

Cooperative Access in Cognitive Radio Networks: Stable Throughput and Delay Tradeoffs

Mahmoud Ashour[†], Amr A. El-Sherif[¶], Tamer ElBatt^{*‡} and Amr Mohamed[†]

[†] Computer Science and Engineering Department, Qatar University, Doha, Qatar.

[¶] Dept. of Electrical Engineering, Alexandria University, Alexandria 21544, Egypt.

^{*} Wireless Intelligent Networks Center (WINC), Nile University, Giza, Egypt.

[‡] EECE Dept., Faculty of Engineering, Cairo University, Giza, Egypt.

{m.ashour@qu.edu.qa, amr.elsherif, telbatt, amrm@ieee.org}

Abstract—In this paper, we study and analyze fundamental throughput-delay tradeoffs in cooperative multiple access for cognitive radio systems. We focus on the class of randomized cooperative policies, whereby the secondary user (SU) serves either the queue of its own data or the queue of the primary user (PU) relayed data with certain service probabilities. The proposed policy opens room for trading the PU delay for enhanced SU delay. Towards this objective, stability conditions for the queues involved in the system are derived. Furthermore, a moment generating function approach is employed to derive closed-form expressions for the average delay encountered by the packets of both users. Results reveal that cooperation expands the stable throughput region of the system and significantly reduces the delay at both users. Moreover, we quantify the gain obtained in terms of the SU delay under the proposed policy, over conventional relaying that gives strict priority to the relay queue.

Index Terms—Cognitive relaying, moment generating function, stable throughput region, average delay.

I. INTRODUCTION

Spectrum scarcity coupled with its under-utilization [1] stimulated the introduction of cognitive radios [2], [3]. The main idea of cognitive radios resides in introducing secondary users (SUs) capable of sensing the spectrum and exploiting periods in which primary users (PUs) are idle. Due to the broadcast nature of wireless channels, cooperative communication in wireless networks has been widely investigated [4], [5]. Incorporating cooperation into cognitive radio networks, the SUs not only seek idle time slots to transmit their own data, but they may also relay the PUs' packets. Thus, cooperation in cognitive radio networks can be viewed as a win-win situation. The SUs help the PUs deliver their packets to the destination. This helps in fulfilling the demand of the PUs and, hence, increases the availability of slots in which SUs can transmit their own packets.

Cooperative communication can be also viewed as a way of implementing the notion of spatial diversity. Analogous to using multiple antennas to achieve spatial diversity in single communication links [6], [7], the resources of multiple nodes can be exploited to induce a similar effect. Many works addressed cooperative communication from a physical layer perspective, such as [4], [5], [8]. However, we are interested

in the implementation of cooperation at higher network layers. For instance, in [9], the PU leases its own bandwidth for a fraction of time to a secondary network in exchange of appropriate gains attributed to cooperation. In [10], protocol-level cooperation is implemented among N nodes in a wireless network, whereby each node is a source and a prospective relay at the same time. In [11], two protocols are developed to implement cooperation in a system of M source terminals, a single destination, and a single cognitive relay. While in [12], multiple protocols which allow cooperation between a PU and a set of SUs are analyzed.

Perhaps the closest to our work is [10] which presents delay analysis for a cognitive relaying scenario in which full priority is given to the relay queue. In this paper, unlike [10], our prime objective is to develop a mathematical framework for the class of randomized cooperative policies that open room for accommodating cognitive radio systems supporting real-time applications. Towards this objective, we propose and analyze a tunable randomized cooperative policy, whereby the SU serves either the queue of its own data with probability (w.p.) a , or the relay queue w.p. $(1 - a)$. The proposed policy is shown to enhance the SU delay at the expense of a slight degradation in the PU delay. The significance of the proposed policy lies in its tunability, whereby a variety of objectives could be realized via performing constrained optimizations over the service probability a . The main contributions of this work are summarized as follows:

- 1) We propose a randomized cooperative policy that enables trading the PU delay for enhanced SU delay, depending on the application and system QoS constraints.
- 2) The stable throughput region of the system is derived. Moreover, we derive closed-form expressions for the average delay experienced by the packets of both users. Furthermore, the effect of varying a on the system's throughput and delay is thoroughly investigated.
- 3) Extensive simulations are conducted to validate the obtained analytical results.
- 4) We study the fundamental throughput-delay tradeoff at both users. At any given point within the stability region of the system, the optimal value of a that minimizes the average delay for the PU and SU is analytically derived.

This paper was made possible by a NPRP grant 4-1034-2-385 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

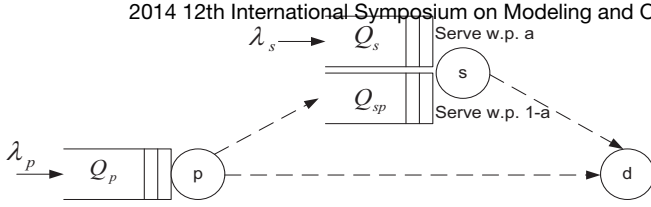


Fig. 1: Cognitive radio network under consideration.

The rest of this paper is organized as follows. Section II presents the system model along with the implemented cooperation strategy. Section III presents the derivation of the stable throughput region of the system. The average delay characterization of the system is provided in section IV. Numerical results are then presented in section V. Finally, concluding remarks are drawn in section VI.

II. SYSTEM MODEL

We consider the cognitive radio system shown in Fig. 1. The system comprises a PU and a SU equipped with infinite capacity buffers, transmitting their packets to a destination d . Time is slotted, and the transmission of a packet takes exactly one time slot. Source burstiness is taken into account through modelling the arrivals at the PU and SU as Bernoulli processes with rates λ_p and λ_s , respectively. The arrival processes at both users are independent of each other, and are independent and identically distributed (i.i.d) across time slots.

We assume that the SU performs perfect sensing. Thus, the system is contention-free, since at most one user is allowed to transmit in a given slot. Hence, the only reason for packet loss is the channel outage event, which is defined as having a signal-to-noise ratio (SNR) at the receiving node below a certain threshold. Let f_{pd} , f_{sd} , and f_{ps} denote the probability of no link outage between the PU and destination, the SU and destination, and the PU and SU, respectively.

