Opportunistic Cooperation in Cognitive Femtocell Networks

Rahul Urgaonkar, *Member, IEEE*, and Michael J. Neely, *Senior Member, IEEE*

***Abstract*—We investigate opportunistic cooperation between secondary (femtocell) users and primary (macrocell) users in cognitive femtocell networks. We consider two models for such cooperation. In the first model, called the *Cooperative Relay Model*, a secondary user cannot transmit its own data concurrently with a primary user. However, it can employ cooperative relaying of primary user data in order to improve the latter's effective transmission rate. In the second model, called the *Interference Model*, a secondary user is allowed to transmit its data concurrently with a primary user. However, the secondary user can "cooperate" by deferring its transmissions when the primary user is busy. In both models, the secondary users must make intelligent cooperation decisions as they seek to maximize their own throughput subject to average power constraints. The decision options are different during idle and busy periods of the primary user, and the decisions in turn influence the durations of these periods according to a controllable infinite state Markov chain. Such problems can be formulated as constrained Markov decision problems, and conventional solution techniques require either extensive knowledge of the system dynamics or learning based approaches that suffer from large convergence times. However, using a generalized Lyapunov optimization technique, we design a novel *greedy* and *online* control algorithm that overcomes these challenges. Remarkably, this algorithm does not require any knowledge of the network arrival rates and is provably optimal.**

***Index Terms*—Resource Allocation, Opportunistic Cooperation, Optimal Control, Cognitive Femtocell**

I. Introduction

We consider a cognitive radio network with one primary user and multiple secondary users. Packets randomly arrive at the primary user and are queued for transmission. The primary user transmits on every slot that it has packets. The success probability is determined by the cooperation decisions made by the secondary users. We consider two models for cooperation: (i) *Cooperative Relay Model*, and (ii) *Interference Model*. In the first model, a secondary user cannot transmit its own data when the primary user is busy. However, it can employ cooperative relaying of primary user data in order to improve the latter's effective success probability. In the second model, a secondary user is allowed to transmit its own data in any slot. However, the secondary user can "cooperate" by deferring its transmissions when the primary user is busy, thereby reducing interference and increasing the primary user's success probability. In both models, the incentive for cooperation is that it reduces the busy periods of the primary user, and allows more idle slots under which the secondary users can transmit without interference. However, such cooperation decisions must be made by the secondary users in an intelligent fashion in order to maximize their own throughput subject to average power constraints.

The decision options and success probabilities are different during idle and busy periods of the primary user. The size of these periods is in turn affected by the cooperation decisions that are made. This creates a non-trivial constrained Markov decision problem (MDP) with infinite state space [1], [2], where the state is the number of packets in the primary user queue. Conventional solution techniques for constrained MDPs have several drawbacks. For example, when the state transition probabilites are known, dynamic programming [4] can be used. However, it is known to suffer from the "curse of dimensionality". In the absense of such knowledge of system dynamics, learning based schemes such as Q-learning may be used [3]. For example, Q-learning based approaches are used in [5], [6] for problems of delay-constrained wireless transmission scheduling and [7] studies distributed Q-learning for interference control in multiuser cognitive femtocell networks. Q-learning algorithms are general solutions for MDPs that involve learning over time. However, they can suffer from long convergence times [1]–[3]. In this work, we use a novel alternative approach that overcomes these limitations. We first transform the problem into a sequence of online *unconstrained* stochastic shortest path problems, using a ratio rule for Lyapunov optimization. This ratio rule is similar to those used in a different context in [8]–[10] for restless bandit and renewal theory problems. Remarkably, we show that for our cognitive radio scenario, the transformed stochastic shortest path problems can be solved *exactly*, and can be implemented without knowledge of the arrival rates of the users. Further, the resulting algorithm does not require any explicit learning phase. This approach is powerful and can likely be applied to other constrained Markov decisions problems with similar structure.

Much prior work on resource allocation in cognitive radio networks has focused on the *dynamic spectrum access* model [11], [12] in which the secondary users seek transmission opportunities for their packets on vacant primary channels in frequency, time, or space. Under this model, the primary users are assumed to be oblivious of the presence of the secondary users and transmit whenever they have data to send. Secondly, a collision model is assumed for the physical layer

Rahul Urgaonkar is with the Network Research Department, Raytheon BBN Technologies, Cambridge, MA 02138. Michael J. Neely is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089. This work was performed when Rahul Urgaonkar was a PhD student at the University of Southern California. (Email: rahul@bbn.com, mjneely@usc.edu, Web: http://www.ir.bbn.com/∼rahul).

This material is supported in part by one or more of the following: the DARPA IT-MANET program grant W911NF-07-0028, the NSF Career grant CCF-0747525, NSF grant 0964479, the Network Science Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory W911NF-09-2-0053.

in which if a secondary user transmits on a busy primary channel, then there is a collision and both packets are lost. We considered a similar model in our prior work [13] where the objective is to design an opportunistic scheduling policy for the secondary users that maximizes their throughput utility while providing tight reliability guarantees on the maximum number of collisions suffered by a primary user over *any* given time interval. We note that this formulation does not consider the possibility of any cooperation between the primary and secondary users. Further, it assumes that the secondary user activity does not affect the primary user channel occupancy process. This allowed [13] to use a greedy "drift-plus-penalty" technique of Lyapunov optimization theory. Our current problem has a more complex Markov decision structure, and the same Lyapunov optimization techniques cannot be used.

There is a growing body of work that investigates alternate models for the interaction between the primary and secondary users in a cognitive radio network. In particular, the idea of cooperation at the physical layer has been considered from an information-theoretic perspective in many works (see the survey paper [14] and the references therein). These are motivated by the work on the classical interference and relay channels [15]. The main idea in these works is that the resources of the secondary user can be utilized to improve the performance of the primary transmissions. In return, the secondary user can obtain more transmission opportunities for its own data when the primary channel is idle. These works mainly treat the problem from a physical layer/information-theoretic perspective and do not consider upper layer issues such as queueing, higher priority for primary user, etc.

Recent work that addresses some of these issues includes [16]–[20]. Specifically, [16] considers the scenario where the secondary user acts as a relay for those packets of the primary user that it receives successfully but which are not received by the primary destination. It derives the stable throughput of the secondary user under this model. [17], [18] use a Stackelberg game framework to study spectrum leasing strategies in *cooperative cognitive radio networks* where the primary users lease a portion of their licensed spectrum to secondary users in return for cooperative relaying. [19], [20] study and compare different physical layer strategies for relaying in such cognitive cooperative systems. An important consequence of this interaction between the primary and secondary users is that the secondary user activity can now potentially influence the primary user channel occupancy process. However, there has been little work in studying this scenario. Exceptions include the work in [21] that considers a two-user setting where collisions caused by the opportunistic transmissions of the secondary user result in retransmissions by the primary user. This works uses a conventional linear programming approach to Markov decision problems that requires finite state space and complete knowledge of the system probabilities. Our current paper develops a new dynamic approach that does not require knowledge of the primary user arrival rate.

The rest of the paper is organized as follows. In Section II, we introduce the problem for the case of a single primary user and a single secondary user. This is extended to multiple secondary users in Section VI. We describe the two

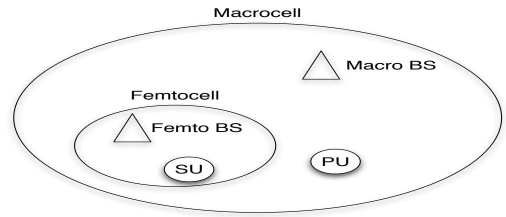


Fig. 1. Example femtocell network with primary and secondary users.

cooperation models in Section II-B and formulate the problem of maximizing the secondary user throughput subject to time average power constraints under these models in Section II-D. In Sections III and IV, we present a solution to this problem using a novel approach based on a generalized Lyapunov optimization technique. Finally, we present simulation results in Section VII, where we also illustrate that our algorithm is adaptive to changes in the arrival rates.

