%)$	$G^{DAARC^+}(\%)$
0.01	0.2081	0.2123	0.2123	2.0183	2.0183
0.05	0.1911	0.2062	0.2043	7.9016	6.9073
0.09	0.1741	0.1916	0.1741	10.0517	0
0.11	0.1656	0.1838	0.1656	10.9903	0
0.13	0.1571	0.1758	0.1571	11.9032	0
0.17	0.1401	0.1592	0.1401	13.6331	0
0.27	0.0977	0.1148	0.0977	17.5266	0
0.37	0.0552	0.0669	0.0669	21.1918	21.1918
0.47	0.0127	0.0157	0.0156	23.1132	22.9560

PU queue first and then assign the remaining resources to the SU. Therefore, the results associated with the poor channel scenario are deliberately not tabulated.

## 7. Conclusions

Two novel QMPs are proposed for an overlay cognitive radio network with delay-constrained PU. The proposed QMPs attempt to maximize the SU throughput while meeting the PU's delay criterion. The first QMP, named HOIS, is a hybrid interweave/overlay model, while the second QMP, called DAARC, deploys a delay-aware adaptive mechanism. Through a novel MRFMQ based model, the closed-form expressions for the exact delay distribution of the PU traffic are derived and validated for DAARC. Moreover, analytical expressions are employed to optimally tune the parameters of DAARC. The proposed QMPs are applied to a secure overlay CRN with delay-constrained PU which deploys physical layer security to keep its messages confidential. The proposed methods are simulated and compared with the conventional interweave and overlay paradigms using this secure overlay CRN. It is shown that by intelligent manipulation of the PU's delay constraint, DAARC consistently outperforms HOIS as well as CIM and COM with the performance gap depending on the channel conditions and the arrival rate to the PU queue. The computational complexity of finding off-line the optimum DAARC parameters and the requirement to keep track of the queuing delay of each packet towards the on-line implementation appear to be the drawbacks of DAARC. More general traffic models for PU and SU are left for future research.

## Appendix A. Proof of Theorem 1

The joint probability density function (pdf) vector of  $X(t)$  for regime  $k$  for  $k = 1, 2$ , when  $T^{(k-1)} \leq D_Q(t) < T^{(k)}$  is defined as follows:

$$f_i^{(k)}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} \Pr \{X(t) \leq x, Z(t) = i\},$$

$$f^{(k)}(x) = (f_1^{(k)}(x), f_2^{(k)}(x), f_3^{(k)}(x)). \quad (A.1)$$

Similarly, the steady state mass accumulation vector for boundary points  $T^{(0)}$  and  $T^{(1)}$  is defined as follows:

$$c_i^{(k)} = \lim_{t \rightarrow \infty} \Pr \{X(t) = T^{(k)}, Z(t) = i\}, \quad c^{(k)} = (c_1^{(k)}, c_2^{(k)}, c_3^{(k)}). \quad (A.2)$$

Following the same procedure in [12], the following set of differential equations holds for the joint pdf vector of  $X(t)$ :

$$\frac{d}{dx} f^{(i)}(x) R^{(i)} = f^{(i)}(x) Q^{(i)}, \text{ for } i = 1, 2. \quad (\text{A.3})$$

In order to derive the boundary conditions, we first identify the set of states with positive, negative and zero drift for each regime and boundary.  $S_+^{(k)}$ ,  $S_-^{(k)}$ , and  $S_0^{(k)}$  represent the states with positive, negative, and zero drift at regime  $k$ , respectively, and  $\tilde{S}_+^{(k)}$ ,  $\tilde{S}_-^{(k)}$ , and  $\tilde{S}_0^{(k)}$  represent the states with positive, negative, and zero drift at boundary  $k$ , respectively. According to aforementioned infinitesimal generator matrices and drift matrices of the system model, these sets are populated as follows:

$$\begin{aligned} S_+^{(1)} &= \{1, 2\}, & S_-^{(1)} &= \{3\}, & S_0^{(1)} &= \emptyset, \\ S_+^{(2)} &= \{1, 2\}, & S_-^{(2)} &= \{3\}, & S_0^{(2)} &= \emptyset, \\ \tilde{S}_+^{(0)} &= \{1, 2\}, & \tilde{S}_-^{(0)} &= \emptyset, & \tilde{S}_0^{(0)} &= \emptyset, \\ \tilde{S}_+^{(1)} &= \{1, 2\}, & \tilde{S}_-^{(1)} &= \{3\}, & \tilde{S}_0^{(1)} &= \emptyset. \end{aligned}$$