A. Queueing Model

There are three queues involved in the system analysis, as shown in Fig. 1. They are described as follows:

- Q_p : stores the packets of the PU corresponding to the external Bernoulli arrival process with rate λ_p .
- Q_{sp} : stores the packets at the SU relayed from the PU.
- Q_s : stores the packets of the SU corresponding to the external Bernoulli arrival process with rate λ_s .

The instantaneous evolution of queue lengths is captured as

$$Q_k^{t+1} = [Q_k^t - Y_k^t]^+ + X_k^t, \quad k \in \{p, sp, s\} \quad (1)$$

where $[x]^+ = \max(x, 0)$, and Q_k^t denotes the number of packets in the k th queue at the beginning of the t th time slot. The binary random variables taking values either 0 or 1, Y_k^t and X_k^t , denote the departures and arrivals corresponding to the k th queue in the t th time slot, respectively.

B. Cooperation Strategy

The proposed cooperative scheme is described as follows:

- 1) The PU transmits a packet whenever Q_p is non empty.

- 2) If the packet is successfully decoded by the destination, it exits the system.
- 3) If the packet is not successfully received by the destination, yet, successfully decoded by the SU, the packet is buffered in Q_{sp} and is dropped from Q_p . It becomes the responsibility of the SU to deliver to the destination.
- 4) If both the destination and the SU fail to decode the packet, it is kept at Q_p for retransmission in the next time slot.
- 5) When the PU is idle, the SU transmits a packet from Q_s w.p. a , or a packet from Q_{sp} w.p. $(1 - a)$.
- 6) If the packet is successfully decoded by the destination, it exits the system. Otherwise, it is kept at its queue for later retransmission.

It is worth noting from the description of the proposed policy that the system at hand is non work-conserving. A system is considered work-conserving if it does not idle whenever it has packets [13]. However, in our system, one case violates this condition, which arises when the SU detects a slot in which the PU is idle, and it randomly selects to transmit a packet from one of its queues which turns out to be empty, while the other queue is non-empty. Accordingly, the slot would go idle and be wasted despite the system having packets awaiting transmission. Clearly, this results in a degradation in the system performance. Nevertheless, we can extend it to a more flexible work-conserving version of the proposed policy that exploits the resources efficiently without the risk of wasting slots. However, its delay analysis is notoriously complex since it involves deriving the moment generating function (MGF) of the joint lengths of the three queues in the system. Thus, we resort to the non work-conserving policy for its mathematical tractability. Consequently, we derive closed-form expressions for the expected packet delay, formulate and solve, analytically, optimization problems with the objective of minimizing delay at both users.

III. STABLE THROUGHPUT REGION

In this section, we characterize the stable throughput region of the system. Moreover, we distill valuable insights related to the effect of tuning a on the stability region of the system.

Theorem 1. The stable throughput region for the system in Fig. 1 under the proposed randomized strategy, for a certain value of a , is given by

$$\mathbf{R} = \left\{ (\lambda_p, \lambda_s) : \lambda_s < a f_{sd} \left[1 - \frac{\lambda_p}{\mu_p} \right], \right. \\ \left. \text{for } \lambda_p < \frac{f_{sd}(1-a)[f_{pd} + f_{ps}(1-f_{pd})]}{f_{sd}(1-a) + f_{ps}(1-f_{pd})} \right\} \quad (2)$$

Proof: We use Loynes' theorem [14] to establish the stability of each queue. The theorem states that if the arrival and service processes of a queue are stationary, then the queue is stable if the arrival rate is strictly less than the service rate.

- For Q_p stability, the following condition must be satisfied

$$\lambda_p < \mu_p \quad (3)$$

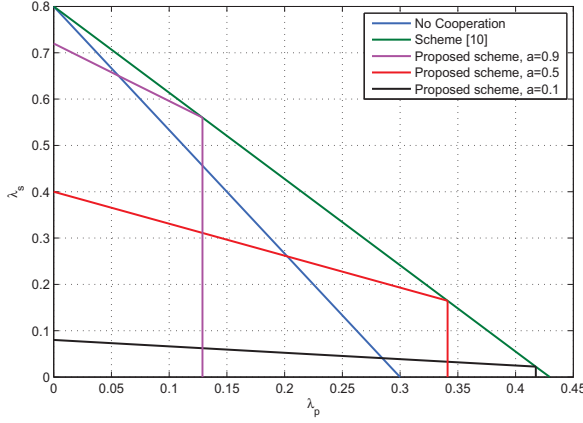


Fig. 2: Stable throughput region under different policies.

where μ_p denotes the service rate of Q_p . A packet departs Q_p if it is successfully decoded by at least one node, i.e., the destination or the SU. Thus, μ_p is given by

$$\mu_p = 1 - (1 - f_{pd})(1 - f_{ps}) = f_{pd} + f_{ps}(1 - f_{pd}) \quad (4)$$

- For Q_{sp} stability, the following condition must be satisfied

$$(1 - f_{pd})f_{ps} \frac{\lambda_p}{\mu_p} < \left[1 - \frac{\lambda_p}{\mu_p}\right] (1 - a)f_{sd} \quad (5)$$

A PU's packet arrives at Q_{sp} if Q_p is not empty, which has a probability of λ_p/μ_p , and an outage occurs in the link between the PU and destination, which happens w.p. $(1 - f_{pd})$, yet, no outage occurs in the link between the PU and SU, which happens w.p. f_{ps} . This explains the left hand side of (5) which is the rate of packet arrivals to the SU relay queue. The right hand side represents the service rate seen by the packets of Q_{sp} . A packet departs Q_{sp} if Q_p is empty, Q_{sp} is selected to transmit a packet, and there is no outage in the link between the SU and destination. Re-arranging the terms of (5) yields the following condition on λ_p

$$\lambda_p < \left[\frac{f_{sd}(1 - a)}{f_{sd}(1 - a) + f_{ps}(1 - f_{pd})} \right] \mu_p \quad (6)$$

Comparing (3) and (6), it becomes clear that (6) provides a tighter bound on λ_p due to the multiplication of μ_p by a term which is less than 1.