II. Basic Model

We consider a network with one primary user, one secondary user and their respective receivers (this is extended to treat multiple secondary users in Section VI). The primary user is the licensed owner of the channel and transmits to its receiver whenever it has data to send. The secondary user does not have any licensed spectrum and seeks transmission opportunities on the primary channel. This model can capture a femtocell scenario [22], [23] where the primary user is a legacy mobile user that communicates with the macro base station over licensed spectrum while the secondary user is the femtocell user that does not have any licensed spectrum of its own (Fig. 1). Within this setting, we consider two models for secondary user transmissions: (i) *Cooperative Relay Model*, and (ii) *Interference Model*. In both models, the secondary user can use cooperation to effectively increase the primary user transmission rate. This can then create more opportunities for the secondary user to transmit its own data when the primary user is idle. These models and their cooperation mechanisms are discussed in detail in Section II-B.

A. Timeslot Structure

We consider a time-slotted model. We assume that the system operates over a frame-based structure. Specifically, the timeline can be divided into successive non-overlapping frames of duration $T[k]$ slots where $k \in \{1, 2, 3, \ldots\}$ represents the frame number (see Fig. 2). The start time of frame k is denoted by t_k with $t_1 = 0$. The length of frame k is given by $T[k] \triangleq t_{k+1} - t_k$. For each k, the frame length $T[k]$ is a random function of the control decisions taken during that frame. Each frame can be further divided into two periods: PU Idle and PU Busy. The "PU Idle" period corresponds to the slots when the primary user does not have any packet to send to its receiver and is idle. The "PU Busy" period corresponds to the slots when the primary user is transmitting its packets to its receiver over the licensed spectrum. As shown in Fig. 2, every frame starts with the "PU Idle" period which is followed by the "PU Busy" period and ends when the primary user becomes idle again.

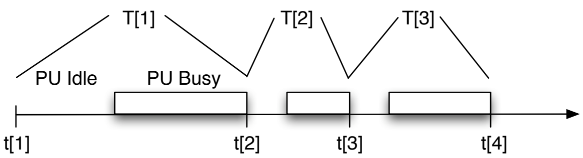


Fig. 2. Frame-based structure of the problem under consideration. Each frame consists of two periods: PU Idle and PU Busy.

We assume that the primary user receives new packets every slot according to an i.i.d. Bernoulli arrival process $A_{pu}(t)$ with rate λ_{pu} packets/slot. This means that the length of the "PU Idle" period of any frame is a geometric random variable with parameter λ_{pu}. However, the length of the "PU Busy" period depends on the secondary user control decisions as discussed in the next section. In any slot t, if the primary user has a non-zero queue backlog, it transmits one packet to its base station. We assume that the transmission of each packet takes one slot. If the transmission is successful, the packet is removed from the primary user queue. However, if the transmission fails, the packet is retained in the queue for future retransmissions.

B. Cooperation Mechanisms and Incentives

In the *Cooperative Relay Model*, the secondary user cannot transmit its packets when the channel is being used by the primary user. It can transmit its packets only during the "PU Idle" period of the frame and must stop its transmissions whenever the primary user becomes active again. This could be because the interference generated at the primary receiver by secondary transmissions is unacceptable. However, in the "PU Busy" period , the secondary user can employ cooperative relaying to improve the success probability of the primary transmissions. This has the effect of *decreasing* the expected length of the "PU Busy" period.

In order to cooperate, the secondary user must allocate its power resources to help relay the primary user packet. This cooperation can take place in several ways depending on the cooperative protocol being used (see [19] for some examples). For example, suppose the Amplify-and-Forward protocol is used for cooperative relaying. Then, the slot is divided into two parts. In the first part, the primary user transmits its packet to both the secondary user and the primary receiver. In the second part, the secondary user transmits an amplified version of the signal that it received in the first part. Finally, the primary receiver uses both the signals jointly to decode the packet. A Decode-and-Forward approach would work similarly. In our model, these details are captured by the resulting probability of successful transmission.

In the *Interference Model*, the secondary user can transmit its packets concurrently with the primary user. However, the resulting interference reduces the success probability of the primary transmission. Further, the primary transmission also causes interference at the secondary receiver reducing the success probability of the secondary transmission. However, if the secondary user defers its transmission, then this again has the effect of *decreasing* the expected length of the "PU Busy" period. Note that this cooperation model does not require any changes in terms of advanced physical layer mechanisms and is consistent with the traditional FCC view of cognitive radio networks.

In both models, the reason why the secondary user may want to cooperate is because this can potentially increase the number of time slots in the future in which the primary user does not have any data to send as compared to a non-cooperative strategy. In the *Cooperative Relay Model*, this creates more opportunities for the secondary user to transmit its own packets. In case of the *Interference Model*, this creates better opportunities for the secondary user to transmit its own packets (due to reduced primary interference). However, the trivial strategy of cooperating whenever possible is not necessarily optimal. For example, in the *Cooperative Relay Model*, this may lead to a scenario where the secondary user does not have enough power left for its own data transmission. Similarly, in the *Interference Model*, this may not maximize the secondary user throughput. Thus, the secondary user needs to make intelligent decisions about cooperation.

C. Control Decisions and Queueing Dynamics

Let $Q_{pu}(t), Q_{su}(t) \in \{0, 1, 2, \ldots\}$ represent the primary and secondary user queues respectively in slot t. New packets arrive at the secondary user according to an i.i.d. process $A_{su}(t)$ of rate λ_{su} packets/slot respectively. We assume that there exists a finite constant A_{max} such that $A_{su}(t) \leq A_{max}$ for all t. Every slot, an admission control decision determines $R_{su}(t)$, the number of new packets to admit into the secondary user queue. Further, every slot, depending on the cooperation model, resource allocation decisions are made as follows. When the primary user is busy (i.e., $Q_{pu}(t) > 0$), the secondary user makes a decision about cooperation. Under the *Cooperative Relay Model*, this represents the secondary user decision on cooperative relaying and the corresponding power allocation $P_{su}(t)$. Under the *Interference Model*, this represents the secondary user decision on deferring its transmission (so that $P_{su}(t) = 0$) or continuing its transmission and the corresponding power allocation $P_{su}(t)$. When the primary user is idle (i.e., $Q_{pu}(t) = 0$), the secondary user makes a decision about its own transmission and the corresponding power allocation $P_{su}(t)$. We assume that in each slot, the secondary user can choose its power allocation $P_{su}(t)$ from a set $\mathcal{P}$ of possible options. Further, this power allocation is subject to a long-term average power constraint P_{avg} and an instantaneous peak power constraint P_{max}. For example, $\mathcal{P}$ may contain only two options $\{0, P_{max}\}$ which represents "Remain Idle" and "Cooperate/Transmit at Full Power". As another example, $\mathcal{P} = [0, P_{max}]$ such that $P_{su}(t)$ can take any value between 0 and P_{max}.

Suppose the primary user is active in slot t and the secondary user allocates power $P(t)$ for cooperative relaying under the *Cooperative Relay Model*. Then the random success/failure outcome of the primary transmission is given by an indicator variable $\mu_{pu}(P(t))$ and the success probability is given by $\Phi_{cr}(P(t)) = \mathbb{E}\{\mu_{pu}(P(t))\}$. The function $\Phi_{cr}(P)$ is known to the network controller and is assumed to be *non-decreasing* in P. However, the value of the random outcome $\mu_{pu}(P(t))$ may not be known beforehand. Note that setting

$P(t) = 0$ corresponds to the secondary user not employing any cooperative relaying in this model.

Similarly, under the *Interference Model*, if the secondary user allocates power $P(t)$ to transmit concurrently with the primary user, then the success probability of the primary transmission is given by $\Phi_{in}(P(t)) = \mathbb{E}\{\mu_{pu}(P(t))\}$. The function $\Phi_{in}(P)$ is known to the network controller and is assumed to be *non-increasing* in P. Note that setting $P(t) = 0$ in this model corresponds to a cooperation from the secondary user (by deferring its own transmission).

