ABSTRACT

Title of dissertation: CROSS-LAYER ASPECTS OF COGNITIVE WIRELESS NETWORKS

Anthony A. Fanous, Doctor of Philosophy, 2013

Dissertation directed by: Professor Anthony Ephremides
Department of Electrical and Computer Engineering

We study cognitive wireless networks from a cross-layer perspective, where we investigate the effects of the PHY layer parameters and enhancements on the MAC layer performance. We quantify the benefit of using sophisticated techniques such as cooperative communications and network coding in cognitive networks.

The first part deals with unicast scenarios. We first study the problem of random access over time varying channels with cognitive nodes adjusting their access probabilities according to the decentralized channel state information they acquire at the PHY layer. We derive the conditions for our random access scheme to outperform orthogonal access.

We then study the case where a set of secondary users (SUs) opportunistically accesses the primary user's (PU) spectrum whenever it is idle. Since sensing errors are unavoidable, we study the effect of the interference from the SUs on the stable throughput of the PU. We then compute the range of the SUs' transmission parameters that guarantees the stability of the PU queue. In order to balance the negative effects of the interference from the SUs, we propose a PHY layer relaying protocol

between the PU and SU networks that is based on distributed orthogonal space-time block codes. Under this protocol, it is shown that the PU's throughput gain from relaying increases with the number of SUs. Moreover, the SUs might benefit from relaying the PU's packets as well.

Next, we propose and analyze access schemes at the SUs aiming at exploiting the SU's knowledge of the statistics of various channels and of the average arrival rate to the PU. The motivation is that although the traditional opportunistic spectrum access (OSA) guarantees full protection to the PUs, it is sometimes too conservative if the interference caused by the SUs at the PU receiver is negligible. We derive the conditions under which schemes without sensing outperform schemes with sensing since they offer to the SU more data transmission duration.

The second part of the dissertation deals with cognitive multicasting networks. First, we study relay assisted multicasting. The relay delivers the unsuccessful packets of the source during the idle slots of the source which are determined by sensing. This avoids allocating any explicit resources to the relay. We then substantiate the benefit of using network coding (NC) at the relay.

Finally, we study the problem of reliable spectrum sensing and opportunistic access on channels with stochastic traffic in batch processing systems such as NC. We show how an SU can leverage the structure induced by block-based NC on PUs' channels to mitigate the effects of channel sensing errors and improve the throughput. We consider two different objectives at the SU: quickest detection of an idle slot and throughput maximization. We validate our results with real radio measurements taken in software-defined radio based wireless network tests.

CROSS-LAYER ASPECTS OF COGNITIVE WIRELESS NETWORKS

by

Anthony A. Fanous

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Anthony Ephremides, Chair and Advisor
Professor Prakash Narayan
Professor Richard J. La
Professor Sennur Ulukus
Professor Neil Spring, Dean's Representative

Acknowledgments

First, I would like to thank my advisor Prof. Anthony Ephremides for giving me the opportunity to work under his supervision during my study at the University of Maryland. His deep insights and understanding of the subtle subjects always inspired me. This thesis would not have been possible without his guidance, encouragement and generous support. Also, I would like thank the members of my dissertation committee, Prof. Prakash Narayan, Prof. Richard J. La, Prof. Sennur Ulukus and Prof. Neil Spring, for taking part in my doctoral defense.

I would like to thank Dr. Yalin E. Sagduyu and Dr. Jason H. Li from Intelligent Automation Inc. for giving me the opportunity to work as an intern within the wireless networks and security group. The work in Chapter 6 was partially accomplished during my internship in Summer 2012.

I also owe a word of thanks to all my friends in Maryland, and in particular, Bichoy Hanna, Peter Soliman, Michel Soliman, Amgad Fahmy, Mariana Shaker, Awny Adly and their families. I also would like to thank my friends that were continuously encouraging me, Dr. George Atia, Dr. Ahmed Sultan and Dr. Maged Barsoum.

Finally, I sincerely thank all the members of my family. I owe my deepest gratitude to my parents, Professor Alfy Fanous and Ebtessam Hanna, and my brother Andrew. Their unlimited love and continuous encouragement during my long years of education have always been the support in all my pursuits.

To them, I am eternally grateful.

Table of Contents

List of Figures

Chapter 1: Introduction

One of the main goals of the science of information, known as "information theory", is to characterize the ultimate rates at which information can be reliably communicated over the physical media. Despite its success in characterizing the information theoretic capacities in several scenarios, characterizing the capacity of wireless networks is, to date, an open problem. A main obstacle to achieving that goal is that some of the fundamental assumptions in information theory such as the continuous availability of traffic and the infinite delay to decoding are not applicable to networks. In networks, source burstiness is a central phenomenon that allows resource sharing between users and delay is a fundamental quantity as both a performance measure and a parameter affecting the rate-accuracy tradeoff [1].

In the scarcity of results on information theoretic capacities for networks, communication networks were traditionally designed in a heuristic *layered* approach, where each function is attributed to a particular layer and different layers are separately designed. Such approach with its standardized framework: the *open system interconnection* (OSI) was proven to be very successful in designing robust wired networks. However, with the proliferation of wireless networks, several shortcomings have been observed in the OSI model [2, 3]. In fact, it has been demonstrated that

joint optimization between multiple layers can lead to performance gains in wireless systems. The reason behind this discrepancy is that the wired medium can well be modeled as a time invariant system whose characteristics are well understood and the separate optimization of layers is effective. However, in wireless channels, the inherent random variations affect all the layers and a *cross-layer* approach that accounts for the interaction between the layers is essential in designing modern wireless networks.

The fast paced advances in communication networks, witnessed in the last decades, were mainly due to both the advances in the design of the protocols and algorithms as well as the breakthroughs in the hardware technology that allowed the creation of computationally powerful devices. Such sophisticated devices allowed the creation of *cognitive radios* defined as [4]: *"A cognitive radio is an intelligent wireless communication system that is aware of its surrounding environment (i.e., outside world), and uses the methodology of understanding-by-building to learn from the environment and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit-power, carrier-frequency, and modulation strategy) in real-time, with two primary objectives in mind: highly reliable communications whenever and wherever needed and efficient utilization of the radio spectrum."* This latter objective and the observation that restricting the spectrum access only to licensed users represents a highly inefficient resource utilization since actual measurements indicated that most of these spectrum bands remain idle for a significant fraction of time [5–7] motivated the idea of cognitive networks [4, 8–10]. In cognitive networks, the spectrum is

made available to both licensed (also called primary) users as well as unlicensed (secondary) cognitive users. Secondary users (SUs) by their capability to explore the spectrum, opportunistically access the licensed spectrum in such a way that the interference on the primary users (PUs) is limited or even completely avoided.

Several approaches to cognitive radio network operation have been suggested in the literature [9, 11, 12]. Two main paradigms exist for cognitive access, namely, spectrum sharing (SS) and opportunistic spectrum access (OSA). In spectrum sharing systems, the SUs are allowed to transmit concurrently with the the PUs given some measures to keep the interference caused on primary users within allowable limits, usually within the primary node's noise floor. However, this restriction leads to a very low SUs allowable transmission power and the throughput of the SUs becomes negligible. Opportunistic spectrum access systems aim at avoiding concurrent transmissions between the PUs and the SUs by restricting the SUs to access the channel only at unoccupied temporal, spectral or spatial holes. In order to achieve that goal, the SUs sense the channel at every slot and access it only if no ongoing primary transmissions are detected [13, 14].

1.1 Cross-Layer Design of Cognitive Wireless Networks

In this dissertation, we aim at presenting a *cross-layer* approach in designing cognitive wireless networks. We mainly focus at jointly designing the physical (PHY) and media access control (MAC) layers to achieve optimal designs that cannot be achieved with the traditional layered approach. We study several cross-layer designs

for both unicast and multicast cognitive networks.

1.1.1 Unicast Cognitive Wireless Networks

Random Access is preferred in large wireless networks since its decentralized operation eliminates the need of coordination between the nodes, which largely simplifies the MAC layer protocol design. Even systems with centralized scheduling use random access for initial admission of the users, where the users use random access for the initial access to the network before being allocated dedicated resources (e.g., the random access channel (RACH) in 3G and 4G-LTE systems). Traditional random access systems (ALOHA) was initiated by the work of Abramson [15] where the access probabilities of the users are assumed fixed. By adding cognition to the nodes that randomly access the channel, the transmission probabilities can be adapted to the channel variations. In that direction, we propose a channel-aware random access scheme where the users, independently, adjust their transmission probabilities according to the channel state information that they acquire. We show that such adaptive scheme might outperform orthogonal access schemes.

We then turn the attention to cognitive networks where the SUs opportunistically access the licensed spectrum of the PUs (OSA). Although several sensing techniques have been proposed in the literature (e.g., energy detectors [16] and cyclostationary detectors [14]), sensing errors are unavoidable due to channel impairments such as fading, path loss, and shadowing. This might lead to undesirable detrimental effects to the performance of the PUs. In order to balance the inter-

ference from the SUs, we propose an access scheme at the secondary network that forces the SUs to *cooperate* with the PUs in delivering their traffic. Cooperative communications was motivated by the effectiveness of space diversity in combatting fading, and hence single antenna users can benefit by the virtual MIMO effect induced by other users relaying their transmissions. Cooperative protocols for two sources- two destinations setup have been proposed and analyzed in [17] and distributed space-time codes for multiple relay scenarios have been developed in [18,19]. The performance study was based on information theoretic metrics such as capacity regions, achievable rates and outage probabilities. A network-level cooperative protocol for an uplink where a single pure cognitive relay is introduced to forward unsuccessful packets from source nodes during their idle slots has been proposed and analyzed in [20] with stable throughput and average delay as performance metrics under the assumption of perfect sensing. The assumption of pure relay has been relaxed in [21, 22] where the relay node is a source node having its own traffic but multi-relay case was not considered. We propose and analyze a cooperative protocol between the SUs and the PUs where multiple SUs can forward the PU's unsuccessful packet using distributed orthogonal space-time block codes (D-OSTBCs). The proposed cooperative protocol has the attractive property that with more SUs in the system, the effect of cooperation becomes more prevalent and a higher PU throughput can be achieved. This can be an incentive for the PUs to share their spectrum with the SUs.

Next, in a cognitive wireless network with SUs and PUs, we study enhanced access schemes that exploit the knowledge of the channel statistics at the SUs. In

fact, the OSA scheme is commonly used when the SUs only know about the statistics of the PUs' received signal at the secondary transmitter but no knowledge about the channels to the receivers nor about the PU traffic dynamics. Although the schemes based on sensing (e.g., OSA) provide full protection to the PUs, a fundamental drawback is that sensing detects the PU transmitters while it is required to protect the PU receivers from SU's interference. This might lead to possible waste of transmission opportunities if a PU is sensed to be busy but the channel between the SU transmitter and PU receiver (the cross channel) is in deep fade, and hence SU transmissions will cause negligible interference to the PU. This latter observation motivated schemes that accurately track the cross channel by overhearing the PU's feedback channel leading to a higher SU throughput for the same PU protection [23–26]. In particular, if the SU has exact information about the PU traffic dynamics as well as exact knowledge of the channel states; then for Markovian dynamics, the optimal access policy can be found through a POMDP formulation as in [27]. However, this latter extreme of exactly tracking the cross channel and exact knowledge of the PU traffic dynamics, might require cooperation between the PU receiver and the SU transmitter as well as higher complexity that may not be feasible in practice. Moreover, sensing and channel probing consume SU's resources such as time duration and processing energy leading to a reduction in the resources used for data transmission [28]. Thus, from the SU perspective, it is preferable to avoid sensing if possible. We propose enhanced access schemes that exploit the knowledge of channel statistics at the SUs. These schemes go beyond the traditional OSA and whenever possible, the sensing duration is exploited for data transmission. In the

analysis, we account for the PHY layer parameters such as the sensing duration, the transmission rates and powers in a PHY/MAC design framework.