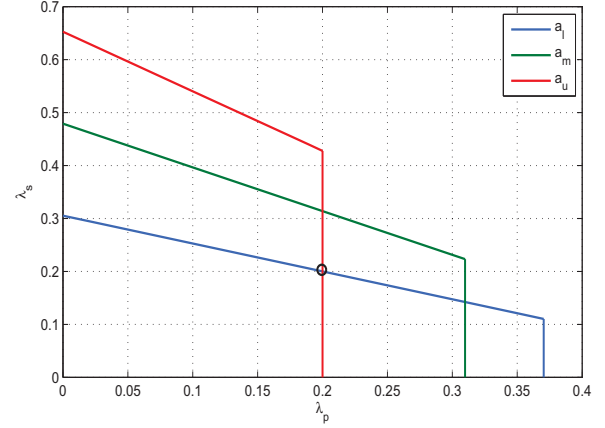
- For Q_s stability, the following condition must be satisfied

$$\lambda_s < a f_{sd} \left[1 - \frac{\lambda_p}{\mu_p} \right] \quad (7)$$

Using the same rationale, a packet departs Q_s if Q_p is empty, Q_s is selected to transmit a packet, and there is no outage in the link between the SU and destination. This explains the service rate seen by the packets of Q_s given in the right hand side of (7).

Conditions in (6) and (7) establish the result in (2). ■

Proposition 2. The maximum sustainable arrival rate at the PU, λ_p , decreases monotonically with a . Conversely, at a


 Fig. 3: Stable throughput region at different values of a for the system to be stable at $\lambda_p = \lambda_s = 0.2$.

fixed λ_p , the maximum sustainable arrival rate at the SU, λ_s , increases monotonically with a .

Proof: From the system stability conditions, the maximum sustainable λ_p for a given value of a is given by (6). Taking the derivative of (6) with respect to (w.r.t.) a yields

$$\frac{\partial \lambda_p}{\partial a} = \frac{-f_{sd}f_{ps}(1 - f_{pd})\mu_p}{[f_{sd}(1 - a) + f_{ps}(1 - f_{pd})]^2} \quad (8)$$

Since f_{sd} , f_{ps} , f_{pd} , and μ_p are all positive numbers less than 1, we conclude from (8) that $\frac{\partial \lambda_p}{\partial a}$ is negative definite. Thus, the maximum sustainable λ_p monotonically decreases with a .

On the other hand, at a fixed λ_p , the maximum sustainable λ_s for a given value of a is given by (7). Taking the derivative of (7) w.r.t. a yields

$$\frac{\partial \lambda_s}{\partial a} = f_{sd} \left[1 - \frac{\lambda_p}{\mu_p} \right] \quad (9)$$

The stability condition in (3) guarantees that $\frac{\lambda_p}{\mu_p}$ is less than 1. Thus, $\frac{\partial \lambda_s}{\partial a}$ is positive definite. Therefore, at a fixed λ_p , the maximum sustainable λ_s monotonically increases with a . ■

In Fig. 2, we plot the stable throughput region of the studied system under different multiple-access policies. Hereafter, the system parameters are chosen as follows: $f_{pd} = 0.3$, $f_{ps} = 0.4$, and $f_{sd} = 0.8$. From Fig. 2, we depict the effect of the probability a on the stability region of the proposed scheme. It can be realized that increasing a decreases the maximum sustainable λ_p . On the contrary, increasing a results in an increase in the maximum sustainable λ_s , for every feasible λ_p . This result is intuitive, since increasing a gives more chance for transmitting the SU own packets as opposed to the PU's relayed packets. This, in turn, reduces the degree of cooperation the PU experiences from the SU and, hence, the maximum sustainable λ_p decreases. On the other hand, since the SU own packets are more likely to be transmitted, the system can sustain higher values of λ_s .

Proposition 3. For any pair of arrival rates (λ_p, λ_s) in the stable throughput region of the system, there is a bounded range of values that a can take for the system to be stable at

the given (λ_p, λ_s) , which is given by

$$\frac{\lambda_s \mu_p}{f_{sd}(\mu_p - \lambda_p)} < a < 1 - \frac{f_{ps}(1 - f_{pd})\lambda_p}{f_{sd}(\mu_p - \lambda_p)} \quad (10)$$

Proof: This follows directly from the equations of the stability region. Using (6), it can be easily shown that

$$f_{sd}(1 - a)(\mu_p - \lambda_p) > f_{ps}(1 - f_{pd})\lambda_p \quad (11)$$

Since the system is assumed stable at (λ_p, λ_s) , then the term $(\mu_p - \lambda_p)$ is positive definite. Thus,

$$a < 1 - \frac{f_{ps}(1 - f_{pd})\lambda_p}{f_{sd}(\mu_p - \lambda_p)} \quad (12)$$

Using (7), it is straightforward to show that

$$a > \frac{\lambda_s \mu_p}{f_{sd}(\mu_p - \lambda_p)} \quad (13)$$

Thus, (12) and (13) establish the result in Proposition 3. ■

To illustrate Propositions 2 and 3, we compute the range of a values that makes the system stable at $\lambda_p = \lambda_s = 0.2$. Using (12) and (13), the upper and lower bounds on a , a_u and a_l , are computed, respectively. In Fig. 3, we plot the stability region at different values of a , specifically at a_u , a_l , and $a_m = \frac{a_u + a_l}{2}$. The curves corresponding to a_u and a_l intersect at the point of interest, $\lambda_p = \lambda_s = 0.2$. From Proposition 2, we know that increasing a reduces the maximum sustainable λ_p . This explains why at a_u the stability region has a maximum $\lambda_p = 0.2$. Exceeding the value of a_u results in excluding any point with $\lambda_p = 0.2$ from the stability region. Similarly, for the curve corresponding to a_l , it is clear that at $\lambda_p = 0.2$, the maximum sustainable λ_s is 0.2. At a given λ_p , decreasing a reduces the maximum sustainable λ_s , thus, a_l is the minimum possible value of a for which the system remains stable at $\lambda_p = \lambda_s = 0.2$. Moreover, the point of interest is strictly inside the stability region for any value of a in the open interval (a_l, a_u) . This is illustrated by the curve corresponding to a_m .