We assume that λ_{pu} is such that it can be supported even when the secondary user never cooperates, i.e., $\lambda_{pu} < \Phi_{cr}(0)$ in the *Cooperative Relay Model* and $\lambda_{pu} < \Phi_{in}(P_{max})$ in the *Interference Model*. This means that the primary user queue is stable even if there is no cooperation. Further, for all k, the frame length $T[k] \geq 1$ and there exist finite constants T_{min}, T_{max} such that under all control policies, we have $1 \leq T_{min} \leq \mathbb{E}\{T[k]\} \leq T_{max}$. For example, in the *Cooperative Relay Model*, T_{min} can be chosen to be the expected frame length when the secondary user always cooperates with full power while T_{max} can be chosen to be the expected frame length when the secondary user never cooperates. Using Little's Theorem, we have $\frac{T_{min}}{T_{min}+1/\lambda_{pu}} = \frac{\lambda_{pu}}{\Phi_{cr}(P_{max})}$. Similarly, we have $\frac{T_{max}}{T_{max}+1/\lambda_{pu}} = \frac{\lambda_{pu}}{\Phi_{cr}(0)}$. Using these, we have:

$$T_{min} \triangleq \frac{\Phi_{cr}(P_{max})}{(\Phi_{cr}(P_{max}) - \lambda_{pu})\lambda_{pu}}, \; T_{max} \triangleq \frac{\Phi_{cr}(0)}{(\Phi_{cr}(0) - \lambda_{pu})\lambda_{pu}}$$

Finally, there exists a finite constant D such that the expectation of the second moment of a frame size, $\mathbb{E}\left\{T^2[k]\right\}$, satisfies the following for all k, regardless of the policy[1]:

$$\mathbb{E}\left\{T^2[k]\right\} \leq D \tag{1}$$

This follows from the assumption that the primary user queue is stable even if there is no cooperation.

If the primary user is idle in slot t and the secondary user allocates power $P(t)$ for its own transmission, the random success/failure outcome of the transmission is given by an indicator variable $\mu_{su}(P(t))$ and the success probability is given by $\Psi_{idle}(P(t)) = \mathbb{E}\{\mu_{su}(P(t))\}$. Recall that under the *Cooperative Relay Model*, the secondary user can transmit its data only when the primary user is idle. However, under the *Interference Model*, the secondary user can transmit concurrently with the primary user. In this case, the success probability of the secondary transmission is given by $\Psi_{busy}(P(t))$. Since primary transmission can interfere at the secondary receiver, we assume that $\Psi_{busy}(P) \leq \Psi_{idle}(P)$ for all P. The functions $\Psi_{idle}(P)$ and $\Psi_{busy}(P)$ are also known to the network controller and are assumed to be non-decreasing in P.

Given these control decisions, the primary and secondary user queues evolve as follows:

$$Q_{pu}(t+1) = [Q_{pu}(t) - \mu_{pu}(P(t)), 0]^+ + A_{pu}(t) \tag{2}$$
$$Q_{su}(t+1) = [Q_{su}(t) - \mu_{su}(P(t)), 0]^+ + R_{su}(t) \tag{3}$$

where $[a, b]^+ \triangleq \max[a, b]$ and $R_{su}(t) \leq A_{su}(t)$.

[1]In [27], we exactly compute such a D that satisfies (1).

D. Control Objective

Consider any control algorithm that makes admission control decision $R_{su}(t)$ and power allocation $P(t)$ every slot subject to the constraints described in Section II-C. Define the time-average rate of the secondary user's admitted traffic under this algorithm as follows: $\overline{R}_{su} \triangleq \lim_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\{R_{su}(\tau)\}$, where the expectation is with respect to the potential randomness of the control algorithm. Define the time-average power allocation $\overline{P}_{su}$ and service rate $\overline{\mu}_{su}$ similarly. Assuming for the time being that these limits exist, our goal is to design a joint admission control and power allocation policy that stabilizes the secondary user queue and maximizes its throughput subject to its average and peak power constraints and the scheduling constraints imposed by the model. Formally, this can be stated as a stochastic optimization problem as follows:

$$\begin{aligned} \text{Maximize: } & \overline{R}_{su} \\ \text{Subject to: } & 0 \leq R_{su}(t) \leq A_{su}(t) \; \forall t \\ & \overline{R}_{su} \leq \overline{\mu}_{su}, \overline{P}_{su} \leq P_{avg}, P(t) \in \mathcal{P} \; \forall t \\ & \text{Scheduling Constraints} \end{aligned} \tag{4}$$

where the constraint $\overline{R}_{su} \leq \overline{\mu}_{su}$ ensures rate stability of the secondary user queue.

It will be useful to define the primary queue backlog $Q_{pu}(t)$ as the "state" for this control problem. This is because the state of this queue (being zero or nonzero) affects the control options as described before. Note that the control decisions on cooperation affect the dynamics of this queue. Therefore, problem (4) is an instance of a constrained Markov decision problem [1]. It is well known that in order to obtain an optimal control policy, it is sufficient to consider only the class of stationary, randomized policies that take control actions only as a function of the current system state (and independent of past history). A general control policy in this class is characterized by a stationary probability distribution over the control action set for each system state. Let υ^* denote the optimal value of the objective in (4). Then using standard results on constrained Markov decision problems [1]–[3], we have the following:

Lemma 1: (Optimal Stationary, Randomized Policy): There exists a stationary, randomized policy *STAT* that takes feasible control decisions $R_{su}^{stat}(t), P_{su}^{stat}(t)$ every slot purely as a (possibly randomized) function of the current state $Q_{pu}(t)$ while satisfying the constraints $R_{su}^{stat}(t) \leq A_{su}(t), P_{su}^{stat}(t) \in \mathcal{P}$ for all t and provides the following guarantees:

$$\overline{R}_{su}^{stat} = \upsilon^* \tag{5}$$
$$\overline{R}_{su}^{stat} \leq \overline{\mu}_{su}^{stat} \tag{6}$$
$$\overline{P}_{su}^{stat} \leq P_{avg} \tag{7}$$

where $\overline{R}_{su}^{stat}, \overline{\mu}_{su}^{stat}, \overline{P}_{su}^{stat}$ denote the time-averages under this policy.

We note that the conventional techniques to solve (4) that are based on dynamic programming [4] require either extensive knowledge of the system dynamics or learning based approaches that suffer from large convergence times. Motivated by the recently developed extension to the technique of Lyapunov optimization in [8]–[10], we take an different approach to this problem in the next section.

III. Solution Using The "Drift-plus-Penalty" Ratio Method

Recall that the start of the k^{th} frame, t_k, is defined as the first slot when the primary user becomes idle after the "PU Busy" period of the $(k-1)^{th}$ frame. Let $Q_{su}(t_k)$ denote the secondary user queue backlog at time t_k. Also let $P(t)$ be the power expenditure incurred by the secondary user in slot t. For notational convenience, in the following we will denote $\mu_{su}(P(t))$ by $\mu_{su}(t)$ noting the dependence on $P(t)$ is implicit. Then the queueing dynamics of $Q_{su}(t_k)$ satisfies the following:

$$Q_{su}(t_{k+1}) \leq [Q_{su}(t_k) - \sum_{t=t_k}^{t_{k+1}-1} \mu_{su}(t), 0]^+ + \sum_{t=t_k}^{t_{k+1}-1} R_{su}(t) \tag{8}$$

where $R_{su}(t)$ denotes the number of new packets admitted in slot t and t_{k+1} denotes the start of the $(k+1)^{th}$ frame. The above expression has an inequality because it may be possible to serve the packets admitted in the k^{th} frame during that frame itself.