Hence, the intermediate boundary point  $T^{(1)}$  is an emitting state [12] and the boundary conditions for the differential equations presented in (A.3) are as follows:

$$c_1^{(0)} = c_2^{(0)} = 0, \quad (\text{A.4a})$$

$$c_m^{(1)} = 0, \text{ for } m \in \{1, 2, 3\}, \quad (\text{A.4b})$$

$$f^{(1)}(0+)R^{(1)} = c^{(0)}\tilde{Q}^{(0)}, \quad (\text{A.4c})$$

$$f^{(2)}(T^{(1)}+)R^{(2)} - f^{(1)}(T^{(1)}-)R^{(1)} = c^{(1)}\tilde{Q}^{(1)}, \quad (\text{A.4d})$$

$$\left( \sum_{k=1}^2 \int_{T^{(k-1)}+}^{T^{(k)}-} f^{(k)}(x) dx + \sum_{k=0}^1 c^{(k)} \right) (1, 1, 1)^T = 1. \quad (\text{A.4e})$$

As the boundary values for the differential equations are not known yet, this problem should be considered as a boundary value problem. The spectral solution to MRMFQs are presented in [12]. For our model, where there is no state with zero drift neither in the regimes, nor in the boundaries, the general solution to (A.3) is given as follows:

$$f^{(k)}(x) = \sum_i a_i^{(k)} e^{\lambda_i^{(k)} x} \phi_i^{(k)}, \text{ for } T^{(k-1)} < x < T^{(k)}, 1 \leq k \leq 2 \quad (\text{A.5})$$

where  $(\lambda_i^{(k)}, \phi_i^{(k)})$  is the  $i$ th eigenvalue-left eigenvector pair of the matrix  $Q^{(k)}(R^{(k)})^{-1}$  [23] which are derived for the first and second regime as follows:

$$\det(\lambda I - Q^{(1)}(R^{(1)})^{-1}) = 0, \quad \det(\lambda I - Q^{(2)}(R^{(2)})^{-1}) = 0,$$

which result in:

$$(\lambda_1^{(1)}, \phi_1^{(1)}) = (\lambda_p - R_p^{(1)}, (0, 1, 1)),$$

$$(\lambda_2^{(1)}, \phi_2^{(1)}) = \left( 0, (0, 1, \frac{R_p^{(1)}}{\lambda_p}) \right),$$

$$(\lambda_3^{(1)}, \phi_3^{(1)}) = (0, (1, 0, 0)),$$

and

$$(\lambda_1^{(2)}, \phi_1^{(2)}) = \left( -R_p^{(1)}, \left( \frac{\lambda_p}{R_p^{(2)} - R_p^{(1)}}, \frac{R_p^{(2)} - R_p^{(1)} - \lambda_p}{R_p^{(2)} - R_p^{(1)}}, 1 \right) \right),$$

$$(\lambda_2^{(2)}, \phi_2^{(2)}) = (\lambda_p - R_p^{(2)}, (1, 0, 1)),$$

$$(\lambda_3^{(2)}, \phi_3^{(2)}) = \left( 0, \left( \frac{\lambda_p}{R_p^{(2)}}, 0, 1 \right) \right).$$

Hence, the pdf of the delay for the first and second regime is derived as follows:

$$f^{(1)}(x) = a_1^{(1)} e^{\lambda_1^{(1)} x} \phi_1^{(1)} + a_2^{(1)} \phi_2^{(1)} + a_3^{(1)} \phi_3^{(1)},$$

$$f^{(2)}(x) = a_1^{(2)} e^{\lambda_1^{(2)} x} \phi_1^{(2)} + a_2^{(2)} e^{\lambda_2^{(2)} x} \phi_2^{(2)} + a_3^{(2)} \phi_3^{(2)}.$$