Theorem 4. The union of the stability regions given by (2) over all possible values of a is the same as that of any work-conserving cooperative scheme, e.g., the one derived in [10], and is given by

$$\lambda_s < f_{sd} - \left[\frac{f_{sd} + f_{ps}(1 - f_{pd})}{f_{pd} + f_{ps}(1 - f_{pd})} \right] \lambda_p \quad (14)$$

Proof: We take the union of (2) over all possible values of a , i.e., $a \in (0, 1)$. A method used to characterize this union is proposed in [11] in an analogous problem. It boils down to solving a constrained optimization problem to find the maximum feasible λ_s corresponding to each feasible λ_p .

Herein, we use the fact that increasing a increases the maximum sustainable λ_s for a given λ_p . Thus, at a fixed λ_p , the maximum over all feasible λ_s is achieved at the highest value of a that keeps the system stable at this given λ_p . Moreover, the point corresponding to the maximum over all possible λ_s and the given λ_p defines the boundary of the stability region, and hence intuitively, there is one and only one value of a that keeps the system stable at this point. Thus,

the upper and lower bounds on a that keep the system stable at this point coincide. Equating (12) and (13) yields

$$\lambda_s \mu_p = f_{sd} \mu_p - (f_{sd} + f_{ps}(1 - f_{pd}))\lambda_p \quad (15)$$

Dividing both sides by μ_p and using (4), the maximum over all feasible λ_s at a fixed λ_p is given by

$$\lambda_s = f_{sd} - \left[\frac{f_{sd} + f_{ps}(1 - f_{pd})}{f_{pd} + f_{ps}(1 - f_{pd})} \right] \lambda_p \quad (16)$$

which represents the boundary of the stability region of any work-conserving cooperative policy that is derived in [10] (Section III-A). Therefore, it has been established that, for any point (λ_p, λ_s) that belongs to the stable throughput region of any work-conserving policy, there exists at least one value of a that guarantees the system stability at this point under the non work-conserving proposed policy. Thus, the union of all stable throughput regions over all values of a of the proposed policy spans the stability region derived in [10]. ■

From Fig. 2, the union of all stable throughput regions of the proposed system (which matches that of [10]) strictly contains the stable throughput region achieved without cooperation.

IV. AVERAGE DELAY CHARACTERIZATION

In this section, we derive closed-form expressions for the average delay encountered by the packets of both users.

Theorem 5. The average delay encountered by the packets of the PU and SU, D_p and D_s , respectively, under the proposed randomized scheme, are given by

$$D_p = \frac{N_p + N_{sp}}{\lambda_p} \quad (17)$$

$$D_s = \frac{N_s}{\lambda_s} \quad (18)$$

where N_p and N_{sp} , the average lengths of Q_p and Q_{sp} , respectively, are given by

$$N_p = \frac{-\lambda_p^2 + \lambda_p}{f_{pd} + f_{ps}(1 - f_{pd}) - \lambda_p} \quad (19)$$

$$N_{sp} = \frac{m\lambda_p^2 + n\lambda_p}{\alpha\lambda_p^2 + \beta\lambda_p + \gamma} \quad (20)$$

where

$$\begin{aligned} m &= f_{ps}(1 - f_{pd}) \left[\frac{(1 - a)f_{sd} - f_{pd}}{f_{pd} + f_{ps}(1 - f_{pd})} \right. \\ &\quad \left. - (1 - a)f_{sd} - f_{ps}(1 - f_{pd}) \right] \\ n &= f_{ps}(1 - f_{pd}) [f_{pd} + f_{ps}(1 - f_{pd})] \\ \alpha &= (1 - a)f_{sd} + f_{ps}(1 - f_{pd}) \\ \beta &= [f_{pd} + f_{ps}(1 - f_{pd})] [-2(1 - a)f_{sd} - f_{ps}(1 - f_{pd})] \\ \gamma &= (1 - a)f_{sd} [f_{pd} + f_{ps}(1 - f_{pd})]^2 \end{aligned} \quad (21)$$

and N_s , the average length of Q_s , is given by

$$N_s = \frac{\lambda_p \lambda_s A + (\lambda_s^2 - \lambda_s)B(B + \lambda_p)}{BC} \quad (22)$$

where

$$\begin{aligned} A &= af_{sd}[f_{pd} + f_{ps}(1 - f_{pd}) - 1] \\ B &= f_{pd} + f_{ps}(1 - f_{pd}) - \lambda_p \\ C &= (\lambda_s - af_{sd})[f_{pd} + f_{ps}(1 - f_{pd})] + af_{sd}\lambda_p \end{aligned} \quad (23)$$

Proof: We start by computing the average delay of the packets of the SU followed by that of the PU.

Applying Little's law on Q_s renders D_s exactly as given by (18). Thus, it remains to calculate N_s . The dependence of the service processes at Q_s and Q_{sp} on the state of Q_p is inherent from the concept of cognitive radios. It is worth noting that the non work-conserving behavior of the proposed policy makes the delay analysis mathematically tractable, since Q_s and Q_{sp} become independent. To analyze the average delays at different queues, we resort to the MGF approach [15]. The MGF of the joint queue lengths Q_p and Q_s is defined as

$$G(x, y) = \lim_{t \rightarrow \infty} \mathbf{E} \left[x^{Q_p^t} y^{Q_s^t} \right] \quad (24)$$

where \mathbf{E} and \mathbf{P} denote the statistical expectation and the probability operators, respectively. Expanding (24), taking its derivative w.r.t. y and substituting by $x = y = 1$ yields

$$G_y(1, 1) = \lim_{t \rightarrow \infty} \sum_{j=0}^{\infty} j \mathbf{P} [Q_s^t = j] = N_s \quad (25)$$

Using the queue evolution form provided by (1), we write

$$\begin{aligned} \mathbf{E} \left[x^{Q_p^{t+1}} y^{Q_s^{t+1}} \right] &= \mathbf{E} \left[x^{(Q_p^t - Y_p^t + X_p^t)} y^{(Q_s^t - Y_s^t + X_s^t)} \right] \\ &= (\lambda_p x + 1 - \lambda_p)(\lambda_s y + 1 - \lambda_s) \mathbf{E} \left[x^{(Q_p^t - Y_p^t)} y^{(Q_s^t - Y_s^t)} \right] \end{aligned} \quad (26)$$