In order to meet the time average power constraint, we make use of a virtual power queue $X_{su}(t_k)$ [25] which evolves over frames as follows:

$$X_{su}(t_{k+1}) = [X_{su}(t_k) - T[k]P_{avg} + \sum_{t=t_k}^{t_{k+1}-1} P(t), 0]^+ \tag{9}$$

where $T[k] = t_{k+1} - t_k$ is the length of the k^{th} frame. Recall that $T[k]$ is a (random) function of the control decisions taken during the k^{th} frame.

In order to construct an optimal dynamic control policy, we use the technique of [8]–[10] where a ratio of "drift-plus-penalty" is maximized over every frame. Specifically, let $\boldsymbol{Q}(t_k) = (Q_{su}(t_k), X_{su}(t_k))$ denote the queueing state of the system at the start of the k^{th} frame. As a measure of the congestion in the system, we use a Lyapunov function $L(\boldsymbol{Q}(t_k)) \stackrel{\triangle}{=} \frac{1}{2}[Q_{su}^2(t_k) + X_{su}^2(t_k)]$. Define the drift $\Delta(t_k)$ as the conditional expected change in $L(\boldsymbol{Q}(t_k))$ over frame k:

$$\Delta(t_k) \stackrel{\triangle}{=} \mathbb{E}\left\{L(\boldsymbol{Q}(t_{k+1})) - L(\boldsymbol{Q}(t_k)) | \boldsymbol{Q}(t_k)\right\} \tag{10}$$

Then, using (8) and (9), we can bound $\Delta(t_k)$ as follows:

$$\Delta(t_k) \leq B - Q_{su}(t_k)\mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} [\mu_{su}(t) - R_{su}(t)] | \boldsymbol{Q}(t_k)\right\} - X_{su}(t_k)\mathbb{E}\left\{T[k]P_{avg} - \sum_{t=t_k}^{t_{k+1}-1} P(t) | \boldsymbol{Q}(t_k)\right\} \tag{11}$$

where B is a finite constant that satisfies the following for all k and $\boldsymbol{Q}(t_k)$ under any control algorithm:

$$B \geq \frac{1}{2}\mathbb{E}\Bigg\{\Big(\sum_{t=t_k}^{t_{k+1}-1} \mu_{su}(t)\Big)^2 + \Big(\sum_{t=t_k}^{t_{k+1}-1} R_{su}(t)\Big)^2 + \Big(\sum_{t=t_k}^{t_{k+1}-1} P(t) - T[k]P_{avg}\Big)^2 | \boldsymbol{Q}(t_k)\Bigg\}$$

Using the fact that $\mu_{su}(t) \leq 1, P(t) \leq P_{max}$ for all t, and using the fact (1), it follows that choosing B as follows satisfies the above:

$$B = \frac{D[1 + A_{max}^2 + (P_{max} - P_{avg})^2]}{2} \tag{12}$$

Adding a penalty term $-V\mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} R_{su}(t) | \boldsymbol{Q}(t_k)\right\}$ (where $V > 0$ is a control parameter that affects a utility-delay trade-off as shown in Theorem 1) to both sides of the above and rearranging yields:

$$\Delta(t_k) - V\mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} R_{su}(t) | \boldsymbol{Q}(t_k)\right\} \leq B + (Q_{su}(t_k) - V) \times \mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} R_{su}(t) | \boldsymbol{Q}(t_k)\right\} - X_{su}(t_k)\mathbb{E}\left\{T[k]P_{avg} | \boldsymbol{Q}(t_k)\right\} - \mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} \Big(Q_{su}(t_k)\mu_{su}(t) - X_{su}(t_k)P(t)\Big) | \boldsymbol{Q}(t_k)\right\} \tag{13}$$

Minimizing the ratio of an upper bound on the right hand side of the above expression and the expected frame length over all control options leads to the following *Frame-Based-Drift-Plus-Penalty-Algorithm*. In each frame $k \in \{1, 2, 3, \ldots\}$, do the following:

1) *Admission Control*: For all $t \in \{t_k, t_k + 1, \ldots, t_{k+1} - 1\}$, choose $R_{su}(t)$ as follows:

$$R_{su}(t) = \begin{cases} A_{su}(t) & \text{if } Q_{su}(t) \leq V \\ 0 & \text{else} \end{cases} \tag{14}$$

2) *Resource Allocation*: Choose a policy that maximizes the following ratio:

$$\frac{\mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} \Big(Q_{su}(t_k)\mu_{su}(t) - X_{su}(t_k)P(t)\Big) | \boldsymbol{Q}(t_k)\right\}}{\mathbb{E}\left\{T[k] | \boldsymbol{Q}(t_k)\right\}} \tag{15}$$

Specifically, every slot t of the frame, the policy observes the queue values $Q_{su}(t_k)$ and $X_{su}(t_k)$ at the beginning of the frame and selects a secondary user power $P(t)$ subject to the constraint $P(t) \in \mathcal{P}$ and the scheduling constraints of the cooperation model being used. For example, under the *Cooperative Relay Model*, this corresponds to the constraint on transmitting own data vs. cooperation depending on whether slot t is in the "PU Idle" or "PU Busy" period of the frame. The objective is to maximize the above frame-based ratio of expectations. Recall that the frame size $T[k]$ is influenced by the policy through the success probabilities that are determined by secondary user power selections. Further recall that these success probabilities are different during the "PU Idle" and "PU Busy" periods of the frame. An explicit policy that maximizes this expectation under both cooperation models is given in the next section.

3) *Queue Update*: After implementing this policy, update the queues as in (3) and (9).

From the above, it can be seen that the admission control part (14) is a simple threshold-based decision that does not require any knowledge of the arrival rates λ_{su} or λ_{pu}. In the next section, we present an explicit solution to the maximizing policy for the resource allocation in (15) under both cooperation models and show that, remarkably, it also does not require knowledge of λ_{su} or λ_{pu} and can be computed easily. We then analyze the performance of the *Frame-Based-Drift-Plus-Penalty-Algorithm* in Section V.

IV. The Maximizing Policy of (15)

Under both cooperation models, the policy that maximizes (15) uses only two numbers that we call P_0^* and P_1^*, defined as follows. P_0^* is given by the solution to the following optimization problem:

$$\begin{aligned} &\text{Maximize: } Q_{su}(t_k)\Psi_{idle}(P_0) - X_{su}(t_k)P_0 \\ &\text{Subject to: } P_0 \in \mathcal{P} \end{aligned} \tag{16}$$

Let $\theta^* \triangleq Q_{su}(t_k)\Psi_{idle}(P_0^*) - X_{su}(t_k)P_0^*$ denote the value of the objective of (16) under the optimal solution. Then, under the *Cooperative Relay Model*, P_1^* is given by the solution to the following optimization problem:

$$\begin{aligned} &\text{Minimize: } \frac{\theta^* + X_{su}(t_k)P_1}{\Phi_{cr}(P_1)} \\ &\text{Subject to: } P_1 \in \mathcal{P} \end{aligned} \tag{17}$$

Under the *Interference Model*, P_1^* is given by the solution to the following optimization problem:

$$\begin{aligned} &\text{Minimize: } \frac{\theta^* - Q_{su}(t_k)\Psi_{busy}(P_1) + X_{su}(t_k)P_1}{\Phi_{in}(P_1)} \\ &\text{Subject to: } P_1 \in \mathcal{P} \end{aligned} \tag{18}$$

Note that (16), (17), and (18) are simple optimization problems in a single variable and can be solved efficiently. Given P_0^* and P_1^*, on every slot t of frame k, the policy that maximizes (15) chooses power $P(t)$ as follows:

$$P(t) = \begin{cases} P_0^* & \text{if } Q_{pu}(t) = 0 \\ P_1^* & \text{if } Q_{pu}(t) > 0 \end{cases} \tag{19}$$

That is, the secondary user uses the constant power P_0^* for its own transmission during the "PU Idle" period of the frame, and uses constant power P_1^* for cooperative relaying (under the *Cooperative Relay Model*) or its own transmission (under the *Interference Model*) during all slots of the "PU busy" period of the frame. Note that P_0^* and P_1^* can be computed easily based on the weights $Q_{su}(t_k), X_{su}(t_k)$ associated with frame k, and do not require knowledge of the arrival rates $\lambda_{su}, \lambda_{pu}$.