In order to solve the differential equation, there are 6 unknown  $a$  coefficients, and 6 unknown  $c$  coefficients that should be determined using the boundary conditions. Since the queue model is of infinite size, the stability condition should also be satisfied for the second regime as  $\pi^{(2)} R^{(2)}(1, 1, 1)^T < 0$  where  $\pi^{(2)}$  is the steady state vector of  $Q^{(2)}$ . Furthermore, in order to acquire a bounded distribution for the second regime, the coefficients corresponding to the zero eigenvalues and the eigenvalues in the right half plane must be equal to zero. To be precise, imposing (A.4a) and (A.4b) results in:

$$c_1^{(0)} = c_2^{(0)} = c_1^{(1)} = c_2^{(1)} = c_3^{(1)} = a_2^{(1)} = a_3^{(1)} = a_3^{(2)} = 0.$$

Then, the Eq. (A.4c) yields the following:

$$\begin{aligned} (a_1^{(1)} \phi_1^{(1)} + a_2^{(1)} \phi_2^{(1)} + a_3^{(1)} \phi_3^{(1)}) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \\ = (0, 0, c_3^{(0)}) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \lambda_p & -\lambda_p \end{bmatrix} \end{aligned}$$

which implies that

$$a_1^{(1)} = c_3^{(0)} \lambda_p, \quad a_2^{(1)} = 0, \quad a_3^{(1)} = 0.$$

In order to derive a bounded distribution,  $a_3^{(2)}$  should be equal to zero, since  $a_3^{(2)}$  is the coefficient associated with the eigenvalue at the origin in the last regime. Hence, employing (A.4d) results in:

$$\begin{aligned} a_1^{(2)} e^{\lambda_1^{(2)} T^{(1)}} \phi_1^{(2)} + a_2^{(2)} e^{\lambda_2^{(2)} T^{(1)}} \phi_2^{(2)} + a_3^{(2)} \phi_3^{(2)} - a_1^{(1)} e^{\lambda_1^{(1)} T^{(1)}} \phi_1^{(1)} \\ = (0, 0, 0) \begin{bmatrix} -R_p^{(2)} & 0 & R_p^{(2)} \\ 0 & -R_p^{(1)} & R_p^{(1)} \\ \lambda_p & 0 & -\lambda_p \end{bmatrix} = 0 \end{aligned}$$

which implies that

$$a_1^{(2)} = \left( \frac{R_p^{(2)} - R_p^{(1)} - \lambda_p}{R_p^{(2)} - R_p^{(1)}} \right) c_3^{(0)} \lambda_p e^{(\lambda_1^{(1)} - \lambda_1^{(2)}) T^{(1)}},$$

$$a_2^{(2)} = \left( \frac{-\lambda_p}{R_p^{(2)} - R_p^{(1)} - \lambda_p} \right) c_3^{(0)} \lambda_p e^{(\lambda_1^{(1)} - \lambda_2^{(2)}) T^{(1)}},$$

$$a_3^{(2)} = 0.$$

Furthermore, the stability condition is imposed as  $\lambda_p < R_p^{(2)}$ . Considering the stability condition and  $R_p^{(1)} > 0$ , we replace all unknown variables in  $c_3^{(0)}$  in Eq. (A.4e) and derive  $c_3^{(0)}$  as in Eq. (6). Moreover, we derive the closed-form expression for the  $a_1^{(1)}$ ,  $a_1^{(2)}$ , and  $a_2^{(2)}$  coefficients as in (7) and for the remaining coefficients as follows:

$$a_2^{(1)} = a_3^{(1)} = a_3^{(2)} = 0,$$

which give the joint pdf vectors as follows:

$$\begin{aligned} f^{(1)}(x) &= a_1^{(1)} e^{(\lambda_p - R_p^{(1)})x} (0, 1, 1), \\ f^{(2)}(x) &= a_1^{(2)} e^{-R_p^{(1)}x} \left( \frac{\lambda_p}{R_p^{(2)} - R_p^{(1)}}, \frac{R_p^{(2)} - R_p^{(1)} - \lambda_p}{R_p^{(2)} - R_p^{(1)}}, 1 \right) \\ &\quad + a_2^{(2)} e^{(\lambda_p - R_p^{(2)})x} (1, 0, 1). \end{aligned}$$