This follows from the independent arrivals at Q_p and Q_s , that yield independent Bernoulli distributed random variables, X_p^t and X_s^t , which produce MGFs of $(\lambda_p x + 1 - \lambda_p)$ and $(\lambda_s y + 1 - \lambda_s)$, respectively. Expanding the above equation, we have

$$\begin{aligned} \mathbf{E} \left[x^{Q_p^{t+1}} y^{Q_s^{t+1}} \right] &= \\ &(\lambda_p x + 1 - \lambda_p)(\lambda_s y + 1 - \lambda_s) \left\{ \mathbf{E}[\mathbf{1}[Q_p^t = 0, Q_s^t = 0]] \right. \\ &+ \left[\frac{f_{pd} + f_{ps}(1 - f_{pd})}{x} + (1 - f_{ps})(1 - f_{pd}) \right] \\ &\times \mathbf{E}[x^{Q_p^t} \mathbf{1}[Q_p^t > 0, Q_s^t = 0]] \\ &+ \left[\frac{af_{sd}}{y} + 1 - af_{sd} \right] \mathbf{E}[y^{Q_s^t} \mathbf{1}[Q_p^t = 0, Q_s^t > 0]] \\ &+ \left[\frac{f_{pd} + f_{ps}(1 - f_{pd})}{x} + (1 - f_{ps})(1 - f_{pd}) \right] \\ &\times \left. \mathbf{E}[x^{Q_p^t} y^{Q_s^t} \mathbf{1}[Q_p^t > 0, Q_s^t > 0]] \right\} \end{aligned} \quad (27)$$

To explain the terms inside the braces of (27), we analyze the 4 possible combinations of the queue states, Q_p^t and Q_s^t

- $Q_p^t = 0, Q_s^t = 0$

Since both queues are empty, $Y_p^t = Y_s^t = 0$. This explains the first term in the braces in (27).

- $Q_p^t > 0, Q_s^t = 0$

Clearly, no departures occur at Q_s since it is empty, i.e., $Y_s^t = 0$. At the PU side, it transmits a packet whenever it has a non-empty queue. Thus, Y_p^t is given by

$$Y_p^t = \begin{cases} 1, & \text{w.p. } f_{pd} + f_{ps}(1 - f_{pd}) \\ 0, & \text{w.p. } (1 - f_{ps})(1 - f_{pd}) \end{cases} \quad (28)$$

This states that a departure occurs at Q_p if the packet is decoded by at least one node, either the destination or the SU. Otherwise, no departures occur and the packet remains at Q_p for retransmission in the next slot. This gives the second term in the braces in (27).

- $Q_p^t = 0, Q_s^t > 0$

The PU is idle, thus, $Y_p^t = 0$. Then, the SU gains access to the system and transmits a packet. It randomly selects the source of this packet to be either Q_s or Q_{sp} . Therefore, Y_s^t is given by

$$Y_s^t = \begin{cases} 1, & \text{w.p. } af_{sd} \\ 0, & \text{w.p. } 1 - af_{sd} \end{cases} \quad (29)$$

This states that a departure occurs at Q_s if it is selected to transmit, which happens w.p. a , and the transmitted packet is successfully decoded by the destination, which happens w.p. f_{sd} . Otherwise, no departures occur. This results in the third term in the braces in (27).

- $Q_p^t > 0, Q_s^t > 0$

Since the PU has the priority to transmit, the SU is silent and $Y_s^t = 0$. The PU transmits a packet and Q_p evolves exactly following the case of $Q_p^t > 0, Q_s^t = 0$ yielding the last term in the braces in (27).

Taking the limit when $t \rightarrow \infty$ at both sides of (27), we get

$$\begin{aligned} G(x, y) &= (\lambda_p x + 1 - \lambda_p)(\lambda_s y + 1 - \lambda_s) \\ &\times \frac{b(x, y)G(0, 0) + c(x, y)G(0, y)}{yd(x, y)} \end{aligned} \quad (30)$$

where

$$\begin{aligned} b(x, y) &= xyaf_{sd} - xaf_{sd} \\ c(x, y) &= xaf_{sd} - y[f_{pd} + f_{ps}(1 - f_{pd})] + xy[f_{sd} + f_{ps}(1 - f_{pd}) - af_{sd}] \\ d(x, y) &= x - (\lambda_p x + 1 - \lambda_p)(\lambda_s y + 1 - \lambda_s) \times \\ &\quad [f_{pd} + f_{ps}(1 - f_{pd}) + x(1 - f_{ps})(1 - f_{pd})] \end{aligned} \quad (31)$$

From the definition of $G(x, y)$, note that

$$\begin{aligned} G(0, 0) &= \lim_{t \rightarrow \infty} \mathbf{E}[\mathbf{1}[Q_p^t = 0, Q_s^t = 0]] \\ G(x, 0) &= G(0, 0) + \lim_{t \rightarrow \infty} \mathbf{E}[x^{Q_p^t} \mathbf{1}[Q_p^t > 0, Q_s^t = 0]] \\ G(0, y) &= G(0, 0) + \lim_{t \rightarrow \infty} \mathbf{E}[y^{Q_s^t} \mathbf{1}[Q_p^t = 0, Q_s^t > 0]] \\ G(x, y) &= G(x, 0) + G(0, y) - G(0, 0) \\ &\quad + \lim_{t \rightarrow \infty} \mathbf{E}[x^{Q_p^t} y^{Q_s^t} \mathbf{1}[Q_p^t > 0, Q_s^t > 0]] \end{aligned} \quad (32)$$

Along the lines of [15], $G(0, 0)$ is evaluated using the normalization condition, $G(1, 1) = 1$, by taking the limit of (30)

when $(x, y) \rightarrow (1, 1)$, which yields

$$G(0, 0) = \frac{af_{sd}[f_{pd} + f_{ps}(1 - f_{pd}) - \lambda_p] - \lambda_s[f_{pd} + f_{ps}(1 - f_{pd})]}{af_{sd}[f_{pd} + f_{ps}(1 - f_{pd})]} \quad (33)$$