Our proof that the above decisions maximize (15) has the following parts: First, we show that the decisions that maximize the ratio of expectations in (15) are the same as the optimal decisions in an equivalent infinite horizon Markov decision problem (MDP). Next, we show that the solution to the infinite horizon MDP uses fixed power P_i for each queue state $Q_{pu}(t) = i$ (for $i \in \{0, 1, 2, \ldots\}$). Then, we show that P_i are the same for all $i \geq 1$. Finally, we show that the optimal powers P_0^* and P_1^* are given as above. The detailed proof is given in the next section. For brevity, we focus on the *Cooperative Relay Model* in the following noting that the proof technique applies to the *Interference Model* as well.

A. Proof Details

Here we examine how to solve (15) in detail under the *Cooperative Relay Model*. First, define the state i in any slot $t \in \{t_k, t_k + 1, \ldots, t_{k+1} - 1\}$ as the value of the primary user queue backlog $Q_{pu}(t)$ in that slot. Note that the state at time t_k is 0 since every frame k starts with the "PU Idle" period. Now let $\mathcal{R}$ denote the class of stationary, randomized policies where every policy $r \in \mathcal{R}$ chooses a power allocation $P_i(r) \in \mathcal{P}$ in each state i according to a stationary distribution. It can be shown that it is sufficient to only consider policies in $\mathcal{R}$ to maximize (15). Now suppose a policy $r \in \mathcal{R}$ is implemented on a separate virtual system with fixed $Q_{su}(t_k)$ and $X_{su}(t_k)$ and with the same state dynamics as our model. Specifically, this system is a Markov Chain with the same state space and control actions per state. However, instead of a single frame, this system is run over an infinite horizon. Then, by basic renewal theory [26], we have that maximizing the ratio of expectations in (15) over the course of the frame is identical to maximizing the infinite horizon time-average cost in the virtual system. Using the fact that $\mu_{su}(t) = 0$ for all t when the state $i \geq 1$, this can be expressed as the following unconstrained MDP problem:

$$\begin{aligned} \text{Maximize: } & Q_{su}(t_k)\mathbb{E}\left\{\Psi_{idle}(P_0(r))\right\}\pi_0(r) \\ & - X_{su}(t_k)\sum_{i\geq 0}\mathbb{E}\left\{P_i(r)\right\}\pi_i(r) \\ \text{Subject to: } & r \in \mathcal{R} \end{aligned} \tag{20}$$

where $\pi_i(r)$ is the resulting steady-state probability of being in state i in the virtual system under the stationary, randomized policy r and where the expectations above are with respect to r. Note that well-defined steady-state probabilities $\pi_i(r)$ exist for all $r \in \mathcal{R}$ because we have assumed that $\lambda_{pu} < \Phi_{cr}(0)$ so that even if no cooperation is used, the primary queue is stable and the system is recurrent. Thus, solving (15) is equivalent to solving the unconstrained time-average maximization problem (20) over the class of stationary, randomized policies. We study this problem in the following.

Consider the optimal stationary, randomized policy that maximizes the objective in (20). Let χ_i denote the probability distribution over $\mathcal{P}$ that is used by this policy to choose a power allocation P_i in state i. Let μ_i denote the resulting effective probability of successful primary transmission in state $i \geq 1$. Then we have that $\mu_i = \mathbb{E}_{\chi_i}\{\Phi_{cr}(P_i)\}$ where $\Phi_{cr}(P_i)$ denotes the probability of successful transmission in state i when the secondary user spends power P_i in cooperative transmission with the primary user. Since the system is stable and has a well-defined steady-state distribution, we can write down the detail equations for the Markov Chain that describes the state transitions of the system as follows (See Fig. 3):

$$\begin{aligned} \pi_0\lambda_{pu} &= \pi_1(1-\lambda_{pu})\mu_1 \\ \pi_1\lambda_{pu}(1-\mu_1) &= \pi_2(1-\lambda_{pu})\mu_2 \\ &\vdots \\ \pi_i\lambda_{pu}(1-\mu_i) &= \pi_{i+1}(1-\lambda_{pu})\mu_{i+1} \qquad \forall i \geq 1 \end{aligned}$$

where π_i denotes the steady-state probability of being in state

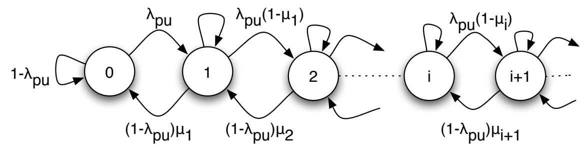


Fig. 3. Birth-Death Markov Chain over the system state where the system state represents the primary user queue backlog.

i under this policy. Summing over all i yields:

$$\lambda_{pu} = \sum_{i\geq 1} \pi_i \mu_i \tag{21}$$

The average power incurred in cooperative transmissions under this policy is given by:

$$\overline{P} = \sum_{i\geq 1} \pi_i \mathbb{E}_{\chi_i}\{P_i\} \tag{22}$$

Now consider an alternate stationary policy that uses the following fixed distribution χ' for choosing control action P' in all states $i \geq 1$:

$$\chi' \triangleq \begin{cases} \chi_1 & \text{with probability } \frac{\pi_1}{\sum_{j\geq 1}\pi_j} \\ \chi_2 & \text{with probability } \frac{\pi_2}{\sum_{j\geq 1}\pi_j} \\ \vdots & \\ \chi_i & \text{with probability } \frac{\pi_i}{\sum_{j\geq 1}\pi_j} \\ \vdots & \end{cases} \tag{23}$$

Let μ' denote the resulting effective probability of a successful primary transmission in any state $i \geq 1$. Note that this is same for all states by the definition (23). Then, we have that:

$$\mu' = \sum_{i\geq 1} \mu_i \frac{\pi_i}{\sum_{j\geq 1}\pi_j} \tag{24}$$

Let π'_i denote the steady-state probability of being in state i under this alternate policy. Note that the system is stable under this alternate policy as well. Thus, using the detail equations for the Markov Chain that describes the state transitions of the system under this policy yields

$$\begin{aligned} \lambda_{pu} &= \sum_{k\geq 1} \pi'_k \mu' = \sum_{k\geq 1} \pi'_k \Big(\sum_{i\geq 1} \mu_i \frac{\pi_i}{\sum_{j\geq 1}\pi_j} \Big) \\ &= \sum_{k\geq 1} \pi'_k \Big(\frac{\sum_{i\geq 1}\mu_i\pi_i}{\sum_{j\geq 1}\pi_j} \Big) = \sum_{k\geq 1} \pi'_k \Big(\frac{\lambda_{pu}}{\sum_{j\geq 1}\pi_j} \Big) \end{aligned} \tag{25}$$

where we used (21) in the last step. This implies that $\sum_{k\geq 1}\pi'_k = \sum_{j\geq 1}\pi_j$ and therefore $\pi'_0 = \pi_0$. Also, the average power incurred in cooperative transmissions under this alternate policy is given by:

$$\begin{aligned} \overline{P}' &= \sum_{k\geq 1} \pi'_k \mathbb{E}_{\chi'}\{P'\} = \sum_{k\geq 1} \pi'_k \Big(\sum_{i\geq 1} \mathbb{E}_{\chi_i}\{P_i\} \frac{\pi_i}{\sum_{j\geq 1}\pi_j} \Big) \\ &= \sum_{k\geq 1} \pi'_k \Big(\frac{\overline{P}}{\sum_{j\geq 1}\pi_j} \Big) = \overline{P} \end{aligned} \tag{26}$$

where we used (22) in the second to last step and $\sum_{k\geq 1}\pi'_k = \sum_{j\geq 1}\pi_j$ in the last step.