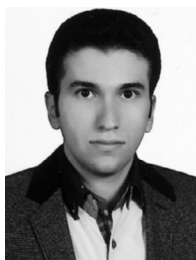
Since the unfinished work process  $A(t)$  determines the amount of delay that newly arriving jobs (which arrive to the system according to a Poisson process) will experience, the steady-state probability distribution of state 3 can be used to obtain the quantities of interest, by a direct consequence of the PASTA (Poisson Arrivals See Time Averages) property. Therefore, in order to obtain the steady-state distribution of  $A(t)$  from that of the fluid process  $X(t)$ , we censor out the states 1 and 2, and subsequently normalize the steady-state distributions. In mathematical terms, we calculate the steady-state CDF of  $A(t)$ , which is equal to the distribution of the queue delay of a newly arrived packet, from that of  $(X(t), Z(t))$  as follows:

$$F(x) = \lim_{t \rightarrow \infty} \Pr\{A(t) \leq x\} = \lim_{t \rightarrow \infty} \frac{\Pr\{Z(t) = 3, X(t) \leq x\}}{\Pr\{Z(t) = 3\}} \quad (\text{A.6})$$

which together with the expressions obtained for the joint pdf vector gives the identity in (5).

## References

- [1] FCC Spectrum Policy Task Force, Report of the Spectrum Efficiency Working Group, 2002
- [2] J. Mitola, G.Q. Maguire, Cognitive radio: making software radios more personal, *IEEE Personal Commun.* 6 (4) (1999) 13–18, doi:10.1109/98.788210.
- [3] A. Goldsmith, S.A. Jafar, I. Maric, S. Srinivasa, Breaking spectrum gridlock with cognitive radios: an information theoretic perspective, *Proc. IEEE* 97 (5) (2009) 894–914, doi:10.1109/jproc.2009.2015717.
- [4] L. Musavian, S. Aïssa, S. Lambetharan, Effective capacity for interference and delay constrained cognitive radio relay channels, *IEEE Trans. Wireless Commun.* 9 (5) (2010) 1698–1707, doi:10.1109/TCOMM.2010.05.090600.
- [5] S.A.R. Zaidi, C.D. McLernon, M. Ghogho, M.A. Imran, Cloud empowered cognitive inter-cell interference coordination for small cellular networks, in: 2015 IEEE International Conference on Communication Workshop (ICCW), 2015, pp. 2218–2224. doi: 10.1109/ICCW.2015.7247511.
- [6] I. Marić, A. Goldsmith, G. Kramer, S. Shamaï, On the capacity of interference channels with one cooperating transmitter, *Eur. Trans. Telecommun.* 19 (4) (2008) 405–420, doi:10.1002/ett.1298.
- [7] Y. Liang, A. Somekh-Baruch, H.V. Poor, S. Shamaï, S. Verdú, Capacity of cognitive interference channels with and without secrecy, *IEEE Trans. Inf. Theory* 55 (2) (2009) 604–619, doi:10.1109/tit.2008.2009584.
- [8] D. Maamari, D. Tuninetti, N. Devroye, Multi-user cognitive interference channels: a survey and new capacity results, *IEEE Trans. Cognit. Commun. Netw.* 1 (1) (2015) 29–44.
- [9] Y. Yang, S. Aïssa, K.N. Salama, Spectrum band selection in delay-qos constrained cognitive radio networks, *Veh. Technol. IEEE Trans.* 64 (7) (2015) 2925–2937, doi:10.1109/TVT.2014.2354076.
- [10] L. Ma, Y. Ma, P. Ma, Delay-qos-driven secrecy power allocation in underlay secure cognitive radio system, in: 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), IEEE, 2016, pp. 1–5. doi: 10.1109/VTCSpring.2016.7504188.
- [11] A.M. Elmahdy, A. El-Keyi, T. ElBatt, K.G. Seddik, Optimizing cooperative cognitive radio networks performance with primary qos provisioning, *IEEE Trans. Commun.* 65 (4) (2017) 1451–1463, doi:10.1109/TCOMM.2016.2621744.
- [12] H.E. Kankaya, N. Akar, Solving multi-regime feedback fluid queues, *Stochastic Models* 24 (3) (2008) 425–450, doi:10.1080/15326340802232285.
- [13] A. Yener, S. Ulukus, Wireless physical-layer security: lessons learned from information theory, *Proc. IEEE* 103 (10) (2015) 1814–1825, doi:10.1109/JPROC.2015.2459592.
- [14] D. Wu, R. Negi, Effective capacity: a wireless link model for support of quality of service, *IEEE Trans. Wireless Commun.* 24 (5) (2003) 630–643, doi:10.1109/TWC.2003.814353.
- [15] C.E. Koksal, O. Ercetin, Y. Sarikaya, Control of wireless networks with secrecy, *IEEE/ACM Trans. Netw.* 21 (1) (2013) 324–337, doi:10.1109/TNET.2012.2197410.
- [16] Y. Sarikaya, O. Ercetin, C.E. Koksal, Confidentiality-preserving control of uplink cellular wireless networks using hybrid ARQ, *IEEE/ACM Trans. Netw.* 23 (5) (2015) 1457–1470, doi:10.1109/TNET.2014.2331077.
- [17] Z. Mao, C.E. Koksal, N.B. Shroff, Achieving full secrecy rate with low packet delays: an optimal control approach, *IEEE J. Sel. Areas Commun.* 31 (9) (2013) 1944–1956, doi:10.1109/jsac.2013.130925.
- [18] Z. Mao, C.E. Koksal, N.B. Shroff, Fair rate allocation for broadcast channel with confidential messages, in: 2013 IEEE Global Conference on Signal and Information Processing, 2013, pp. 799–802. Doi: 10.1109/GlobalSIP.2013.6737012.
- [19] Y. Sarikaya, C.E. Koksal, O. Ercetin, Dynamic network control for confidential multi-hop communications, *IEEE/ACM Trans. Netw.* 24 (2) (2016) 1181–1195, doi:10.1109/TNET.2015.2414945.
- [20] D. Gross, C.M. Harris, *Fundamentals of Queueing Theory*, third ed., Wiley, New York, 1998.
- [21] C. Tunc, N. Akar, Performance modeling of delay-based dynamic speed scaling, in: 9th International Conference on Matrix-Analytic Methods in Stochastic Models, Budapest, Hungary, 2016, pp. 7–13.
- [22] M. Bloch, J. Barros, M.R.D. Rodrigues, S.W. McLaughlin, Wireless information-theoretic security, *IEEE Trans. Inf. Theory* 54 (6) (2008) 2515–2534, doi:10.1109/TIT.2008.921908.
- [23] W. Scheinhardt, M. Mandjes, N. Van Foreest, Continuous feedback fluid queues, *Operat. Res. Lett.* 33 (6) (2005) 551–559, doi:10.1016/j.orl.2004.11.008.