1.1.2 Multicast Cognitive Wireless Networks

We then turn the attention to the cross-layer design of multicast networks. Unlike the unicast case, in multicast networks the source node has to deliver its traffic to all of its destinations. With ordinary retransmission schemes, the bottleneck is the destination with the worst channel. For the destinations with stronger channels that already successfully received a packet, no throughput benefit is achieved during the retransmissions to the weakest destination. Mitigating this weakness by using *network coding* was one of the key advances in network theory in the last decade.

Network coding has emerged as a powerful scheme to improve the throughput in multicasting networks by coding over the packet traffic [29] and it is well-known that it can improve the throughput in both multi-hop [30] and single-hop [31] multicasting networks. Linear network coding is sufficient to achieve the Min-cut Max-flow capacity [32] for a single multicasting source; and for low complexity network operation, random linear network coding has been introduced in [30].

In the first work in that direction, we aim at combining network coding with cognitive cooperation. We study the case where a cognitive relay assists a multicasting source to deliver its traffic. The cognitive relay delivers the source's unsuccessful packets during the idle slots of the source which are determined by sensing. For further throughput improvements, the relay uses network coding when multicasting

the packets it has in queue.

Spectrum sensing techniques are typically applied in a "memoryless" way without taking into account the spectrum dynamics of the PUs or the channel sensing history. Sensing techniques that exploit the correlation between the PUs' states have been proposed, where the memory is limited to one step (first order Markov chain) as in [33–35], or memory of arbitrary length by introducing the concept of age as in [36].

Our second work merging network coding with cognitive networks is based on the fundamental observation that when applied to PUs' communications, the possible throughput benefits of network coding may lead to more idle slots available to the SUs (i.e., higher spectrum availability). In addition, network coding "shapes" the spectrum and induces a structure to the PUs' states such that the network-coded transmissions occur in batches rather than sporadically (i.e., higher spectrum predictability). Note that the spectrum predictability gain (i.e., "shaping effect") is present even when there is no spectrum availability gain, e.g., when the PU transmitters have perfect channels to multiple receivers or when there is a single receiver with an imperfect channel (where retransmission is still optimal). Figure 1.1 illustrates the spectrum of one PU with and without network coding. With network coding, the busy periods on the PUs' channels are lower-bounded by the coding block size, K, and the idle periods must accumulate a block of K packets to start transmissions, leading to a more predictable spectrum. Such a structure would be observed for any block-based coding at the packet level. These systems include systems using network coding [32], Fountain codes [37] or traffic shaping techniques.

Throughout our work, we use network coding as an example for discussion but the results apply to the other cases as well.

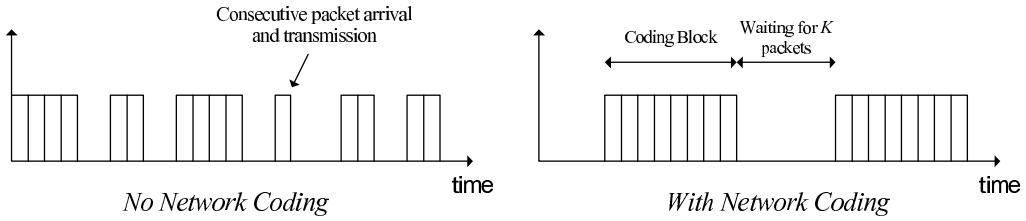

Figure 1.1: Illustration of typical busy/idle periods with and without network coding.

In our work [38], the question that we answer in that part is how an SU, trying to opportunistically access the unintentionally *shaped* PU spectrum due to network coding, can leverage such structure for better inference of the PU state and hence mitigating the negative effects of sensing errors. We will see that the structure induced on a network-coded spectrum can lead to higher SU throughput and more PU protection from misdetections at the SUs.

1.2 Performance Metrics

As mentioned earlier, information theoretic capacity which is commonly used as a PHY-layer throughput metric assumes backlogged nodes and infinite delays in contrast to the burstiness of the traffic as well as the finite packet delay required in network operation. In fact, traffic burstiness is crucial for the operation of some of our protocols and analysis such as the cognitive cooperation to be discussed.

Throughout the dissertation, we use several performance measures that are adequate to capturing the PHY/MAC interactions of the system. The first metric is the *stable throughput*, which is a network-layer metric that is largely affected by

the physical layer parameters. The stable throughput at a queue is the maximum average arrival rate that can be sustained by that queue while remaining finite at all times and emptying infinitely often.[1] The second metric is the *average delay* of a packet which accounts for the queueing and transmission delays.[2] Although queue stability guarantees finiteness of delay, a stable queue can experience very long delays if the average arrival rate is close to the stable throughput of that queue; and hence average delay is a more strict metric for performance.

1.3 Interference Models

A main challenge in devising wireless systems is the existence of fading and undesirable interference from other nodes in the system. In a cognitive environment, interference can also exist between the SUs and the PUs due to sensing errors. Since we are interested in network-level (MAC) performance metrics, the communication unit is considered to be a packet regardless of its bit-content. If a transmission is successful, the entire packet is considered to be decoded without error; otherwise, the packet is not successfully decoded and is discarded. In order to capture the PHY-layer effects on the MAC layer performance, we consider several models for signal reception. In the absence of interference, it is assumed that the receiver is able to successfully decode a transmitted packet if the Signal-to-Noise Ratio (SNR) exceeds some threshold β throughout the packet duration. Hence, the probability of success is given by $\Pr[\text{SNR} > \beta]$. The threshold β depends on the modulation

[1] A precise technical definition of queueing stability is given in Section 2.3.

[2] For all practical purposes, propagation delay is negligible.

scheme, the coding and the target bit-error-rate (BER) set by the receiving node as well as other features of the detector structure. In case of simultaneous interfering transmissions, the capability of reception at the receiver is modeled according to the sophistication and complexity handled at that receiver. For simple receivers with no multipacket reception (MPR) capability, we use the traditional *collision model* where it is assumed that simultaneous transmissions fail with probability one. On the other hand, for receivers which can handle MPR, we use the SINR-threshold model for reception; where the receiver is able to successfully decode a packet if the Signal-to-Interference plus Noise Ratio (SINR) exceeds some threshold throughout the packet duration. Clearly, both the SNR and the SINR models incorporate the effects of the physical layer parameters such as the transmission powers, the channel gains and the additive noise power.

1.4 Outline of the Dissertation

The first part of the dissertation is devoted to unicast cognitive networks. In Chapter 2, we study channel-aware random access and substantiate its usefulness over time varying channels. In Chapter 3, we study the effects of interference induced in a cognitive network, due to sensing errors at the SUs, on the stable throughput of the PU. We then propose a PHY-layer multinode relaying protocol between the SUs and the PU that leads to throughput gain to the PU and hence balancing the effect of interference due to sensing errors. In Chapter 4, we propose and analyze several MAC layer access schemes in cognitive networks and study them from a

PHY/MAC- layer perspective. The target of these access schemes is to maximize the SU throughput subject to guaranteeing some MAC layer performance level at the PU. In particular, we focus on the cases where the PU queue must remain stable and on the more restrictive case of guaranteeing an average delay to the PU.

The second part of the dissertation aims at illuminating the connection between network coding and cognitive radio and how these two emerging technologies can be jointly exploited in wireless networks. In Chapter 5, we study the effect of network coding at a cognitive node relaying the unsuccessful packets of a multicasting source. In Chapter 6, we study the spectrum shaping effects due to network coding and show how they can be leveraged for better inference of the PU's state. Finally, in Chapter 7 we summarize the contributions of the dissertation.

Chapter 2: Channel-Aware Random Access

2.1 Introduction

Random Access schemes were known to be suboptimal to the orthogonal access schemes over the collision channel. However, there has been a continuous interest in studying random access systems due to their simple decentralized operation. A main difficulty in studying random access systems is the inherent interaction between the queues. To bypass that difficulty, the authors in [39] used the idea of stochastic dominance to derive sufficient conditions on the stability of two user slotted ALOHA (S-ALOHA). In [40], the authors used the idea of dominant systems to decouple the interaction between the queues and derive the exact stability region of two user S-ALOHA over the collision channel as well as inner bounds for more than two users ($N > 2$). In [41], the idea of stability ranks was introduced to derive tight bounds on the stability region over collision channel for $N > 2$. Deviating from the oversimplified collision channel, a Multipacket Reception (MPR) model was introduced in [42], where ALOHA with MPR capability under statistically identical infinite users with single buffers was analyzed. Stable throughput region of S-ALOHA with MPR capability in an asymmetric configuration was first derived in [43], where it was shown that for strong MPR capability, the stability region undergoes a phase transition

from a concave region to a convex polyhedron, and in that case, S-ALOHA outperforms orthogonal access. The effect of knowing the channel state information (CSI) on the maximum aggregate stable throughput rate was studied in [44] for an N user statistically symmetric S-ALOHA system. However, the practical case of asymmetric users and channels, and the general case of MPR were not considered. In [45] and [46], the case of asymmetric network was considered under the assumption of collision channel, that is, no MPR capability. Moreover, the delay was considered only under the transmission policy that maximizes the stability region. The policy which minimizes the average delay and its relation to the one maximizing the stability region were not identified. In this part [47], we study channel-aware random access with MPR capability and we identify the policy that minimizes the average delay. We consider an asymmetric two-user S-ALOHA system with i.i.d. two-state time varying links. One state is the good state where transmitted packets are likely to be successfully decoded and the other is the bad state where packets are unsuccessful with high probability. Users have perfect channel knowledge and adjust their transmission probabilities according to the channel state (transmission control). We calculate the stable throughput region with and without MPR capability; as well as the average delay without MPR. The main result is that S-ALOHA with transmission control over a collision channel, from a stability or delay point of view, outperforms orthogonal access whenever the channels tend to be in the bad state. Moreover, in this case, the optimal strategy is to transmit whenever backlogged. By enhancing the physical layer by allowing MPR capability, S-ALOHA with transmission control can outperform orthogonal access even if the channels are not in the bad

state for long proportion of time. This highlights the advantage of using random access with transmission control over time varying channels which is suitable to use over networks that lack strong coordination between the users.

This chapter is organized as follows. In Section 2.2, we introduce the channel model. In Section 2.3, we rigorously define queueing stability and introduce Loynes'theorem to be used throughout the dissertation. In Section 2.4, we calculate the stabile throughput region of a controlled two user S-ALOHA without MPR; while in Section 2.5, we consider the effect of Multipacket reception capability (MPR) on the stability region. In Section 2.6, we consider the minimum average delay per packet without MPR and drive the delay optimal transmission policy, and in Section 2.7 we conclude the chapter.

2.2 System Model

The system consists of an uplink with two source nodes and one destination node as shown in Fig. 2.1. Time is slotted with slot duration equal to one packet transmission duration. User $i \in \{1, 2\}$ receives (or generates) packets according to a stationary process with average rate λ_i, and the arrival processes at the two users are assumed to be independent. Each user has a buffer of infinite capacity to store the packets. The channels are assumed to be independent among links. The channel of a particular link can be in one of two states at any given time slot: the good state that we denote by '1' and the bad state that we denote by '0'. The channel state is assumed to be fixed during a slot duration and varies in

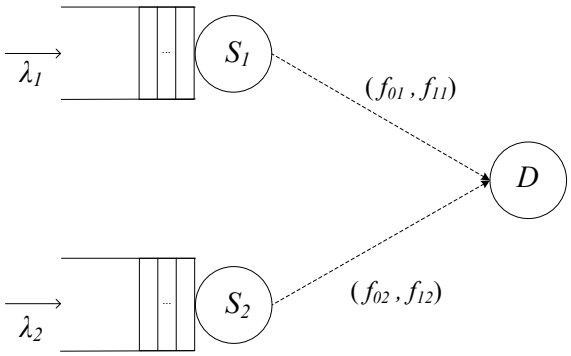

Figure 2.1: System model.

an independent and identically distributed (i.i.d.) manner between slots. The long term proportion of time in which user j's channel is in state i is denoted by $\pi_i^{(j)}$, $i \in \{0, 1\}$, $j \in \{1, 2\}$; and can be obtained either through channel measurements or through a physical model of the channels. We denote by q_{ij} the probability that user j transmits given that his channel is in state i. We denote by f_{ij} the success probability of user j's transmission when his channel is in state i. In this work, we specialize to the case where the channel in 'bad' state of the channel is a deep fade condition, and any transmission through that channel is assumed to fail with probability one, i.e. $f_{01} = f_{02} = 0$. This assumption is in conform with the commonly used SNR threshold model for reception in which a packet is successfully decoded at a destination if and only if the received SNR at that destination exceeds some threshold value. In the bad state, the received SNR is below the threshold and hence the success probability is zero; while in the good state, the received SNR is above the required threshold and hence the reception is successful with probability one. We relax the latter assumption by allowing some positive success probability whenever the channel is in the 'good' state.