In the derivation of (33), we use the fact that

$$G(0, 1) = \lim_{t \rightarrow \infty} \mathbf{P}[Q_p^t = 0] = 1 - \frac{\lambda_p}{f_{pd} + f_{ps}(1 - f_{pd})} \quad (34)$$

To find N_s , we solve for $G_y(1, 1)$. We evaluate the derivative of (30) w.r.t. y , then take the limit of the result when $(x, y) \rightarrow (1, 1)$. Applying L'Hopital's rule twice, we obtain an equation relating $G_y(1, 1)$ to $G_y(0, 1)$ as

$$G_y(1, 1) = \lambda_s - 1 + \frac{af_{sd}}{\lambda_s} G_y(0, 1) \quad (35)$$

Next, we compute $\left. \frac{\partial G(y, y)}{\partial y} \right|_{y=1}$. We make use of the fact that $\left. \frac{\partial G(y, y)}{\partial y} \right|_{y=1} = N_p + N_s$, and $G_y(1, 1) = N_s$. After some algebraic manipulation, we get

$$G_y(1, 1) = \frac{-(\lambda_p + \lambda_s)^2 + \lambda_p \lambda_s + \lambda_p + \lambda_s}{f_{pd} + f_{ps}(1 - f_{pd}) - \lambda_p - \lambda_s} - N_p + \left[\frac{f_{pd} + f_{ps}(1 - f_{pd}) - af_{sd}}{f_{pd} + f_{ps}(1 - f_{pd}) - \lambda_p - \lambda_s} \right] G_y(0, 1) \quad (36)$$

We can easily calculate N_p by observing that Q_p is a discrete-time $M/M/1$ queue with arrival rate λ_p and service rate μ_p . Thus, applying the Pollaczek-Khinchine formula [16], N_p is directly given by (19). Solving (35) and (36) together using the result obtained by (19), the term $G_y(0, 1)$ is eliminated and N_s is exactly given by (22) in Theorem 5.

Next, we calculate the average delay of the PU's packets. A PU's packet, if directly delivered to the destination, experiences the queueing delay at Q_p only. This happens w.p. $1 - \epsilon = \frac{f_{pd}}{1 - (1 - f_{ps})(1 - f_{pd})}$, which is the probability that the packet is successfully decoded by the destination given that it is dropped from Q_p . Otherwise, if the transmission through the direct link between the PU and destination fails, the packet is relayed through Q_{sp} and, hence, experiences the total queueing delay at both Q_p and Q_{sp} . This happens w.p. ϵ . Therefore, the average delay that a PU's packet experiences is given by

$$D_p = (1 - \epsilon)\tau_p + \epsilon(\tau_p + \tau_{sp}) = \tau_p + \epsilon\tau_{sp} \quad (37)$$

where τ_p and τ_{sp} denote the average queueing delays at Q_p and Q_{sp} , respectively. Since the arrival rates at Q_p and Q_{sp} are given by λ_p and $\epsilon\lambda_p$, respectively. Then, applying Little's law yields

$$\tau_p = N_p / \lambda_p, \quad \tau_{sp} = N_{sp} / \epsilon\lambda_p \quad (38)$$

Substituting (38) in (37) renders D_p exactly matching (17). Provided that N_p is shown to be given by (19), the calculation of D_p boils down to evaluating N_{sp} . As indicated earlier, the service process at Q_{sp} depends on the state of Q_p , so we again employ the MGF approach to compute N_{sp} . Let $H(x, y) = \lim_{t \rightarrow \infty} \mathbf{E}[x^{Q_p^t} y^{Q_{sp}^t}]$ be defined as the MGF of the joint queue lengths of Q_p and Q_{sp} . Using an analogous

derivation employed to evaluate $G(x, y)$, we write $H(x, y)$ as

$$H(x, y) = (\lambda_p x + 1 - \lambda_p) \frac{b'(x, y)G(0, 0) + c'(x, y)G(0, y)}{y d'(x, y)} \quad (39)$$

where

$$\begin{aligned} b'(x, y) &= x(1 - a)f_{sd}(y - 1) \\ c'(x, y) &= x(1 - a)f_{sd} - yf_{pd} - y^2f_{ps}(1 - f_{pd}) \\ &\quad + xy[f_{pd} + f_{ps}(1 - f_{pd}) - (1 - a)f_{sd}] \\ d'(x, y) &= x - (\lambda_p x + 1 - \lambda_p)[f_{pd} + yf_{ps}(1 - f_{pd}) + x(1 - f_{ps})(1 - f_{pd})] \end{aligned} \quad (40)$$

Following the same footsteps of the approach employed to evaluate N_s , N_{sp} is shown to be given by (20). ■

In order to be able to solve the optimization problem formulated later, we check how D_p and D_s behave in response to variations in a .

Proposition 6. Under the proposed randomized cooperative policy, if the system is stable, the average delay experienced by the packets of the PU, D_p , is a monotonically increasing function in a , while the average delay encountered by the packets of the SU, D_s , decreases monotonically with a .

Proof: The closed-form expressions of D_p and D_s as functions of the parameter a are given in Theorem 5 by (17) and (18), respectively. Thus, to check how D_p and D_s behave in response to changes in a , we compute their derivatives w.r.t. a . First, we take the derivative of (17) w.r.t. a which yields

$$\frac{\partial D_p}{\partial a} = \Omega \left\{ (\mu_p^3 - \lambda_p^3) - \lambda_p \omega (\mu_p - \lambda_p) \right\} \quad (41)$$

where

$$\begin{aligned} \Omega &= \frac{f_{sd}f_{ps}(1 - f_{pd})}{(\alpha\lambda_p^2 + \beta\lambda_p + \gamma)^2} \\ \omega &= f_{pd} + [f_{ps}(1 - f_{pd}) + 2]\mu_p \end{aligned} \quad (42)$$

After some manipulations, we obtain

$$\frac{\partial D_p}{\partial a} = \Omega(\mu_p - \lambda_p) [(\mu_p - \lambda_p)^2 + \lambda_p \mu_p - \lambda_p [f_{pd} + \mu_p f_{ps}(1 - f_{pd})]] \quad (43)$$

Since $\mu_p < 1$, we note that $\lambda_p [f_{pd} + \mu_p f_{ps}(1 - f_{pd})] < \lambda_p [f_{pd} + f_{ps}(1 - f_{pd})] = \lambda_p \mu_p$. Using this fact along with (43), it can be seen that

$$\frac{\partial D_p}{\partial a} > \Omega(\mu_p - \lambda_p)^3 \quad (44)$$

Since it has been established in the proof of Theorem 1 by (3) that $(\mu_p - \lambda_p)$ is positive definite as long as Q_p is stable. Furthermore, it can be noticed that Ω is positive by definition. Therefore, we conclude from (44) that the derivative of D_p w.r.t. a is positive, irrespective of the choice of a . This proves that D_p is a monotonically increasing function in a .