Thus, if we choose $\chi' = \chi_0$ in state $i = 0$ and choose χ' as defined in (23) in all other states, it can be seen that *the alternate policy achieves the same time-average value of the objective (20) as the optimal policy*. This implies that to maximize (20), it is sufficient to optimize over the class of stationary policies that use the *same* distribution for choosing P_i for all states $i \geq 1$. Denote this class by $\mathcal{R}'$. Then for all $i > 1$, we have that $\mathbb{E}\{P_i(r)\} = \mathbb{E}\{P_1(r)\}$ for all $r \in \mathcal{R}'$. Using this and the fact that $1 - \pi_0(r) = \sum_{i\geq 1}\pi_i(r)$, (20) can be simplified as follows:

$$\begin{aligned} &\text{Max: } (Q_{su}(t_k)\mathbb{E}\{\Psi_{idle}(P_0(r))\} - X_{su}(t_k)\mathbb{E}\{P_0(r)\})\pi_0(r) \\ &\qquad - X_{su}(t_k)\mathbb{E}\{P_1(r)\}(1-\pi_0(r)) \\ &\text{Subject to: } r \in \mathcal{R}' \end{aligned} \tag{27}$$

where $\pi_0(r)$ is the resulting steady-state probability of being in state 0 and where $\mathbb{E}\{P_1(r)\}$ is the average power incurred in cooperative transmission in state $i = 1$ (same for all states $i \geq 1$). Next, note that the control decisions taken by the secondary user in state $i = 0$ do not affect the length of the frame and therefore $\pi_0(r)$. Further, the expectations can be removed. Therefore the first term in the problem above can be maximized separately as follows:

$$\begin{aligned} &\text{Maximize: } Q_{su}(t_k)\Psi_{idle}(P_0) - X_{su}(t_k)P_0 \\ &\text{Subject to: } P_0 \in \mathcal{P} \end{aligned} \tag{28}$$

This is the same as (16). Let P_0^* denote the optimal solution to (28) and let $\theta^* = Q_{su}(t_k)\Psi_{idle}(P_0^*) - X_{su}(t_k)P_0^*$ denote the value of the objective of (28) under the optimal solution. Note that we must have that $\theta^* \geq 0$ because the value of the objective when the secondary user chooses $P_0 = 0$ (i.e., stays idle) is 0. Then, (27) can be written as:

$$\begin{aligned} &\text{Maximize: } \theta^*\pi_0(r) - X_{su}(t_k)\mathbb{E}\{P_1(r)\}(1-\pi_0(r)) \\ &\text{Subject to: } r \in \mathcal{R}' \end{aligned} \tag{29}$$

The effective probability of a successful primary transmission in any state $i \geq 1$ is $\mathbb{E}\{\Phi_{cr}(P_1(r))\}$. Using Little's Theorem, we have $\pi_0(r) = 1 - \frac{\lambda_{pu}}{\mathbb{E}\{\Phi_{cr}(P_1(r))\}}$. Using this and rearranging the objective in (29) and ignoring the constant terms, we have the following equivalent problem:

$$\begin{aligned} &\text{Minimize: } \frac{\theta^* + X_{su}(t_k)\mathbb{E}\{P_1(r)\}}{\mathbb{E}\{\Phi_{cr}(P_1(r))\}} \\ &\text{Subject to: } r \in \mathcal{R}' \end{aligned} \tag{30}$$

It can be shown that it is sufficient to consider only *deterministic* power allocations to solve (30) (see, for example, [10, Section 7.3.2]). This yields the following problem:

$$\begin{aligned} &\text{Minimize: } \frac{\theta^* + X_{su}(t_k)P_1}{\Phi_{cr}(P_1)} \\ &\text{Subject to: } P_1 \in \mathcal{P} \end{aligned} \tag{31}$$

This is the same as (17). Note that solving this problem does not require knowledge of λ_{pu} or λ_{su} and can be solved easily for general power allocation options $\mathcal{P}$. We present an example that admits a particularly simple solution to this problem.

Suppose $\mathcal{P} = \{0, P_{max}\}$ so that the secondary user can either cooperate with full power P_{max} or not cooperate (with power expenditure 0) with the primary user. Then, the optimal solution to (31) can be calculated by comparing the value of its objective for $P_1 \in \{0, P_{max}\}$. This yields the following simple threshold-based rule:

$$P_1^* = \begin{cases} 0 & \text{if } X_{su}(t_k) \geq \frac{\theta^*(\Phi_{cr}(P_{max}) - \Phi_{cr}(0))}{P_{max}\Phi_{cr}(0)} \\ P_{max} & \text{else} \end{cases} \quad (32)$$

We also note that this threshold can be computed without any knowledge of the input rates $\lambda_{pu}, \lambda_{su}$.

To summarize, the overall solution to (15) is given by the pair (P_0^*, P_1^*) where P_0^* denotes the power allocation used by the secondary user for its own transmission when the primary user is idle and P_1^* denotes the power used by the secondary user for cooperative transmission (under the *Cooperative Relay Model*) or its own transmission (under the *Interference Model*). Note that these values remain fixed for the entire duration of frame k. However, these can change from one frame to another depending on the values of the queues $Q_{su}(t_k), X_{su}(t_k)$. The computation of (P_0^*, P_1^*) can be carried out using a two-step process as follows:

1) First, compute P_0^* by solving problem (16). Let θ^* be the value of the objective of (16) under the optimal solution P_0^*.
2) Then compute P_1^* by solving problem (17) for the *Cooperative Relay Model* or (18) for the *Interference Model*.

It is interesting to note that in order to implement this algorithm, the secondary user does not require knowledge of the current queue backlog value of the primary user. Rather, it only needs to know the values of its own queues and whether the current slot is in the "PU Idle" or "PU Busy" part of the frame. This is quite different from the conventional solution to the MDP (4) which is typically a different randomized policy for each value of the state (i.e., the primary queue backlog).

V. Performance Analysis

To analyze the performance of the *Frame-Based-Drift-Plus-Penalty-Algorithm*, we compare its Lyapunov drift with that of the optimal stationary, randomized policy *STAT* of Lemma 1.

Theorem 1: (Performance Theorem) Suppose the *Frame-Based-Drift-Plus-Penalty-Algorithm* is implemented over all frames $k \in \{1, 2, 3, \ldots\}$ with initial condition $Q_{su}(0) = 0, X_{su}(0) = 0$ and with a control parameter $V > 0$. Let $\mu_{su}^{fab}(t), P_{su}^{fab}(t)$ denote the resource allocation decisions under this algorithm. Then, we have:

1) The secondary user queue backlog $Q_{su}(t)$ is upper bounded for all t:

$$Q_{su}(t) \leq Q_{max} \triangleq A_{max} + V \quad (33)$$

2) The virtual power queue $X_{su}(t_k)$ is mean rate stable, i.e.,

$$\lim_{K\to\infty} \frac{\mathbb{E}\{X_{su}(t_K)\}}{K} = 0 \quad (34)$$

Further, we have:

$$\limsup_{K\to\infty} \left(\frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left\{ \sum_{t=t_k}^{t_{k+1}-1} (P_{su}^{fab}(t) - P_{avg}) \right\} \right) \leq 0 \quad (35)$$

$$\limsup_{K\to\infty} \frac{\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} P_{su}^{fab}(t)\right\}}{\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\{T[k]\}} \leq P_{avg} \quad (36)$$

3) The time-average secondary user throughput (defined over frames) satisfies the following bound for all $K > 0$:

$$\frac{\sum_{k=1}^{K} \mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} R_{su}^{fab}(t)\right\}}{\sum_{k=1}^{K} \mathbb{E}\{T[k]\}} \geq v^* - \frac{B+C}{VT_{min}} \quad (37)$$

where $B = \frac{D[1 + A_{max}^2 + (P_{max} - P_{avg})^2]}{2}$ and $C = \frac{D(A_{max}+1)A_{max}}{2}$.