2.3 Queue Stability

We adopt the definition of stability used by Szpankowski in [48].

Definition 2.1

A multidimensional stochastic process $\mathbf{Q}^t = (Q_1^t, ..., Q_M^t)$ is *stable* if for every $\mathbf{x}\in \mathbb{N}_0^M$ the following holds

$$\lim_{t\to\infty} \Pr[\mathbf{Q}^t < \mathbf{x}] = F(\mathbf{x}) \text{ and } \lim_{\mathbf{x}\to\infty} F(\mathbf{x}) = 1, \tag{2.1}$$

where $F(\mathbf{x})$ is the limiting distribution function and the limit $\mathbf{x}\to \infty$ is taken componentwise. If $\mathbf{Q}^t$ is an irreducible Markov chain, then stability is equivalent to its ergodicity. Roughly speaking, a queue is stable if its length is finite at all times and it empties infinitely often, while an unstable queue grows in length to infinity. The stable throughput of a queue is the maximum average arrival rate that can be handled at the queue while keeping it stable.

If a weaker condition holds, namely,

$$\lim_{\mathbf{x}\to\infty} \liminf_{t\to\infty} \Pr[\mathbf{Q}^t < \mathbf{x}] = 1, \tag{2.2}$$

then the process is called *substable*.

The ith queue, $i \in \{1, 2\}$ evolves as

$$Q_i^{t+1} = \left(Q_i^t - Y_i^t\right)^+ + X_i^t, \tag{2.3}$$

where Q_i^t is the length of the ith queue at the beginning of time slot t. X_i^t and Y_i^t are the arrival and the service processes at the ith queue in time slot t respectively and $(x)^+$=max$(x, 0)$.

Throughout the dissertation, we use the following lemma [48,49], sometimes referred to as Loynes' theorem.

Lemma 2.1. *For a queue evolving as in* Eq. (2.3)*, if the pair* $\{(X_i^t, Y_i^t)\}$ *is a strictly stationary process (i.e.* $\{X_i^t\}$ *and* $\{Y_i^t\}$ *are jointly stationary), then*

(i) If $\mathbb{E}[X_i^t] < \mathbb{E}[Y_i^t]$*, then the queue is stable in the sense of the definition in* Eq. (2.1)*.*

(ii) If $\mathbb{E}[X_i^t] > \mathbb{E}[Y_i^t]$*, then the queue is unstable and* $\lim_{t\to\infty} Q_i^t = \infty$ *almost surely, where* $\mathbb{E}$ *denotes the expectation operator.*

2.4 Stable Throughput Region (No MPR)

In this section, we consider the case where the destination uses a simple receiver that does not have any MPR capability, thus, simultaneous transmissions result in a collision and both packets are lost. Due to the random access scheme used, the success probability of one user depends on whether the queue of the other user is empty or not. Hence, the two queues are called *interacting*. Although the stability of the queues is equivalent to the ergodicity of the Markov chain according to which they evolve, a closed form solution of the stationary distribution of interacting queues is not easy to obtain, even for the simplest case of random access over a collision channel [40].[1] Hence, the straightforward method of computing the stationary distribution is not applicable. In order to calculate the stable throughput

[1]For systems with MPR capability, stable throughput region is not generally known for more than two interacting queues [43].

region without solving for the stationary distribution, we will use the technique of dominant systems as in [40,41] to decouple the interaction between the two queues. The idea is to come up with hypothetical systems in which the queues do not interact (hence we can exactly solve for their stability region), and yet, the boundary of their stability region coincides with the boundary of the stability region of the original system where the queues interact.

-First Dominant System (S_1)

In S_1, the arrivals to the queues as well as the channel variations are assumed to be identical to those in the original system. However, in S_1, whenever the queue of user 1 empties, he continues transmitting *dummy* packets causing more collisions with the packets of user 2.

The dominant system has the following properties [40]: (i) the queue lengths in the dominant system are no shorter than the queues in the original system; hence if the queues in the dominant system are stable then the queues in the original system are stable as well, (ii) the two systems coincide at saturation, that is, if the queue of user 1 never empties (that is, if it is saturated or unstable), then the dominant system and the original system are indistinguishable; and thus, the instability of the dominant system implies the instability of the original system. Clearly, (i) and (ii) imply that the stability of the dominant system is a necessary and sufficient condition for the stability of the original system and hence, the stable throughput regions of both systems coincide for fixed transmission probabilities. It is clear that in S_1, Q_1 never empties and hence Q_2 sees a constant service rate while Q_1 service rate depends on the state of Q_2 (empty or not). Specifically, Q_2 in the first

dominant system evolves as a discrete M/M/1 queue with stability condition (by Loynes' theorem)

$$\lambda_2 < \mu_2 = \Pr[\text{User 2 is successful in a slot}] = \pi_1^{(2)} q_{12} f_{12} \left(1 - \pi_0^{(1)} q_{01} - \pi_1^{(1)} q_{11}\right). \tag{2.4}$$

Since the service rate of Q_1 depends on the state of Q_2, we have

$$\lambda_1 < \pi_1^{(1)} q_{11} f_{11} \left(1 - \frac{\lambda_2}{\mu_2}\right) + \pi_1^{(1)} q_{11} f_{11} \left(\frac{\lambda_2}{\mu_2}\right) \left(1 - \pi_0^{(2)} q_{02} - \pi_1^{(2)} q_{12}\right), \tag{2.5}$$

where $\left(\frac{\lambda_2}{\mu_2}\right)$ and $\left(1 - \frac{\lambda_2}{\mu_2}\right)$ are the probabilities that Q_2 is busy and idle in a slot, respectively.

Equivalently, the stability conditions in S_1 can be written as

$$\lambda_2 < \pi_1^{(2)} q_{12} f_{12} \left(1 - \pi_0^{(1)} q_{01} - \pi_1^{(1)} q_{11}\right), \tag{2.6}$$

$$\lambda_1 < \pi_1^{(1)} q_{11} f_{11} \left[1 - \left(\frac{\lambda_2}{\mu_2}\right) \left(\pi_0^{(2)} q_{02} + \pi_1^{(2)} q_{12}\right)\right]. \tag{2.7}$$

Similarly for S_2 in which Q_2 transmits dummy packets whenever it empties

$$\lambda_1 < \mu_1 = \pi_1^{(1)} q_{11} f_{11} \left(1 - \pi_0^{(2)} q_{02} - \pi_1^{(2)} q_{12}\right), \tag{2.8}$$

$$\lambda_2 < \pi_1^{(2)} q_{12} f_{12} \left[1 - \left(\frac{\lambda_1}{\mu_1}\right) \left(\pi_0^{(1)} q_{01} + \pi_1^{(1)} q_{11}\right)\right]. \tag{2.9}$$

It can be easily shown that for optimality: $q_{01}^* = q_{02}^* = 0$, since this leads to a strictly higher stability region because transmissions whenever the channel is in the bad state fail with probability one, as expected.

In order to obtain the stable throughput region, we need to find the union over all $(q_{11}, q_{12}) \in [0,1]^2$ of the regions given by the previous equations for fixed q_{11} and

q_{12}. The stability region is thus given by

$$\mathcal{S} = \bigcup_{(q_{11},q_{12})\in[0,1]^2} \mathcal{S}\left(q_{11}, q_{12}\right), \tag{2.10}$$

where $\mathcal{S}\left(q_{11}, q_{12}\right)$ is the stability region for fixed transmission probabilities q_{11} and q_{12} and is given by Eqs. (2.6), (2.7), (2.8) and (2.9). Calculating the boundary of the stability region can be formulated as a constrained optimization problem which can be directly solved by using the same technique as in [50]. Refer to Section 2.8.1 for details.

Theorem 2.1. *If $\pi_0^{(1)} + \pi_0^{(2)} < 1$: The boundary of the stability region is characterized by straight lines near the axes, and by strictly convex function in the middle part. The resulting stability region is given by $\mathcal{R} = \mathcal{L}_1 \bigcup \mathcal{L}_2 \bigcup \mathcal{L}_3$, where*

$$\mathcal{L}_1 = \left\{ (\lambda_1, \lambda_2) : \lambda_2 < \pi_1^{(2)} f_{12} - \left(\frac{\pi_1^{(2)} f_{12}}{\pi_0^{(2)} f_{11}} \right) \lambda_1, \text{for } \lambda_1 \in \left[0, (\pi_0^{(2)})^2 f_{11} \right) \right\},$$

$$\mathcal{L}_2 = \left\{ (\lambda_1, \lambda_2) : \sqrt{\frac{\lambda_1}{f_{11}}} + \sqrt{\frac{\lambda_2}{f_{12}}} < 1, \text{for } \lambda_1 \in \left[(\pi_0^{(2)})^2 f_{11}, (\pi_1^{(1)})^2 f_{11} \right) \right\},$$

$$\mathcal{L}_3 = \left\{ (\lambda_1, \lambda_2) : \lambda_2 < \pi_0^{(1)} f_{12} - \left(\frac{\pi_0^{(1)} f_{12}}{\pi_1^{(1)} f_{11}} \right) \lambda_1, \text{for } \lambda_1 \in \left[(\pi_1^{(1)})^2 f_{11}, \pi_1^{(1)} f_{11} \right) \right\}.$$

If $\pi_0^{(1)} + \pi_0^{(2)} \geq 1$: The stability region is a convex polyhedron whose boundary is determined by two lines. The optimal transmission probabilities are $(q_{11}^*, q_{12}^*) = (1, 1)$. The resulting stability region is convex and given by $\mathcal{R} = \mathcal{L}_1 \bigcup \mathcal{L}_2$, where

$$\mathcal{L}_1 = \left\{ (\lambda_1, \lambda_2) : \lambda_2 < \pi_1^{(2)} f_{12} - \left(\frac{\pi_1^{(2)} f_{12}}{\pi_0^{(2)} f_{11}} \right) \lambda_1, \text{for } \lambda_1 \in \left[0, \pi_1^{(1)} (1 - \pi_1^{(2)}) f_{11} \right) \right\},$$

$$\mathcal{L}_2 = \left\{ (\lambda_1, \lambda_2) : \lambda_2 < \pi_0^{(1)} f_{12} - \left(\frac{\pi_0^{(1)} f_{12}}{\pi_1^{(1)} f_{11}} \right) \lambda_1, \text{for } \lambda_1 \in \left[\pi_1^{(1)} (1 - \pi_1^{(2)}) f_{11}, \pi_1^{(1)} f_{11} \right) \right\}.$$

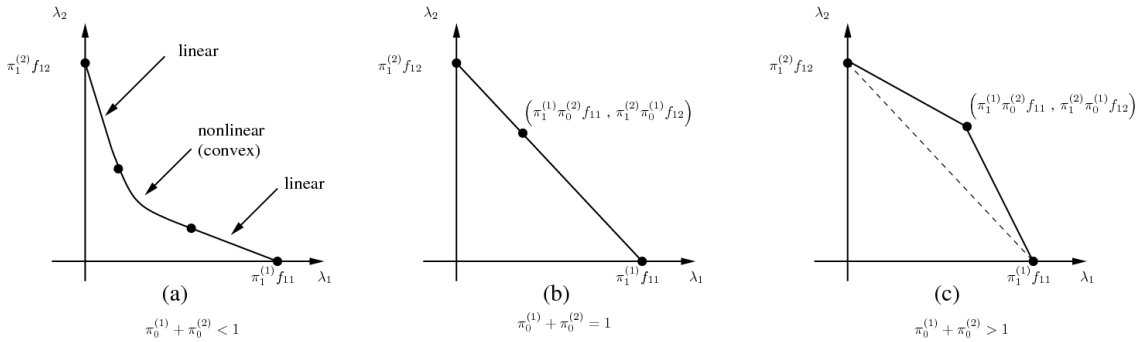

Figure 2.2: Stable throughput regions (no MPR) for various values of stationary probabilities.