Next, we proceed with taking the derivative of (18) w.r.t. a ,

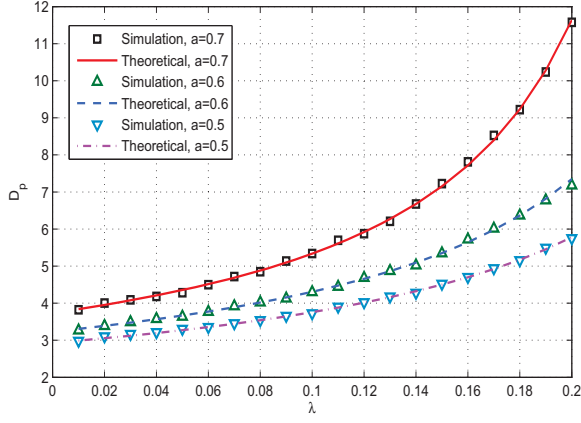


Fig. 4: Average delay of the PU's packets for different values of a .

which yields

$$\frac{\partial D_s}{\partial a} = f_{sd} \mu_p (\mu_p - \lambda_p) \left\{ \lambda_s (\lambda_p^2 - \lambda_p) + (\mu_p - \lambda_p) [\lambda_s \mu_p - (\mu_p - \lambda_p)] \right\} \quad (45)$$

Notice that $f_{sd} \mu_p (\mu_p - \lambda_p)$ is positive definite. Thus, we focus our attention on analyzing the term in the braces. The term $(\lambda_p^2 - \lambda_p)$ is obviously negative definite since $\lambda_p^2 < \lambda_p$. This follows from the fact that λ_p is a positive number which is less than 1. Finally, for checking the sign of $[\lambda_s \mu_p - (\mu_p - \lambda_p)]$, we resort to the stability condition given by (7) which states that $\lambda_s \mu_p < a f_{sd} (\mu_p - \lambda_p)$. Since each of a and f_{sd} is a positive number less than 1, then $a f_{sd}$ is also a positive number less than 1. Thus, it is guaranteed that $\lambda_s \mu_p < (\mu_p - \lambda_p)$. This establishes that $[\lambda_s \mu_p - (\mu_p - \lambda_p)]$ is negative definite. Therefore, it has been proven that $\frac{\partial D_s}{\partial a}$ is negative, irrespective of the choice of a . This, in turn, proves that D_s is a monotonically decreasing function in a . ■

V. NUMERICAL RESULTS

In this section, we investigate the performance of the system under the proposed randomized policy. Extensive simulations are conducted to validate the closed-form expressions obtained for the average delay experienced by the packets of the PU and SU. Furthermore, we characterize and analyze a fundamental tradeoff between the average delay and the throughput at both the PU and SU. Moreover, performance comparisons to the no-cooperation scenario as well as [10] are done to show the merits of the proposed randomized scheme.

In Figs. 4 and 5, we plot the average delay experienced by the packets of the PU and the SU, respectively, versus λ , where we choose $\lambda_p = \lambda_s = \lambda$ for ease of exposition. It can be viewed that the results obtained through simulations exactly match the results of the closed-form expressions derived in Theorem 5. This validates the soundness of the mathematical model and the MGF approach. Moreover, at a given λ , when a increases, D_p is shown to increase, while D_s decreases. This matches the result stated by Proposition 6.

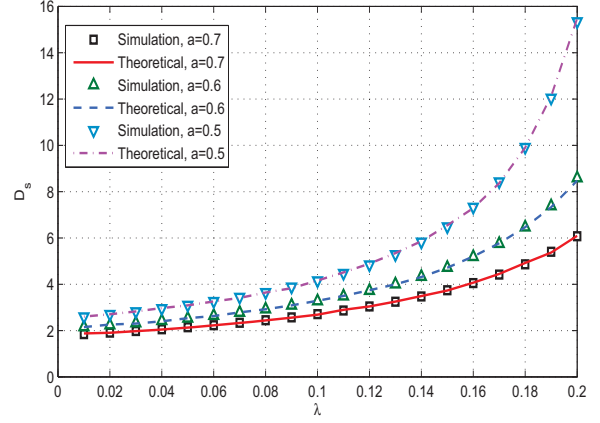


Fig. 5: Average delay of the SU's packets for different values of a .

Next, we characterize a fundamental tradeoff that arises between the average delay and the throughput at both the PU and SU. Intuitively, when a node needs to maintain a higher throughput, it loses in terms of the average delay encountered by its packets. Given that the system is stable, the node's throughput equals its packet arrival rate. Thus, increased throughput means injecting more packets into the system which yields a higher delay.