Theorem 1 shows that the time-average secondary user throughput can be pushed to within $O(1/V)$ of the optimal value with a trade-off in the worst case queue backlog. By Little's Theorem, this leads to an $O(1/V, V)$ utility-delay tradeoff.

Proof: Omitted for brevity. See [27] for full proof. ■

VI. Extensions to Multiple Secondary Users

Consider the scenario with one primary user as before, but with $N > 1$ secondary users. For brevity, we only consider the *Cooperative Relay Model*. The primary user channel occupancy process evolves as before where the secondary users can transmit their own data only when the primary user is idle. However, they may cooperatively transmit with the primary user to increase its transmission success probability. In general, multiple secondary users may cooperatively transmit with the primary in one timeslot. However, for simplicity, here we assume that at most one secondary user can take part in a cooperative transmission per slot. Further, we also assume that at most one secondary user can transmit its data when the primary user is idle. Our formulation can be easily extended to this scenario. Let $\mathcal{P}_i$ denote the set of power allocation options for secondary user i. Suppose each secondary user i is subject to average and peak power constraints $P_{avg,i}$ and $P_{max,i}$ respectively. Let $\Phi_{cr,i}(P)$ denote the success probability of the primary transmission when secondary user i spends power P in cooperative transmission. Also, let $\Psi_{idle,i}(P)$ denote the success probability of the secondary user i when it spends power P for its own transmission in the "PU Idle" phase. Now consider the objective of maximizing the sum total throughput of the secondary users subject to each user's average and peak power constraints and the scheduling constraints of the model. In order to apply the "drift-plus-penalty" ratio method, we use the following queues:

$$Q_i(t_{k+1}) \leq [Q_i(t_k) - \sum_{t=t_k}^{t_{k+1}-1} \mu_i(t), 0]^+ + \sum_{t=t_k}^{t_{k+1}-1} R_i(t) \quad (38)$$

$$X_i(t_{k+1}) = [X_i(t_k) - T[k]P_{avg,i} + \sum_{t=t_k}^{t_{k+1}-1} P_i(t), 0]^+ \quad (39)$$

where $Q_i(t_k)$ is the queue backlog of secondary user i at the beginning of the k^{th} frame, $\mu_i(t)$ is the service rate of secondary user i in slot t, $R_i(t)$ and $P_i(t)$ denote the number of new packets admitted and the power expenditure incurred by the secondary user i in slot t. Finally, t_{k+1} denotes the start of the $(k+1)^{th}$ frame and $T[k] = t_{k+1} - t_k$ is the length of the k^{th} frame as before.

Let $\boldsymbol{Q}(t_k) = (Q_1(t_k), \ldots, Q_N(t_k), X_1(t_k), \ldots, X_N(t_k))$ denote the queueing state of the system at the start of the k^{th} frame. Using a Lyapunov function $L(\boldsymbol{Q}(t_k)) \overset{\Delta}{=} \frac{1}{2}\Big[\sum_{i=1}^N Q_i^2(t_k) + \sum_{i=1}^N X_i^2(t_k)\Big]$ and following the steps in Section III yields the following *Multi-User Frame-Based-Drift-Plus-Penalty-Algorithm*. In each frame $k \in \{1, 2, 3, \ldots\}$, do the following:

1) *Admission Control*: For all $t \in \{t_k, t_k+1, \ldots, t_{k+1}-1\}$, for each secondary user $i \in \{1, 2, \ldots, N\}$, choose $R_i(t)$ as follows:

$$R_i(t) = \begin{cases} A_i(t) & \text{if } Q_i(t) \le V \\ 0 & \text{else} \end{cases} \tag{40}$$

where $A_i(t)$ is the number of new arrivals to secondary user i in slot t.

2) *Resource Allocation*: Choose a policy that maximizes the following ratio:

$$\frac{\sum_{i=1}^N \mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1}(Q_i(t_k)\mu_i(t) - X_i(t_k)P_i(t))|\boldsymbol{Q}(t_k)\right\}}{\mathbb{E}\left\{T[k]|\boldsymbol{Q}(t_k)\right\}} \tag{41}$$

3) *Queue Update*: After implementing this policy, update the queues as in (38) and (39).

Similar to the basic model, this algorithm can be implemented without any knowledge of the arrival rates λ_i or λ_{pu}. Further, using the techniques developed in Section IV, it can be shown that the solution to (41) can be computed in two steps as follows. First, we solve the following problem for each $i \in \{1, 2, \ldots, N\}$:

$$\begin{aligned} &\text{Maximize: } Q_i(t_k)\Psi_{idle,i}(P) - X_i(t_k)P \\ &\text{Subject to: } P \in \mathcal{P}_i \end{aligned} \tag{42}$$

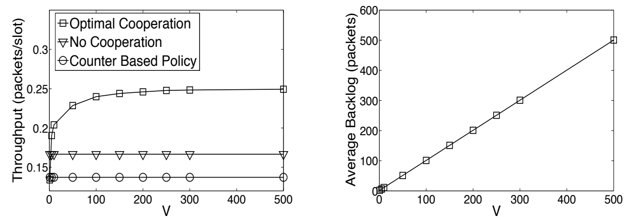
Let P_0^* denote the optimal solution to (42) achieved by user i^* and let θ^* denote the optimal objective value. This means user i^* transmits on all idle slots of frame k with power P_0^*. Next, to determine the optimal cooperative transmission strategy, we solve the following problem for each $i \in \{1, 2, \ldots, N\}$:

$$\begin{aligned} &\text{Minimize: } \frac{\theta^* + X_i(t_k)P}{\Phi_{cr,i}(P)} \\ &\text{Subject to: } P \in \mathcal{P}_i \end{aligned} \tag{43}$$

Let P_1^* denote the optimal solution to (43) achieved by user j^*. This means user j^* cooperatively relays on all busy slots of frame k with power P_1^*, while all others are idle.

VII. Simulations

In this section, we evaluate the performance of the *Frame-Based-Drift-Plus-Penalty-Algorithm* using simulations. We consider the *Cooperative Relay Model* of Section II with one primary and one secondary user. The set $\mathcal{P}$ consists of only two options $\{0, P_{max}\}$. We assume that $P_{avg} = 0.5$ and $P_{max} = 1$. We set $\Phi_{cr}(0) = 0.6$ and $\Phi_{cr}(P_{max}) = 0.8$. For simplicity, we assume that $\Psi_{idle}(P_{max}) = 1$.



(a) Average Throughput of Secondary User vs. V.

(b) Average Secondary User Queue Occupancy vs. V.

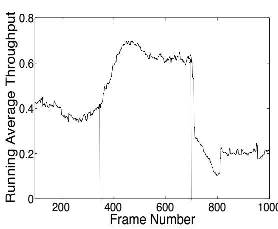
In the first set of simulations, we fix the input rates $\lambda_{pu} = \lambda_{su} = 0.5$ packets/slot. For these parameters, we can compute the optimal offline solution by linear programming. This yields the maximum secondary user throughput as 0.25 packets/slot. We now simulate the *Frame-Based-Drift-Plus-Penalty-Algorithm* for different values of the control parameter V over 1000 frames. In Fig. 4(a), we plot the average throughput achieved by the secondary user over this period. It can be seen that the average throughput increases with V and converges to the optimal value 0.25 packets/slot, with the difference exhibiting a $O(1/V)$ behavior as predicted by Theorem 1. In Fig. 4(b), we plot the average queue backlog of the secondary user over this period. It can be see that the average queue backlog grows linearly in V, again as predicted by Theorem 1. Also, for all V, the average secondary user power consumption over this period was found not to exceed $P_{avg} = 0.5$ units/slot.