Proof: Refer to Section 2.8.1. ■

From Fig. 2.2, we notice that whenever $\pi_0^{(1)} + \pi_0^{(2)} < 1$, which is the case when the channels tend to be in the good state, the stability region is a strict subset of the stability region of the orthogonal access, but is a strict superset of the stability region of ordinary S-ALOHA without transmission control given by $\sqrt{\frac{\lambda_1}{\pi_1^{(1)} f_{11}}} + \sqrt{\frac{\lambda_2}{\pi_1^{(2)} f_{12}}} = 1$. If $\pi_0^{(1)} + \pi_0^{(2)} = 1$, the boundary of the stability region becomes linear and the region coincides with the stability region of the orthogonal access. Finally, if $\pi_0^{(1)} + \pi_0^{(2)} > 1$, the stability region undergoes a phase transition to a convex polyhedron strictly containing the stability region of orthogonal access. This means that, for collision channels, whenever the channels have tendency to be in the bad state, random access with transmission control outperforms orthogonal access while keeping the advantage of simple distributed operation.

2.5 Effect of MPR Capability

So far, we have seen that only if $\pi_0^{(1)} + \pi_0^{(2)} > 1$, does the random access scheme with transmission control outperform orthogonal access. In this section, we show

that, even if $\pi_0^{(1)} + \pi_0^{(2)} > 1$, we can still get such advantage for random access if the receiver has multipacket reception capability.[2] Depending on the strength of the MPR, its effect can be either a strict increase of the stability region without a phase transition (from a concave region to a convex polygon) or an increase of the stability region with phase transition.

Let $\tilde{f}_i$ be the probability of success of the ith user whenever both users transmit simultaneously,[3] which is zero for collision channels. Using the dominant system approach and using that for optimality $q_{01}^* = q_{02}^* = 0$, we obtain the stability conditions in the first dominant system in which Q_1 transmits dummy packets as

$$\lambda_2 < \pi_1^{(2)} q_{12} f_{12} \left[1 - \pi_1^{(1)} q_{11} \left(1 - \frac{\tilde{f}_2}{f_{12}} \right) \right], \tag{2.11}$$

$$\lambda_1 < \pi_1^{(1)} q_{11} f_{11} \left[1 - \frac{\left(1 - \frac{\tilde{f}_1}{f_{11}} \right) \lambda_2}{f_{12} \left[1 - \pi_1^{(1)} q_{11} \left(1 - \frac{\tilde{f}_2}{f_{12}} \right) \right]} \right]. \tag{2.12}$$

Similarly for S_2 in which Q_2 transmits dummy packets:

$$\lambda_1 < \pi_1^{(1)} q_{11} f_{11} \left[1 - \pi_1^{(2)} q_{12} \left(1 - \frac{\tilde{f}_1}{f_{11}} \right) \right], \tag{2.13}$$

$$\lambda_2 < \pi_1^{(2)} q_{12} f_{12} \left[1 - \frac{\left(1 - \frac{\tilde{f}_2}{f_{12}} \right) \lambda_1}{f_{11} \left[1 - \pi_1^{(2)} q_{12} \left(1 - \frac{\tilde{f}_1}{f_{11}} \right) \right]} \right]. \tag{2.14}$$

Following similar steps as in Section 2.8.1, we obtain the stable throughput region with MPR capability as

[2] This is possible for instance in a cellular uplink where the receiver (base station) can handle highly complex receiving algorithms allowing MPR capability.

[3] According to our earlier assumptions, the two users can transmit simultaneously only if their channels are in good state.

Theorem 2.2. *If* $\pi_0^{(1)} + \pi_0^{(2)} + \pi_1^{(2)} \frac{\tilde{f}_1}{f_{11}} + \pi_1^{(1)} \frac{\tilde{f}_2}{f_{12}} < 1$ *The boundary of the stability region is characterized by straight lines near the axes, and by strictly convex function in the middle part. The resulting stability region is given by* $\mathcal{R} = \mathcal{L}_1 \bigcup \mathcal{L}_2 \bigcup \mathcal{L}_3$, *where*

$$\mathcal{L}_1 = \left\{ (\lambda_1, \lambda_2) : \frac{\lambda_2}{\pi_1^{(2)} f_{12}} + \frac{\Psi_2 \lambda_1}{f_{11}\left[1 - \pi_1^{(2)}\Psi_1\right]} < 1, \text{for } \lambda_1 \in \left[0, \frac{f_{11}\left[1 - \pi_1^{(2)}\Psi_1\right]^2}{\Psi_2}\right) \right\},$$

$$\mathcal{L}_2 = \left\{ (\lambda_1, \lambda_2) : \sqrt{\frac{\Psi_2 \lambda_1}{f_{11}}} + \sqrt{\frac{\Psi_1 \lambda_2}{f_{12}}} < 1, \text{for } \lambda_1 \in \left[\frac{f_{11}\left[1 - \pi_1^{(2)}\Psi_1\right]^2}{\Psi_2}, (\pi_1^{(1)})^2 f_{11}\Psi_2\right) \right\},$$

$$\mathcal{L}_3 = \left\{ (\lambda_1, \lambda_2) : \frac{\lambda_1}{\pi_1^{(1)} f_{11}} + \frac{\Psi_1 \lambda_2}{f_{12}\left[1 - \pi_1^{(1)}\Psi_2\right]} < 1, \text{for } \lambda_1 \in \left[(\pi_1^{(1)})^2 f_{11}\Psi_2, \pi_1^{(1)} f_{11}\right) \right\}.$$

and $\Psi_i = 1 - \frac{\tilde{f}_i}{f_{1i}}$ *for* $i \in \{1, 2\}$.

If $\pi_0^{(1)} + \pi_0^{(2)} + \pi_1^{(2)} \frac{\tilde{f}_1}{f_{11}} + \pi_1^{(1)} \frac{\tilde{f}_2}{f_{12}} \geq 1$ *The stability region is a convex polygon whose boundary is determined by two lines. The optimal transmission probabilities are* $(q_{11}^*, q_{12}^*) = (1, 1)$. *The resulting stability region is convex and given by* $\mathcal{R} = \mathcal{L}_1 \bigcup \mathcal{L}_2$, *where*

$$\mathcal{L}_1 = \left\{ (\lambda_1, \lambda_2) : \frac{\lambda_2}{\pi_1^{(2)} f_{12}} + \frac{\Psi_2 \lambda_1}{f_{11}\left[1 - \pi_1^{(2)}\Psi_1\right]} < 1, \text{for } \lambda_1 \in \left[0, \pi_1^{(1)} f_{11}\left[1 - \pi_1^{(2)}\Psi_1\right]\right) \right\},$$

$$\mathcal{L}_2 = \left\{ (\lambda_1, \lambda_2) : \frac{\lambda_1}{\pi_1^{(1)} f_{11}} + \frac{\Psi_1 \lambda_2}{f_{12}\left[1 - \pi_1^{(1)}\Psi_2\right]} < 1, \text{for } \lambda_1 \in \left[\pi_1^{(1)} f_{11}\left[1 - \pi_1^{(2)}\Psi_1\right], \pi_1^{(1)} f_{11}\right) \right\}.$$

We thus conclude that by enhancing the physical layer characteristics of the receiver by allowing MPR capability, random access with transmission control can outperform orthogonal access over time varying channels despite its simple decentralized operation.

2.6 Delay Analysis

In this section, we analyze the delay of a *symmetric* two-user S-ALOHA system with transmission control over i.i.d. time varying channels without MPR capability at the receiver. By *symmetry* we mean that the arrival processes to the two users as well as the channels of both links are statistically identical ($\lambda_1 = \lambda_2 = \lambda$ and $\pi_1^{(2)} = \pi_1^{(1)}$), and hence the users are indistinguishable ($q_{12} = q_{11}$). The need for symmetry is to allow the calculation of the average delay without exactly solving for the queue length distributions which is, to date, an open problem. In [51], authors computed the average delay of two user symmetric S-ALOHA over the collision channel as well as the optimal transmission probability to minimize the delay. In [43], the authors computed the average delay of symmetric S-ALOHA over a class of channels with MPR capability, namely, channels with capture. We follow a similar approach to these works to calculate the average delay of S-ALOHA with transmission control without MPR capability. Our results show that if the channels are more likely to be in the 'bad' states rather than the 'good' states, then the optimal transmission probability is equal to one over all possible arrival rates. Hence in this case, the strategy of transmitting whenever backlogged if the channel is in the 'good' state is both throughput and delay optimal. On the other hand, if the channels tend to be in the 'good' states, then transmission probability equal to one is delay optimal only over a certain range of arrival rates. We make this more specific in the following theorem.

Theorem 2.3. *For symmetric two-user S-ALOHA with transmission control under*

the above assumptions, the average delay is given by

$$D_{avg} = \frac{(1-\lambda) - \left(1-\frac{\lambda}{2}\right)\pi_1^{(1)} q_{11}}{\pi_1^{(1)} q_{11} f_{11}\left(1-\pi_1^{(1)} q_{11}\right) - \lambda}.$$

The optimal transmission probability that minimizes the delay is given by

If $\pi_0^{(1)} \geq 0.5$

$$q_{11}^* = 1, \textit{ for } \lambda \in \left[0, \pi_1^{(1)}\left(1-\pi_1^{(1)}\right) f_{11}\right).$$

If $\pi_0^{(1)} \leq 0.5$

$$q_{11}^* = \begin{cases} 1 & \text{for } \lambda \in \left[0, \tilde{\lambda_1}\right), \\ p_1 & \text{for } \lambda \in \left[\tilde{\lambda_1}, \frac{f_{11}}{4}\right), \end{cases}$$

where

$$\tilde{\lambda_1} = 1 + f_{11}\left(1 - 2\pi_1^{(1)} + 0.5\left(\pi_1^{(1)}\right)^2\right) - \sqrt{1 - \left(\pi_1^{(1)}\right)^2 f_{11} + f_{11}^2\left(1 - 2\pi_1^{(1)} + 0.5\left(\pi_1^{(1)}\right)^2\right)^2},$$

$$p_1 = \frac{1}{(2-\lambda)\pi_1^{(1)}}\left[2(1-\lambda) - \sqrt{4(1-\lambda)^2 - \frac{4}{f_{11}}\left(1-\frac{\lambda}{2}\right)\left[(1-\lambda)f_{11} - \lambda\left(1-\frac{\lambda}{2}\right)\right]}\right].$$

Proof: Refer to Section 2.8.2. ■

In Figs. 2.3 and 2.4, we plot the minimum average delay of a symmetric S-ALOHA with and without transmission control for different values of success probability f_{11} and different values of stationary probability of the channel $\pi_1^{(1)}$. The success probability f_{11} only affects the maximum stable throughput rate that can be handled at the queues. On the other hand, the channel stationary probability $\pi_1^{(1)}$ plays a major role in the relative advantage of transmission control from a delay point of view: transmission control has more significant advantage whenever the channels tend to be in the bad states for a longer proportion of time as can be

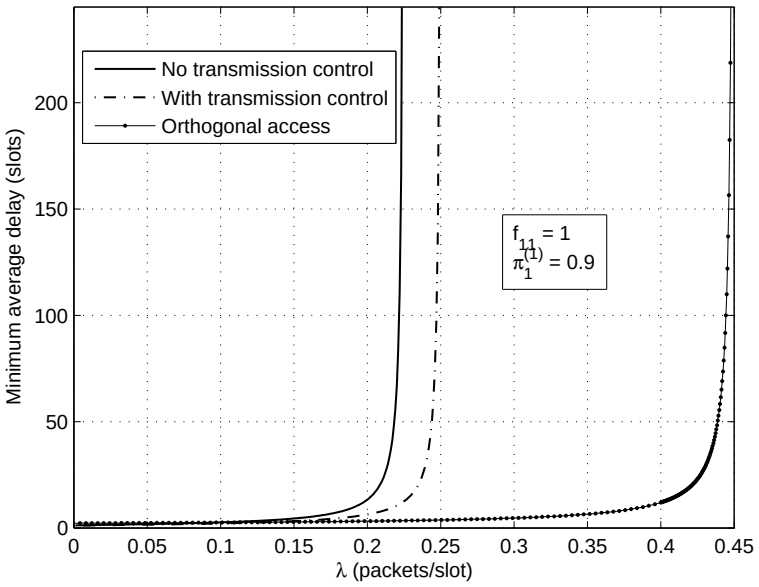

Figure 2.3: Minimum average delay vs. throughput per user ($\pi_1^{(1)} > 0.5$).

inferred from Theorems 2.1 and 2.3. This can be explained by noting that in that case, since the transmissions fail with probability one when the channels are in bad states, it is more advantageous for the users to use transmission control to avoid transmitting in those slots, and thus leaving the channel contention-free to the other user.