In Fig. 6, we illustrate the throughput-delay tradeoff at the PU. Note that, given the stability of the system, the throughput of the PU equals λ_p . We fix the value of λ_s at 0.2. Then, at every point λ_p , we formulate and solve the following optimization problem

$$\begin{aligned} & \underset{a}{\text{minimize}} && D_p \\ & \text{subject to} && \lambda_p < \left[\frac{f_{sd}(1-a)}{f_{sd}(1-a) + f_{ps}(1-f_{pd})} \right] \mu_p, \\ & && \lambda_s < a f_{sd} \left[1 - \frac{\lambda_p}{\mu_p} \right], \\ & && 0 < a < 1. \end{aligned} \quad (46)$$

Thus, at every point (λ_p, λ_s) , we solve for the optimal value of a that minimizes D_p , while simultaneously keeping the system stable at this point. The solution of the problem is easily done using the results obtained in Propositions 3 and 6. First, we compute the feasible set, which is the set of a values that guarantee the system stability at (λ_p, λ_s) . This set is given by the interval between the lower and upper bounds of a defined by (13) and (12), respectively. To minimize D_p , we choose the minimum feasible value of a , i.e., the lower bound given by (13), since it has been established in Proposition 6 that D_p is a monotonically increasing function of a . Finally, we calculate D_p using the closed-form expression obtained in Theorem 5 at this optimal value of a . Thus, in Fig. 6, we show the best achievable performance of the system under the proposed randomized policy in terms of the average delay at the PU side. The tradeoff is now obvious, as the delay is shown to increase when the throughput increases. We also plot the

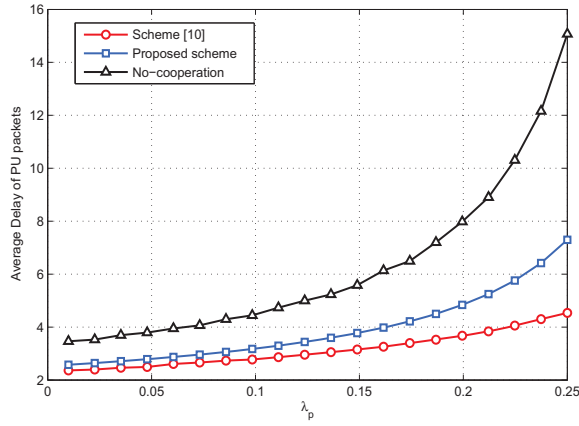


Fig. 6: Throughput-delay tradeoff at the PU.

throughput-delay curves corresponding to the no-cooperation scenario and the policy proposed in [10]. The proposed policy is shown to outperform the no-cooperation scenario, however, being outperformed by [10]. This is expected since in [10], the priority is always given to the packets of the PU, i.e., Q_p has the highest priority to transmit followed by Q_{sp} , while Q_s never transmits except when both Q_p and Q_{sp} are empty.

We follow the same steps at the SU side. Fixing $\lambda_p = 0.2$, we vary the throughput of the SU, λ_s , to investigate its effect on D_s . For every point (λ_p, λ_s) , we minimize D_s subject to keeping the system stable at this point. The resulting throughput-delay curves for the proposed policy as well as for [10] are shown in Fig. 7. We avoided plotting the no-cooperation baseline case to have a clear view for the comparison between the plotted policies, since the no-cooperation performance is way worse than both. It can be viewed that at the SU, the best achievable performance of the system under the proposed randomized policy in terms of the average delay at the SU side, is superior to the performance of the system under the policy proposed in [10].

VI. CONCLUSION

We propose a randomized cooperative policy, whereby the SU serves either the queue of its own data or the relay queue w.p. a and $(1 - a)$, respectively. Results unveil performance gains of cooperation as opposed to the no-cooperation case. Cooperation is shown to expand the stable throughput region of the system as well as reducing the delay encountered by the packets of both users. Moreover, the proposed randomized policy opens room for trading the PU delay for enhanced SU delay. Furthermore, the system can be adjusted to satisfy various objectives under a set of constraints. This degree of freedom is inherited from the tunability of a .

REFERENCES

- [1] P. J. Kolodzy, "Spectrum policy task force," *Federal Commun. Comm., Washington, DC, Rep. ET Docket*, no. 02-135, 2002.
- [2] J. Mitola *et al.*, "Cognitive radio: An integrated agent architecture for software defined radio," in *PhD thesis, Royal Institute of Technology (KTH)*, pp. 271–350, 2000.

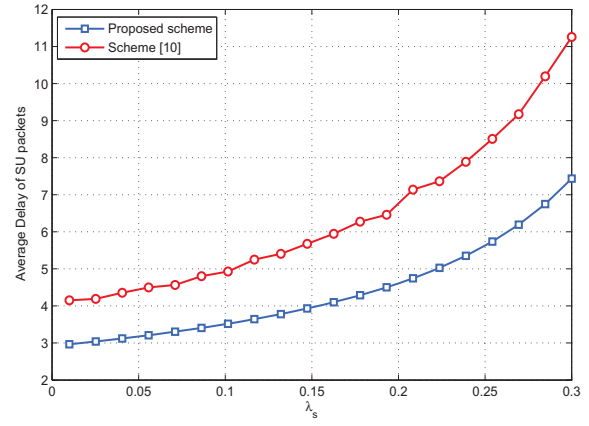


Fig. 7: Throughput-delay tradeoff at the SU.

- [3] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.
- [4] J. N. Laneman, D. N. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.
- [5] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *Information Theory, IEEE Transactions on*, vol. 51, no. 9, pp. 3037–3063, 2005.
- [6] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.
- [7] E. Telatar, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [8] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4152–4172, 2005.
- [9] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 203–213, 2008.
- [10] B. Rong and A. Ephremides, "Cooperative access in wireless networks: stable throughput and delay," *IEEE Transactions on Information Theory*, vol. 58, no. 9, pp. 5890–5907, 2012.
- [11] A. K. Sadek, K. R. Liu, and A. Ephremides, "Cognitive multiple access via cooperation: protocol design and performance analysis," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3677–3696, 2007.
- [12] I. Krikidis, J. N. Laneman, J. S. Thompson, and S. McLaughlin, "Protocol design and throughput analysis for multi-user cognitive cooperative systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4740–4751, 2009.
- [13] R. W. Wolff, *Stochastic modeling and the theory of queues*. Prentice Hall, 1989.
- [14] R. Loynes, "The stability of a queue with non-independent inter-arrival and service times," in *Proc. Cambridge Philos. Soc.*, vol. 58, no. 3. Cambridge Univ Press, 1962, pp. 497–520.
- [15] M. Sidi and A. Segall, "Two interfering queues in packet-radio networks," *IEEE Transactions on Communications*, vol. 31, no. 1, pp. 123–129, 1983.
- [16] L. Kleinrock, *Queueing systems, vol. 1*. New York: Wiley, 1975.