For comparison, we also simulate three alternate algorithms. In the first algorithm "No Cooperation", the secondary user never cooperates with the primary user and only attempts to maximize its throughput over the resulting idle periods. The secondary user throughput under this algorithm was found to be 0.166 packets/slot as shown in Fig. 4(a). Note that using Little's Theorem, the resulting fraction of time the primary user is idle is $1 - \lambda_{pu}/\Phi_{cr}(0) = 1 - 0.5/0.6 = 0.166$. This limits the maximum secondary user throughput under the "No Cooperation" case to 0.166 packets/slot.

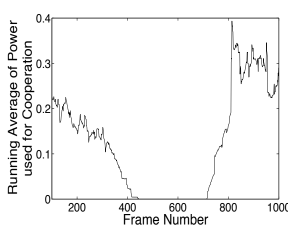
In the second algorithm, we consider the "Always Cooperate" case where the secondary user always cooperates with the primary user. For the example under consideration, this uses up all the secondary user power and thus, the secondary user achieves zero throughput.

In the third algorithm "Counter Based Policy", a running average of the total secondary user power consumption so far is maintained. In each slot, the secondary user decides to transmit/cooperate only if this running average is smaller than P_{avg}. The maximum secondary user throughput under this algorithm was found to be 0.137 packets/slot. This demonstrates that simply satisfying the average power constraint is not sufficient to achieve maximum throughput. For example, it may be the case that under the "Counter Based Policy", the running average condition is usually satisfied when the primary user is busy. This causes the secondary user to cooperate. However, by the time the primary user next becomes idle, the running average exceeds P_{avg} so that the secondary user does not transmit its own data. In contrast, the *Frame-Based-Drift-Plus-Penalty-Algorithm* is able to find the opportune moments to cooperate/transmit optimally.

In the second set of simulations, we fix the input rate $\lambda_{su} = 0.8$ packets/slot, $V = 500$, and simulate the *Frame-Based-Drift-Plus-Penalty-Algorithm* over 1000 frames. At the start of the simulation, we set $\lambda_{pu} = 0.4$ packets/slot. The values of the other parameters remain the same. However, during the course of the simulation, we change λ_{pu} to 0.2 packets/slot after the first 350 frames and then again to 0.55 packets/slot after the first 700 frames. In Figs. 4(c) and 4(d), we plot the running average (over 100 frames) of the secondary user throughput and the average power used for cooperation. These show that the *Frame-Based-Drift-Plus-Penalty-Algorithm* automatically adapts to the changes in λ_{pu}. Further, it quickly approaches the optimal performance corresponding to the new λ_{pu} by adaptively spending more or less power (as required) on cooperation. For example, when λ_{pu} reduces to 0.2 packets/slot after frame number 350, the fraction of time the primary is idle even with no cooperation is $1 - 0.2/0.6 = 0.66$. With $P_{avg} = 0.5$, there is no need to cooperate anymore. This is precisely what the *Frame-Based-Drift-Plus-Penalty-Algorithm* does as shown in Fig. 4(d). Similarly, when when λ_{pu} increases to 0.55 packets/slot after frame number 700, the *Frame-Based-Drift-Plus-Penalty-Algorithm* starts to spend more power on cooperative transmissions.



(c) Running Average of Secondary User Throughput over Frames.



(d) Running Average of Power used by the Secondary User for Cooperative Transmissions over Frames.

VIII. Conclusions

In this paper, we studied the problem of opportunistic cooperation in a cognitive femtocell network. Specifically, we considered two models for such cooperation. In both models, a secondary user can cooperate with the primary user to increase the latter's transmission success probability. In return, the secondary user can get more opportunities for transmitting its own data when the primary user is idle. A key feature of this problem is that here, the evolution of the system state depends on the control actions taken by the secondary user. This dependence makes it a constrained Markov decision problem (MDP). Traditional MDP solutions require either extensive knowledge of the system dynamics or learning based approaches that suffer from large convergence times. However, using the technique of Lyaunov optimization, we designed a novel greedy and online control algorithm that overcomes these challenges and is provably optimal.

References

[1] E. Altman. *Constrained Markov Decision Processes*. Boca Raton, FL: Chapman and Hall/CRC Press, 1999.

[2] M. L. Puterman. *Markov Decision Processes*. John Wiley & Sons, 2005.

[3] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[4] D. P. Bertsekas. *Dynamic Programming and Optimal Control, vols. 1 & 2*, Belmont, MA: Athena Scientific, 2007.

[5] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar. An on-line learning algorithm for energy efficient delay constrained scheduling over a fading channel. *IEEE Journal on Selected Areas in Communications*, 26(4):732-742, May 2008.

[6] F. Fu and M. van der Schaar. Decomposition principles and online learning in cross-layer optimization for delay-sensitive applications. *IEEE Trans. Signal Processing*, 58(3):1401-1415, March 2010.

[7] A. Galindo-Serrano, L. Giupponi, and M. Dohler. Cognition and docition in OFDMA-based femtocell networks. *Proc. GLOBECOM*, Dec. 2010.

[8] C. Li and M. J. Neely. Network utility maximization over partially observable markovian channels. *arXiv:1008.3421v1*, Aug. 2010.

[9] M. J. Neely. Dynamic optimization and learning for renewal systems. *Proc. Asilomar Conference*, Nov. 2010.

[10] M. J. Neely. *Stochastic Network Optimization with Application to Communication & Queueing Systems*. Morgan&Claypool, 2010.

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

[12] Q. Zhao and B. Sadler. A survey of dynamic spectrum access. *IEEE Signal Processing Magazine*, 24(3):79-89, May 2007.

[13] R. Urgaonkar and M. J. Neely. Opportunistic scheduling with reliability guarantees in cognitive radio networks. *IEEE Trans. Mobile Computing*, 8(6):766-777, June 2009.

[14] A. J. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proc. of the IEEE*, 97(5):894-914, May 2009.

[15] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[16] O. Simeone, Y. Bar-Ness, and U. Spagnolini. Stable throughput of cognitive radios with and without relaying capability. *IEEE Trans. Communications*, 55(12):2351-2360, Dec. 2007.

[17] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz. Spectrum leasing to cooperating secondary ad hoc networks. *IEEE JSAC Special Issue on Cognitive Radio: Theory and Applications*, 26(1):203-213, Jan. 2008.

[18] J. Zhang and Q. Zhang. Stackelberg game for utility-based cooperative cognitive radio networks. *Proc. ACM MobiHoc*, May 2009.

[19] I. Krikidis, J. N. Laneman, J. Thompson, and S. McLaughlin. Protocol design and throughput analysis for multi-user cognitive cooperative systems. *IEEE Trans. Wireless Commun.*, 8(9):4740-4751, Sept. 2009.

[20] B. Rong, I. Krikidis, and A. Ephremides. Network-level cooperation with enhancements based on the physical layer. *IEEE Information Theory Workshop*, Cairo, Egypt, Jan. 2010.

[21] M. Levorato, U. Mitra, and M. Zorzi. Cognitive interference management in retransmission-based wireless networks. *Proc. 47th Allerton Conference on Communication, Control, and Computing*, Sept. 2009.

[22] G. Gur, S. Bayhan, and F. Alagoz. Cognitive femtocell networks: An overlay architecture for localized dynamic spectrum access. *IEEE Wireless Communications*, 17(4):62-70, Aug. 2010.

[23] Z. Bharucha, H. Haas, A. Saul, and G. Auer. Throughput enhancement through femto-cell deployment. *European Transactions on Telecommunications*, 21(5):469-477, Aug. 2010.

[24] L. Georgiadis, M. J. Neely, and L. Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, 1(1):1-149, 2006.

[25] M. J. Neely. Energy optimal control for time varying wireless networks. *IEEE Trans. Inform. Theory*, 52(7):2915-2934, July 2006.

[26] R. Gallager. *Discrete Stochastic Processes*. Kluwer Academic Publishers, Boston, 1996.

[27] R. Urgaonkar and M. J. Neely. Opportunistic cooperation in cognitive femtocell networks. *arXiv:1103.1401*, March 2011.