2.7 Summary and Conclusions

In this chapter, we characterized the stable throughput region and the average delay of two user random access over i.i.d. time varying channels, where the users exploit their knowledge about their channel states to adjust their channel access probabilities. We showed that random access with transmission control is very effective and can outperform orthogonal access whenever the channels tend to be in

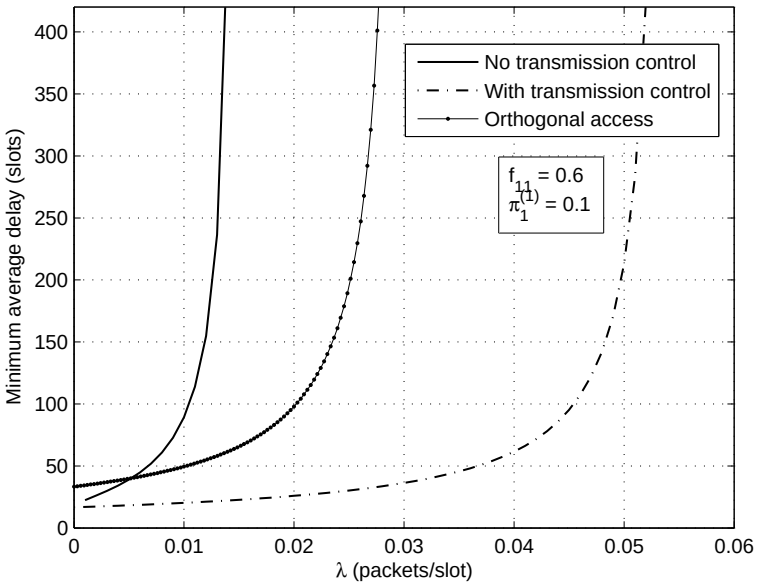

Figure 2.4: Minimum average delay vs. throughput per user ($\pi_1^{(1)} < 0.5$).

the bad states and in that case, the optimal transmission probabilities are one, which eliminates the need of scheduling and simplifies the design of the MAC layer protocol. Moreover, we showed that these transmission probabilities are delay optimal. If the channels tend to be in the good states, transmission control strictly improves the stability region compared to ordinary S-ALOHA but in this case, orthogonal access is better. Furthermore, we showed that enhancing the physical layer by allowing MPR capability can alleviate this downside, attracting the attention that transmission control can make S-ALOHA very suitable to use over time varying channels in networks lacking the capability of coordination between the nodes.

2.8 Appendix

2.8.1 Proof of Theorem 2.1

In Section 2.4, we computed the stability region for a fixed probability pair (q_{11}, q_{12}) by using the dominant system approach. We use the constrained optimization technique as in [50] to derive the boundary of the stability region. After replacing λ_1 by x and λ_2 by y, the boundary of the stability region for fixed transmission probability pair can be written as

$$y = \pi_1^{(2)} q_{12} f_{12} \left[1 - \frac{x}{f_{11}\left(1 - \pi_1^{(2)} q_{12}\right)} \right], \text{ for } 0 \leq x < \pi_1^{(1)} q_{11} f_{11} \left(1 - \pi_1^{(2)} q_{12}\right), \tag{2.15}$$

$$x = \pi_1^{(1)} q_{11} f_{11} \left[1 - \frac{y}{f_{12}\left(1 - \pi_1^{(1)} q_{11}\right)} \right], \text{ for } 0 \leq y < \pi_1^{(2)} q_{12} f_{12} \left(1 - \pi_1^{(1)} q_{11}\right). \tag{2.16}$$

First we consider the constrained optimization problem as given by Eq. (2.15). It can be written as

$$\max_{q_{12} \in [0,1]} y = \pi_1^{(2)} q_{12} f_{12} - \frac{\pi_1^{(2)} q_{12} f_{12} x}{\left(1 - \pi_1^{(2)} q_{12}\right) f_{11}},$$

$$\text{s.t. } 0 \leq x < \pi_1^{(1)} q_{11} f_{11} \left(1 - \pi_1^{(2)} q_{12}\right). \tag{2.17}$$

Differentiating the objective function with respect to q_{12}, we obtain

$$\frac{dy}{dq_{12}} = \pi_1^{(2)} f_{12} - \frac{\pi_1^{(2)} f_{12} x}{\left(1 - \pi_1^{(2)} q_{12}\right)^2 f_{11}}. \tag{2.18}$$

Setting Eq. (2.18) to zero, we obtain

$$q_{12}^{*} = \frac{1}{\pi_1^{(2)}} \left(1 - \sqrt{\frac{x}{f_{11}}}\right). \tag{2.19}$$

For q_{12}^{*} to be a valid probability, we should have

$$\left(1 - \pi_1^{(2)}\right)^2 f_{11} \leq x \leq f_{11}. \tag{2.20}$$

Also, for the constraint in Eq. (2.17) to be satisfied, x must satisfy

$$x \leq (\pi_1^{(1)})^2 f_{11}. \tag{2.21}$$

Combining the two conditions, x must satisfy

$$\left(1 - \pi_1^{(2)}\right)^2 f_{11} \leq x \leq (\pi_1^{(1)})^2 f_{11}. \tag{2.22}$$

Substituting in the objective function in Eq. (2.17), we find that the boundary of the stability region within this range is given by

$$\sqrt{\frac{\lambda_1}{f_{11}}} + \sqrt{\frac{\lambda_2}{f_{12}}} = 1. \tag{2.23}$$

Next, we consider the values of x for which $x < \left(1 - \pi_1^{(2)}\right)^2 f_{11}$. It can be easily shown that

$$\frac{dy}{dq_{12}} > 0, \forall q_{12} \in [0, 1], \tag{2.24}$$

therefore, $q_{12}^{*} = 1$. For the constraint in Eq. (2.17) to be satisfied, $x < \pi_1^{(1)} \pi_0^{(2)} f_{11}$. Hence, by substituting in the objective function of Eq. (2.17), we get that for $0 < x < \min\left(\pi_1^{(1)} \pi_0^{(2)} f_{11}, \left(1 - \pi_1^{(2)}\right)^2 f_{11}\right)$

$$\frac{\lambda_1}{\pi_0^{(2)} f_{11}} + \frac{\lambda_2}{\pi_1^{(2)} f_{12}} = 1. \tag{2.25}$$

Finally, for $x > \left(\pi_1^{(1)}\right)^2 f_{11}$, and noting that $x < \pi_1^{(1)} f_{11}$, in order for the constraint in Eq. (2.17) to be satisfied, we should have that $q_{12} < \frac{1}{\pi_1^{(2)}}\left(1 - \frac{x}{\pi_1^{(1)} f_{11}}\right)$. It can be easily shown that $\frac{dy}{dq_{12}} > 0$ over the range $\left(\pi_1^{(1)}\right)^2 f_{11} < x < \pi_1^{(1)} f_{11}$. Hence, $q_{12}^* = \frac{1}{\pi_1^{(2)}}\left(1 - \frac{x}{\pi_1^{(1)} f_{11}}\right)$. For q_{12}^* to be a valid probability, we should have that $\pi_1^{(1)} f_{11}(1 - \pi_1^{(2)}) < x < \pi_1^{(1)} f_{11}$. Combining both conditions, we get that it is valid for $\max\left(\pi_1^{(1)}(1 - \pi_1^{(2)}) f_{11}, \left(\pi_1^{(1)}\right)^2 f_{11}\right) < x < \pi_1^{(1)} f_{11}$. Substituting in the objective function in Eq. (2.17), we get that for $\max\left(\pi_1^{(1)}(1 - \pi_1^{(2)}) f_{11}, \left(\pi_1^{(1)}\right)^2 f_{11}\right) < x < \pi_1^{(1)} f_{11}$

$$\frac{\lambda_2}{\pi_0^{(1)} f_{12}} + \frac{\lambda_1}{\pi_1^{(1)} f_{11}} = 1. \tag{2.26}$$

By similar arguments, it can be shown that the other dominant system leads to exactly the same stable throughput region, hence the proof is complete.

It should be finally noted that the shape of the stability region depends on whether $\pi_0^{(1)} + \pi_0^{(2)} > 1$ or not. If $\pi_0^{(1)} + \pi_0^{(2)} > 1 \Leftrightarrow \pi_1^{(1)}\pi_0^{(2)} f_{11} < \left(1 - \pi_1^{(2)}\right)^2 f_{11} \Leftrightarrow \pi_1^{(1)}(1 - \pi_1^{(2)}) f_{11} > \left(\pi_1^{(1)}\right)^2 f_{11}$, the stability region consists of two linear parts while if $\pi_0^{(1)} + \pi_0^{(2)} < 1 \Leftrightarrow \pi_1^{(1)}\pi_0^{(2)} f_{11} > \left(1 - \pi_1^{(2)}\right)^2 f_{11} \Leftrightarrow \pi_1^{(1)}(1 - \pi_1^{(2)}) f_{11} < \left(\pi_1^{(1)}\right)^2 f_{11}$, the stability region consists of three parts as in Theorem 2.1.

2.8.2 Proof of Theorem 2.3

The proof follows a similar approach to [51] and [43] in order to solve for the average delay in a symmetric configuration without explicitly solving for the joint queue statistics. The queue of the ith user evolves as

$$Q_i^{t+1} = \left[Q_i^t - D_i^t\right]^+ + A_i^t,\ i \in \{1, 2\}, \tag{2.27}$$

where Q_i^t is the queue length of user i at any time slot t, A_i^t is the number of arrivals to the queue of the ith user in time slot t and D_i^t is the number of departures from the queue of the ith user in time slot t.