**Kamal Adli Mehr** received the B.S. degree in electrical engineering and the M.S. degree in communication systems from the University of Tabriz, Tabriz, Iran, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include wireless communication systems, cognitive radio, performance evaluation of communication systems, physical layer security, network security, and cryptography.



**Javad Musevi Niya** was born in Ahar, Iran. He received the B.Sc. degree from the University of Tehran, the M.Sc. degree from the Sharif University of Technology, and the Ph.D. degree in communications from the University of Tabriz. Since 2006, he has been with the Faculty of Electrical and Computer Engineering, University of Tabriz. His current research interests include wireless communication systems, multimedia networks, and signal processing for communication systems.



**Nail Akar** received his B.S. degree from Middle East Technical University, Turkey, in 1987 and M.S. and Ph.D. degrees from Bilkent University, Ankara, Turkey, in 1989 and 1994, respectively, all in electrical and electronics engineering. From 1994 to 1996, he was a visiting scholar and a visiting assistant professor in the Computer Science Telecommunications program at the University of Missouri - Kansas City, USA. He joined the Technology Planning and Integration group at Long Distance Division, Sprint, Overland Park, Kansas, in 1996, where he held a senior member of technical staff position from 1999 to 2000. Since 2000, he has been with Bilkent University, Turkey, currently as a Professor, at the Electrical and Electronics Engineering Department. He visited the School of Computing, University of Missouri - Kansas City, as a Fulbright scholar in 2010 for a period of six months. His current research interests include performance analysis of computer and communication systems and networks, performance evaluation tools and methodologies, modeling, designing, and engineering of optical and wireless networks, queuing systems, and resource allocation.