In order to compute the delay, we solve for the moment generating function of the joint queue lengths of Q_1 and Q_2 denoted by $G(x,y) = \lim_{t\to\infty} \mathbb{E}\left[x^{Q_1^t} y^{Q_2^t}\right]$. Using the queue evolution equation, we have

$$\mathbb{E}\left[x^{Q_1^t} y^{Q_2^t}\right] = \mathbb{E}\left[x^{A_1^t} y^{A_2^t}\right]\Bigg\{\mathbb{E}\left[\mathbb{1}\left[Q_1^t = 0, Q_2^t = 0\right]\right] + \mathbb{E}\left[x^{Q_1^t - D_1^t}\mathbb{1}\left[Q_1^t > 0, Q_2^t = 0\right]\right]$$
$$+ \mathbb{E}\left[y^{Q_2^t - D_2^t}\mathbb{1}\left[Q_1^t = 0, Q_2^t > 0\right]\right] + \mathbb{E}\left[x^{Q_1^t - D_1^t} y^{Q_2^t - D_2^t}\mathbb{1}\left[Q_1^t > 0, Q_2^t > 0\right]\right]\Bigg\}. \quad (2.28)$$

Taking the limit as $t \to \infty$, we obtain

$$G(x,y) = F(x,y)\Bigg\{G(0,0) + \left[G(x,0) - G(0,0)\right]\Phi_x + \left[G(0,y) - G(0,0)\right]\Phi_y$$
$$+ \left[G(x,y) - G(x,0) - G(0,y) + G(0,0)\right]\Phi_{xy}\Bigg\}, \quad (2.29)$$

where $\Phi_{xy} = \left[\Phi\left(1 - \pi_1^{(1)} q_{11}\right)\left(\frac{1}{x} + \frac{1}{y}\right)\right] + \left[1 - 2\Phi\left(1 - \pi_1^{(1)} q_{11}\right)\right]$, $\Phi_x = \frac{\Phi}{x} + (1 - \Phi)$, $\Phi_y = \frac{\Phi}{y} + (1 - \Phi)$, $\Phi = \pi_1^{(1)} q_{11} f_{11}$ and $F(x,y) = (\lambda x + 1 - \lambda)(\lambda y + 1 - \lambda)$.

It is obvious that $G(1,1) = 1$ and by symmetry $G(0,1) = G(1,0)$. Using L'Hôpital rule

$$G(0,0)\left(\pi_1^{(1)}\right)^2 q_{11}^2 f_{11} + G(1,0)\pi_1^{(1)} q_{11} f_{11}\left(1 - 2\pi_1^{(1)} q_{11}\right) = \left[\pi_1^{(1)} q_{11} f_{11}\left(1 - \pi_1^{(1)} q_{11}\right) - \lambda\right]. \quad (2.30)$$

Calculating $G_1(1,1) = \frac{\partial G(x,y)}{\partial x}$ at $(x,y) = (1,1)$, we get by L'Hôpital rule

$$\left[\pi_1^{(1)} q_{11} f_{11}\left(1 - \pi_1^{(1)} q_{11}\right)\right] G_1(1,1) = \lambda(1 - \lambda) - G_1(1,0)\left(\pi_1^{(1)}\right)^2 q_{11}^2 f_{11}. \quad (2.31)$$

Calculating $\frac{dG(x,x)}{dx}$ at $x = 1$ and using L'Hôpital rule, we get

$$\frac{d}{dx}G(x,x)\bigg|_{x=1} = \frac{G_1(1,0)\,\pi_1^{(1)}q_{11}f_{11}\left(1-2\pi_1^{(1)}q_{11}\right)+\lambda-\frac{3}{2}\lambda^2}{\left[\pi_1^{(1)}q_{11}f_{11}\left(1-\pi_1^{(1)}q_{11}\right)-\lambda\right]}. \tag{2.32}$$

Using that $\frac{dG(x,x)}{dx}\bigg|_{x=1} = 2G_1(1,1)$ and after some manipulations, we get

$$G_1(1,1) = \frac{\lambda(1-\lambda)+\pi_1^{(1)}q_{11}\left(\frac{\lambda^2}{2}-\lambda\right)}{\left[\pi_1^{(1)}q_{11}f_{11}\left(1-\pi_1^{(1)}q_{11}\right)-\lambda\right]}. \tag{2.33}$$

By using Little's law, we get the average delay per packet as

$$D_{avg} = \frac{G_1(1,1)}{\lambda} = \frac{(1-\lambda)+\pi_1^{(1)}q_{11}\left(\frac{\lambda}{2}-1\right)}{\left[\pi_1^{(1)}q_{11}f_{11}\left(1-\pi_1^{(1)}q_{11}\right)-\lambda\right]}. \tag{2.34}$$

We next seek $q_{11} \in [0,1]$ which minimizes D_{avg} while ensuring the stability of the queues. Specifically, we need to solve

$$\begin{aligned}\min_{q_{11}\in[0,1]} D_{avg} &= \frac{(1-\lambda)+\pi_1^{(1)}q_{11}\left(\frac{\lambda}{2}-1\right)}{\left[\pi_1^{(1)}q_{11}f_{11}\left(1-\pi_1^{(1)}q_{11}\right)-\lambda\right]} \\ \text{s.t. } \lambda &< \pi_1^{(1)}q_{11}f_{11}\left(1-\pi_1^{(1)}q_{11}\right).\end{aligned} \tag{2.35}$$

The constraint for stability can be written as

$$\left(\pi_1^{(1)}\right)^2 q_{11}^2 f_{11} - \pi_1^{(1)}f_{11}q_{11} + \lambda < 0. \tag{2.36}$$

The roots of this equation that we denote by s_1 and s_2 are given by

$$s_1, s_2 = \frac{1 \mp \sqrt{1-4\lambda/f_{11}}}{2\pi_1^{(1)}}. \tag{2.37}$$

The stability constraint implies that the optimal probability q_{11}^* satisfies $s_1 \le q_{11}^* \le s_2$. Ignoring for the moment the constraints and equating the derivative of the objective function to zero, we get the optimal transmission probabilities as

$$p_1, p_2 = \frac{(1-\lambda) \mp \sqrt{\lambda/2}\sqrt{\frac{2}{f_{11}}\left(1-\lambda/2\right)^2-(1-\lambda)}}{(1-\lambda/2)\,\pi_1^{(1)}}. \tag{2.38}$$

After some algebraic manipulations, we can show that $0 \leq s_1 \leq 1$ and that $s_1 \leq p_1 \leq s_2 \leq p_2$. Since the objective function is strictly decreasing on (s_1, p_1), we can conclude that the optimal transmission probability q_{11}^* that minimizes the delay is given by $q_{11}^* = \min(p_1, 1)$.

The stable throughput condition yields that $\lambda < \lambda_{\max}$, where

$$\lambda_{\max} = \begin{cases} f_{11}/4, & \text{if } \pi_1^{(1)} \geq 1/2 \Leftrightarrow \pi_0^{(1)} \leq 1/2, \\ \pi_1^{(1)} \left(1 - \pi_1^{(1)}\right) f_{11}, & \text{if } \pi_1^{(1)} \geq 1/2 \Leftrightarrow \pi_0^{(1)} \leq 1/2. \end{cases} \tag{2.39}$$

It can be shown after some manipulations that $p_1 > 1 \Leftrightarrow \lambda \in [0, \lambda^*]$ and that $p_1 < 1 \Leftrightarrow \lambda \in [\lambda^*, \lambda_{\max})$, where

$$\lambda^* = 1 + f_{11}\left(1 - 2\pi_1^{(1)} + \left(\pi_1^{(1)}\right)^2/2\right) - \sqrt{1 - \left(\pi_1^{(1)}\right)^2 f_{11} + f_{11}^2\left(1 - 2\pi_1^{(1)} + \left(\pi_1^{(1)}\right)^2/2\right)^2}. \tag{2.40}$$

Noting that $\lambda^*(\pi_1^{(1)} = 1/2) = f_{11}/4$ and that $\lambda^* < \lambda_{\max} = f_{11}/4$ only if $\pi_0^{(1)} \leq 1/2$, the proof is complete.

Chapter 3: Cross-Layer Cooperation in Cognitive Networks

3.1 Introduction

Cognitive radio has been studied from an information theoretic point of view in [52, 53]. However, such formulation mainly focuses on sophisticated coding techniques at the physical layer and does not take into account the bursty nature of the traffic.[1] In [54], authors studied the stable throughput of a simple cognitive network consisting of one primary and one secondary source-destination pairs under the SINR threshold model for reception with and without relaying for perfect and erroneous sensing. However, such simplified model does not capture the effect of the potential interference induced in a real network with many secondary nodes sharing the spectrum with the primary [55], or the effect of the multiple access protocol used at the secondary network. Moreover, relaying is limited to single node relaying, and the case where the secondary node can be successful when both the primary and the secondary transmit simultaneously was not considered. Similar simple models were considered in [56, 57]. In [58], the authors considered the stable throughput of a more realistic model consisting of a primary TDMA uplink with some dedicated cognitive relays deployed to help the primary, and a secondary network consisting

[1] Refer to Chapter 1 for more details.

of an Ad-Hoc network. However, only the case of perfect sensing of the primary nodes was considered and the analysis is limited to the oversimplified collision model for reception. Relaying is restricted to single node relaying despite the presence of several dedicated relays in the system. In [59], the stable throughput of a network consisting of one primary link and a symmetric secondary cluster with common destination under perfect sensing assumption is considered. The secondary cluster is controlled via a central controller with one secondary node scheduled for transmission at each slot, and communication within the cluster is assumed to be perfect. However, the assumption of having a secondary cluster is not appropriate for Ad-Hoc networks where the presence of a central controller is not generally feasible and the transmissions of the secondary nodes interfere.

In this part, we focus on the effect of the interference in a cognitive network where many secondary nodes share the spectrum with a primary on both the primary's stable throughput and secondary's throughputs. Secondary interference on the primary may occur due to sensing errors or even with perfect sensing in the presence of malicious attacks; while interference between secondary nodes is due to the random access protocol used in the secondary network. We adopt the SINR threshold model for reception which allows for the possibility of successful simultaneous transmissions and captures the effect of the physical layer parameters on the performance. In order to mitigate the effect of the secondary interference on the primary, we propose a multinode relaying protocol that exploits all the SUs that can decode a PU's unsuccessful packet to relay that packet using orthogonal space-time block codes [60]. It is shown that under this protocol, the throughput gain

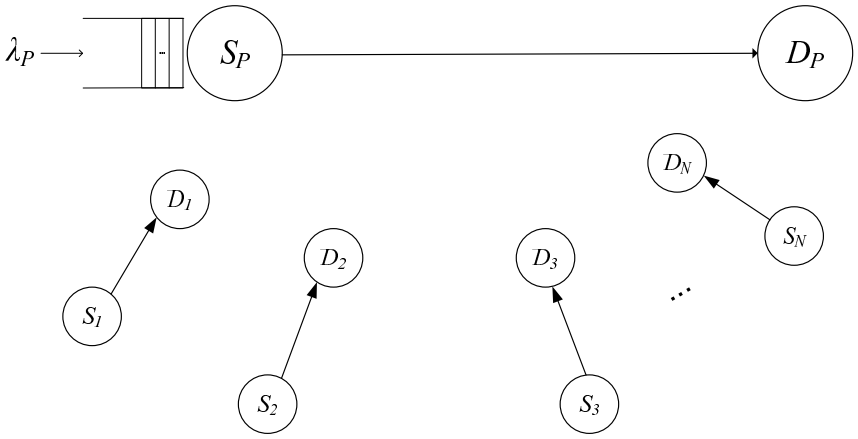

Figure 3.1: System Model

of the primary user from relaying increases with more secondary nodes present in the network [61]. Meanwhile, the secondary nodes might benefit from relaying. The primary node benefits by having more nodes relaying its packets and the secondary might benefit by helping the primary to empty his queue and hence having access to a larger number of idle slots.

This chapter is organized as follows: in Section 3.2, we describe the network and channel models. In Section 3.3, we study the stable throughput of the PU and throughputs of the SUs in the perfect sensing case which will serve as an upper bound on performance, while in Section 3.4, we analyze the effect of erroneous sensing on the throughputs of the PU and of the SUs. In Section 3.5, we propose the relaying protocol to benefit from the large population of secondary nodes. Section 3.6 presents the numerical results and in Section 3.7, we conclude the chapter.

3.2 System Model

The system consists of one primary link and a secondary network consisting of N secondary source-destination pairs forming an interference network as shown in Fig. 3.1. All the nodes use the same frequency band for transmission. This situation arises when that band is licensed to the primary node, while to improve the spectral efficiency, some secondary nodes are allowed to access that band in an opportunistic way. All nodes have buffers of infinite capacity to store their packets to be transmitted. Time is slotted with one packet transmission duration equal to the slot duration. Arrival process to the primary source node is assumed to be stationary with an average rate λ_P packets/slot, while secondary source nodes are assumed to be saturated. Throughout this chapter, we designate the primary node by the subscript P and the ith secondary node by the subscript i with $i \in \mathcal{S} = \{1, 2, ..., N\}$. The ith source node is denoted by S_i and the ith destination node is denoted by D_i, $i \in \{P, 1, 2, ..., N\}$. The ith source node transmits to its destination D_i at power P_i, $i \in \{P, 1, 2, ..., N\}$.

3.2.1 Channel Model

The physical distance between node i and node j is denoted by r_{ij}, where $i, j \in \{S_k, D_k | k = P, 1, 2, ..., N\}$. For instance, $r_{S_P D_j}$ denotes the physical distance between the primary source node and the jth secondary destination node. Path loss exponent is assumed to be equal to α throughout the network. The link between the (i, j) pair of nodes is subject to stationary block fading with fad-

ing coefficients $\{h_{ij}^t\}$ which are independent over slots and mutually independent among links. All receivers are subject to independent additive white complex Gaussian noise with zero mean and variance N_0. Under the adopted SINR-threshold model for reception, node j is able to successfully decode a packet if the received signal-to-interference plus noise ratio (SINR) remains larger than some threshold β_j throughout the packet duration [62]. The threshold β_j depends on the modulation scheme, the coding and the target BER set by the receiving node as well as other features of the detector structure. Upon the success or failure of a packet reception at a node, an Acknowledgment/Non-Acknowledgment (ACK/NACK) packet is fedback to the corresponding transmitter. The ACK/NACK packets are assumed to be instantaneous and error free. This is a reasonable assumption for short length ACK/NACK packets that have negligible delay, and small error rate achieved by using low rate codes on the feedback channel.

Under this model, the transmitted signal by the ith node in the presence of an interfering set of nodes $\mathcal{I}$ at time slot t is received at the jth node as

$$y_j^t = \sqrt{P_i r_{ij}^{-\alpha}} h_{ij}^t x_i^t + \sum_{k \in \mathcal{I}} \sqrt{P_k r_{kj}^{-\alpha}} h_{kj}^t x_k^t + n_j^t, \tag{3.1}$$

where x_k^t is the transmitted packet by the kth node at time slot t and is of unit power and $n_j^t \sim \mathcal{CN}(0, N_0)$ is the additive white complex Gaussian noise at node j. In this case, the success probability of the transmission of the ith node at the jth receiver is given by

$$\Pr\left[\text{SINR}_{ij} > \beta_j\right] = \Pr\left[\frac{P_i r_{ij}^{-\alpha} |h_{ij}^t|^2}{N_0 + \sum_{k \in \mathcal{I}} P_k r_{kj}^{-\alpha} |h_{kj}^t|^2} > \beta_j\right]. \tag{3.2}$$

3.2.2 Multiple Access Protocol

Both the PU and the SUs transmit over the same frequency band. We adopt the OSA scheme and hence, the SUs are restricted to use the idle slots of the PU. The primary node has the priority for transmission. At the beginning of each slot, the SUs sense the channel and only if a slot is detected to be idle, do they access the channel in a random access way. The ith SU will transmit in a slot with a probability q_i whenever that slot is detected to be idle. We assume that there is sufficient guard time at the beginning of each slot to allow sensing the channel at the SUs.

According to the cognitive radio principle, the SUs should be "transparent" to the primary in the sense that their transmissions should not affect some performance criterion (here, the queueing stability) of the primary node. If the sensing is perfect, the SUs never interfere with the PU and can use any values for their transmission parameters (power/channel access probability) that maximize their sum throughput without affecting the stability of the PU. However, if the sensing is not perfect, the SUs must limit their interference on the PU by controlling their transmission parameters to achieve that goal while maximizing their opportunistic throughput. We discuss the constraints on the secondary transmission parameters in case of imperfect sensing in Section IV.

Throughout this chapter, we consider both perfect and imperfect sensing. Perfect sensing is an optimistic case and only serves as an upper bound on performance. We also consider both cases of asymmetric network with arbitrary fading distribu-

tions and symmetric network with Rayleigh fading where $h_{ij} \sim \mathcal{CN}(0, \sigma_{ij}^2)$. We will find that most results apply for both cases but in the asymmetric case, some results are not in closed form.

The symmetric case with Rayleigh fading that we will consider is characterized as follows

- $P_j = P_0$, for $j \in \mathcal{S} = \{1, 2, ..., N\}$.
- $q_j = q$, for $j \in \mathcal{S} = \{1, 2, ..., N\}$.
- $\beta_j = \beta$, for $j \in \mathcal{S} = \{1, 2, ..., N\}$.
- $r_{S_k D_j} = r_j$, for $j, k \in \mathcal{S} = \{1, 2, ..., N\}$.
- $r_{S_j D_P} = r_0$, for $j \in \mathcal{S} = \{1, 2, ..., N\}$.
- $r_{S_P S_k} = r$, for $k \in \mathcal{S} = \{1, 2, ..., N\}$.
- $h_{S_j D_k} = \tilde{h_j} \sim \mathcal{CN}(0, \tilde{\sigma}^2)$, for $j, k \in \mathcal{S} = \{1, 2, ..., N\}$.
- $h_{S_j D_P} = \bar{h_j} \sim \mathcal{CN}(0, \sigma_0^2)$, for $j \in \mathcal{S} = \{1, 2, ..., N\}$.
- $h_{S_P S_k} = h_k \sim \mathcal{CN}(0, \sigma^2)$, for $k \in \mathcal{S} = \{1, 2, ..., N\}$.

This geometry, for instance, arises whenever the secondary sources lie on a circle and secondary destinations, along with primary source-destination pair lie on a line passing by the center of that circle and perpendicular to its plane. This is an approximation for the case where the secondary network is an uplink and the primary network is an uplink or downlink, and the secondary nodes are located far away from their base station while the primary base station is close to its receiver. We adopt the definition of queueing stability as in Section 2.3.

3.3 Perfect Sensing Case

In this case, the SUs are able to perfectly identify the PU's idle slots where they can access the channel, from the PU's busy slots where they must remain silent to avoid interfering with the PU. In this case, the PU gets his maximum possible service rate. Clearly, this is an ideal situation serving as an upper bound on the performance.

3.3.1 Primary User's Queue

Theorem 3.1. *The stability condition for the PU's queue in the perfect sensing case is given by*

$$\lambda_P < \mu_P^{\max} = \Pr\left[\frac{P_P|h_{S_PD_P}|^2 r_{S_PD_P}^{-\alpha}}{N_0} > \beta_P\right]. \tag{3.3}$$

For the case of Rayleigh fading, we have that $\mu_P^{\max} = \exp\left(\frac{-N_0\beta_P}{P_P\sigma^2_{S_PD_P} r^{-\alpha}_{S_PD_P}}\right)$.

Proof: The service process at the PU's queue is given by $Y_P^t = \mathbb{1}\left\{\overline{\mathcal{O}^t_{S_PD_P}}\right\}$, where $\overline{\mathcal{O}^t_{S_PD_P}}$ denotes the event of no outage on the PU source-destination link in slot t and $\mathbb{1}\{\cdot\}$ is the indicator function which takes the value of one if its argument is true and zero otherwise. This event depends on the fading process on the (S_P, D_P) link which is stationary and hence, the process $\{Y_P^t\}$ is stationary. The average service rate of the PU's queue in this case, that we denote by $\mu_P^{\max}$, is given by

$$\mu_P^{\max} = \mathbb{E}\left[Y_P^t\right] = \Pr\left[\overline{\mathcal{O}^t_{S_PD_P}}\right] = \Pr\left[\frac{P_P|h_{S_PD_P}|^2 r_{S_PD_P}^{-\alpha}}{N_0} > \beta_P\right]. \tag{3.4}$$

For the Rayleigh fading case, $|h_{S_PD_P}|^2$ is exponentially distributed with mean $\sigma^2_{S_PD_P}$, and the results follows.

Finally, by using Loynes' theorem, we can get the stability condition of the primary node as given in Eq. (3.3). ■

3.3.2 Secondary Users' Queues

Source node $i \in \mathcal{S} = \{1, 2, ..., N\}$ of the secondary network transmits with probability q_i, independently of the other secondary nodes, whenever a slot is detected to be idle.

Theorem 3.2. *In the perfect sensing case, the throughput region of the SUs is*

$$\mathcal{R} = \bigcup_{\boldsymbol{q}\in[0,1]^N} \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^N : \lambda_j = \left(1 - \frac{\lambda_P}{\mu_P^{\max}}\right) \sum_{\substack{\mathcal{T}\subseteq\{1,2,...,N\} \\ j\in\mathcal{T}}} \prod_{k\in\mathcal{T}} q_k \prod_{l\in\mathcal{S}\setminus\mathcal{T}} (1-q_l) P_{S_j}^{\mathcal{T}},\ j \in \mathcal{S} \right\}, \tag{3.5}$$

where

$$P_{S_j}^{\mathcal{T}} = \Pr\left[\frac{P_j |h_{S_j D_j}|^2 r_{S_j D_j}^{-\alpha}}{N_0 + \sum_{k\in\mathcal{T}, k\neq j} P_k |h_{S_k D_j}|^2 r_{S_k D_j}^{-\alpha}} > \beta_j \right],$$

which, for the Rayleigh fading case, is equal to

$$P_{S_j}^{\mathcal{T}} = \exp\left(\frac{-N_0 \beta_j}{\sigma_{S_j D_j}^2 P_j r_{S_j D_j}^{-\alpha}} \right) \sum_{\substack{k\in\mathcal{T} \\ k\neq j}} \left(\prod_{l\neq k} \frac{\theta_l \theta_k}{\theta_l - \theta_k} \right) \frac{1}{\left(\theta_k + 1/\sigma_{S_j D_j}^2\right)}.$$

Proof: Let $\mathcal{A}_{\mathcal{T}^t}^t$ be the event that only nodes in subset $\mathcal{T}^t \subseteq \mathcal{S}$ of secondary nodes transmit in slot t and let $\overline{\mathcal{O}_{S_j D_j, \mathcal{T}^t}^t}$ be the event of no outage on the (S_j, D_j) link in slot t when all nodes in the set $\mathcal{T}^t$ transmit.

The departure process of the jth secondary node can be written as

$$Y_j^t = \sum_{\mathcal{T}^t\subseteq\{1,2,...,N\}, j\in\mathcal{T}^t} \mathbb{1}\left\{ \{Q_P^t = 0\} \bigcap \mathcal{A}_{\mathcal{T}_t}^t \bigcap \overline{\mathcal{O}_{S_j D_j, \mathcal{T}^t}^t} \right\}. \tag{3.6}$$

By using the fact that if the primary queue is stable, then the process $\mathbb{1}\{Q_P^t = 0\}$ is stationary [41,48]; it can be easily shown that the process $\{Y_j^t\}$ is stationary. Hence, we drop the time indices. By Little's law [63], it follows that $\Pr[Q_P = 0] = 1 - \frac{\lambda_P}{\mu_P^{\max}}$. Given a set $\mathcal{T} \subseteq \{1, 2, ..., N\}$ of secondary nodes transmitting in a slot, the probability that the jth secondary destination node $j \in \mathcal{T}$ is able to successfully decode the jth secondary source node transmission is given by

$$P_{S_j}^{\mathcal{T}} = \Pr\left[\overline{\mathcal{O}_{S_j D_j, \mathcal{T}}}\right] = \Pr\left[\frac{P_j |h_{S_j D_j}|^2 r_{S_j D_j}^{-\alpha}}{N_0 + \sum_{k \in \mathcal{T}, k \neq j} P_k |h_{S_k D_j}|^2 r_{S_k D_j}^{-\alpha}} > \beta_j\right]. \tag{3.7}$$

For the case of Rayleigh fading (Refer to Section 3.8.1 for the proof), $P_{S_j}^{\mathcal{T}}$ is

$$P_{S_j}^{\mathcal{T}} = \exp\left(\frac{-N_0 \beta_j}{\sigma_{S_j D_j}^2 P_j r_{S_j D_j}^{-\alpha}}\right) \sum_{\substack{k \in \mathcal{T} \\ k \neq j}} \left(\prod_{l \neq k} \frac{\theta_l \theta_k}{\theta_l - \theta_k}\right) \frac{1}{\left(\theta_k + 1/\sigma_{S_j D_j}^2\right)}, \tag{3.8}$$

where $\theta_k = \frac{P_j r_{S_j D_j}^{-\alpha}}{P_k r_{S_k D_j}^{-\alpha} \beta_j \sigma_{S_k D_j}^2}$.

By independence of the events in Eq. (3.6), the throughput rate of the jth secondary source node is given by

$$\lambda_j = \mathbb{E}\left[Y_j^t\right] = \left(1 - \frac{\lambda_P}{\mu_P^{\max}}\right) \sum_{\substack{\mathcal{T} \subseteq \{1,2,...,N\} \\ j \in \mathcal{T}}} \prod_{k \in \mathcal{T}} q_k \prod_{l \in \mathcal{S} \setminus \mathcal{T}} (1 - q_l) P_{S_j}^{\mathcal{T}}. \tag{3.9}$$

Finally, the throughput region of the secondary network is obtained by taking the union over all possible transmission probability vectors $\mathbf{q} = (q_1, q_2, ..., q_N) \in [0, 1]^N$ as in Eq. (3.5). ■

Next, we consider the symmetric case introduced in Section 3.2. In that case, the probability of success of the jth SU in the presence of k other interfering transmissions is given by

$$P^{(k)} = \Pr\left[\frac{P_0 |h_{S_j D_j}|^2 r_j^{-\alpha}}{N_0 + \sum_{m=1}^{k} P_0 |\tilde{h_m}|^2 r_j^{-\alpha}} > \beta\right] = \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 r_j^{-\alpha} P_0}\right) \frac{1}{(1 + \beta)^k}. \tag{3.10}$$

Proof: Refer to Section 3.8.1. ■

In this case, the throughput rate of the jth secondary node is given by

$$\begin{aligned}\lambda_j &= \left(1-\frac{\lambda_P}{\mu_P^{\max}}\right) q \sum_{k=0}^{N-1} \binom{N-1}{k} q^k (1-q)^{N-1-k} P^{(k)} \\ &= \left(1-\frac{\lambda_P}{\mu_P^{\max}}\right) \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 r_j^{-\alpha} P_0}\right) q \sum_{k=0}^{N-1} \binom{N-1}{k} q^k (1-q)^{N-1-k} \frac{1}{(1+\beta)^k} \\ &= \left(1-\frac{\lambda_P}{\mu_P^{\max}}\right) \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 r_j^{-\alpha} P_0}\right) q \left[1 - q\frac{\beta}{1+\beta}\right]^{N-1}. \end{aligned} \tag{3.11}$$

We note that due to perfect sensing, the SUs do not interfere with the PU. Hence, they can transmit at their maximum power in order to maximize their throughput rate $\left(\frac{\partial \lambda_j}{\partial P_0} > 0\right)$ without affecting the stability of the primary queue. This is not necessary true in case of imperfect sensing as will be seen in Section 3.4.

Next, we calculate the optimum transmission probability q^* at which the SUs should transmit in order to maximize their throughput. A very small q limits the interference between the SUs but at the same time reduces the throughput, while a large value of q causes high interference between the SUs; leading to a degradation in the throughput. By setting $\frac{\partial \lambda_j}{\partial q} = 0$, we get $q^* = \min\{1, \frac{1}{\chi N}\}$, where $\chi = \frac{\beta}{1+\beta}$. Thus, for a small number of secondary nodes N, it is beneficial to transmit with probability one, while for a large value of N, the SUs should backoff to limit the interference on each other. Moreover, it is beneficial for both primary and secondary nodes that the PU transmits at its maximum power to achieve its maximum service rate and hence maximize the fraction $\left(1-\frac{\lambda_P}{\mu_P^{\max}}\right)$ of idle slots available to the secondary transmissions. This may not be true if the sensing is not perfect because of the interference between the PU and the SUs that may suffer from degradation of their throughput

if the PU increases its transmission power.

3.4 Imperfect Sensing Case

Due to fading and other channel impairments, secondary nodes can encounter errors while sensing the channel, leading to interference with the primary node that might cause drastic reduction of its stable throughput. In this section, we quantify the effect of imperfect sensing on the throughput of primary and secondary nodes.

3.4.1 Channel Sensing

Two errors may occur at the secondary nodes while sensing the channel, namely, false alarm and misdetection errors. All subsequent results are applicable for any sensing method since they are given in terms of general false alarm $P_f^{(i)}$ and misdetection $P_e^{(i)}$ probabilities at the ith SU. It should be noted that for a particular detector, $P_f^{(i)}$ and $P_e^{(i)}$ are related by its receiver operating characteristics (ROC) [64]. False alarm occurs whenever the primary node is idle but is sensed to be busy. Clearly, false alarm errors do not affect the stable throughput of the PU but degrades the throughput of the SUs. Misdetection occurs when the PU is busy but is sensed by some SUs to be idle. Those SUs will simultaneously transmit with the PU causing some interference at the primary destination. If the interference is strong enough, it may lead to instability of the primary queue. Note that by the independence of the fading processes between nodes, the misdetection and false alarm events are independent between secondary nodes.

3.4.2 Primary Queue Analysis

Theorem 3.3. *In the imperfect sensing case, the stability condition of the primary queue is given by*

$$\lambda_P < \mu_P = \sum_{\mathcal{U}\subseteq\{1,2,\ldots,N\}} P_e^{\mathcal{U}} \left[\sum_{\mathcal{T}\subseteq\mathcal{U}} \left[\prod_{k\in\mathcal{T}} q_k \prod_{k\in\mathcal{U}\setminus\mathcal{T}} (1-q_k) \right] \mu_P^{(\mathcal{T})} \right], \tag{3.12}$$

where $P_e^{\mathcal{U}}$ *and* $\mu_P^{(\mathcal{T})}$ *are given by* Eqs. (3.14) *and* (3.15).

Proof: Let $\mathcal{U}^t \subseteq \mathcal{S} = \{1, 2, ..., N\}$ be the set (possibly empty) of SUs that have misdetections at time slot t. A subset $\mathcal{T}^t \subseteq \mathcal{U}^t$ of these nodes chooses to transmit at that time slot. The service process of the PU can be expressed as

$$Y_P^t = \sum_{\mathcal{U}^t\subseteq\{1,2,\ldots,N\}} \sum_{\mathcal{T}^t\subseteq\mathcal{U}^t} \mathbb{1}\left\{ \mathcal{E}_{\mathcal{U}^t}^t \bigcap \mathcal{A}_{\mathcal{T}^t}^t \bigcap \overline{\mathcal{O}_{S_PD_P,\mathcal{T}^t}^t} \right\}, \tag{3.13}$$

where $\mathcal{E}_{\mathcal{U}^t}^t$ denotes the event that only nodes in the set $\mathcal{U}^t$ misdetect the primary node in time slot t; $\mathcal{A}_{\mathcal{T}^t}^t$ is the event that only nodes in the set $\mathcal{T}^t$ transmit at time slot t and $\overline{\mathcal{O}_{S_PD_P,\mathcal{T}^t}^t}$ is the event of no outage on the (S_P, D_P) link in the presence of an interfering set $\mathcal{T}^t$ of secondary nodes. The process $\{Y_P^t\}$ is clearly stationary, thus we drop the time indices t subsequently.

By independence, the probability that only nodes in the set $\mathcal{U}$ have misdetection is

$$P_e^{\mathcal{U}} = \Pr\left[\mathcal{E}_{\mathcal{U}}\right] = \prod_{j\in\mathcal{U}} P_e^{(j)} \prod_{j\in\mathcal{S}\setminus\mathcal{U}} \left(1 - P_e^{(j)}\right). \tag{3.14}$$

The service rate at the primary node given a set $\mathcal{T}$ of transmitting nodes, as defined above, is given by

$$\mu_P^{(\mathcal{T})} = \Pr\left[\overline{\mathcal{O}_{S_PD_P,\mathcal{T}}}\right] = \Pr\left[\frac{P_P |h_{S_PD_P}|^2 r_{S_PD_P}^{-\alpha}}{N_0 + \sum_{k\in\mathcal{T}} P_{S_k} |h_{S_kD_P}|^2 r_{S_kD_P}^{-\alpha}} > \beta_P \right]. \tag{3.15}$$

We then obtain the average service rate at the primary queue as given in Eq. (3.15), and hence by Loynes' theorem, the proof is complete. ■

Note that from Eq. (3.15), $\mu_P^{(\mathcal{T})} < \mu_P^{\max}$ unless $\mathcal{T}$ is the empty set. Since $\sum_{\mathcal{U}\subseteq\{1,2,\ldots,N\}} P_e^{\mathcal{U}} \left[\sum_{\mathcal{T}\subseteq\mathcal{U}} \left[\prod_{k\in\mathcal{T}} q_k \prod_{k\in\mathcal{U}\setminus\mathcal{T}}(1-q_k)\right]\right] = 1$, it is clear that Eq. (3.12) is a convex combination of terms less than or equal to $\mu_P^{\max}$, hence μ_P given in Eq. (3.12) is strictly less than $\mu_P^{\max}$, which is an expected result due to secondary interference.

Next, we specialize to the symmetric case with Rayleigh fading. In that case, by symmetry, the probability of misdetection is the same for all the secondary nodes, i.e., $P_e^{(j)} = P_e,\ j \in \mathcal{S} = \{1, 2, ..., N\}$.

Let $\mu_P^{(k)}$ be the success probability of the PU given k secondary concurrent transmissions, then by similar analysis as in Section 3.8.1, we get $\mu_P^{(k)}$ as

$$\begin{aligned}\mu_P^{(k)} &= \Pr\left[\frac{P_P|h_{S_pD_P}|^2 r_{S_PD_P}^{-\alpha}}{N_0 + \sum_{j=1}^{k} P_0|h_j|^2 r_0^{-\alpha}} > \beta_P\right] \\ &= \exp\left(\frac{-N_0\beta_P}{\sigma_{S_pD_P}^2 P_P r_{S_PD_P}^{-\alpha}}\right) \frac{1}{\left(1 + \frac{P_0 r_0^{-\alpha}\beta_P\sigma_0^2}{\sigma_{S_PD_P}^2 P_P r_{S_pD_P}^{-\alpha}}\right)^k} = \mu_P^{\max}\left(\frac{a}{1+a}\right)^k, \end{aligned} \tag{3.16}$$

where

$$a = \frac{\sigma_{S_PD_P}^2 P_P r_{S_PD_P}^{-\alpha}}{\sigma_0^2\beta_P P_0 r_0^{-\alpha}}. \tag{3.17}$$

By symmetry, the average service rate of the primary queue is given by

$$\begin{aligned}\mu_P &= \sum_{L=0}^{N}\binom{N}{L} P_e^L (1-P_e)^{N-L}\left[\sum_{k=0}^{L}\binom{L}{k} q^k (1-q)^{L-k}\mu_P^{(k)}\right] = \mu_P^{\max}\left[1 - \frac{qP_e}{a+1}\right]^N \\ &= \mu_P^{\max}\left[1 - qP_e\frac{P_0 r_0^{-\alpha}\beta_P\sigma_0^2}{\sigma_{S_PD_P}^2 P_P r_{S_PD_P}^{-\alpha} + \sigma_0^2 P_0 r_0^{-\alpha}\beta_P}\right]^N. \end{aligned} \tag{3.18}$$

The effect of imperfect sensing is shown in the multiplication of $\mu_P^{\max}$ by a term less

than one. By Loynes' theorem, for stability, we should have

$$\lambda_P < \mu_P^{\max} \left[1 - qP_e \frac{P_0 r_0^{-\alpha} \beta_P \sigma_0^2}{\sigma_{S_P D_P}^2 P_P r_{S_P D_P}^{-\alpha} + \sigma_0^2 P_0 r_0^{-\alpha} \beta_P} \right]^N . \tag{3.19}$$

The primary user chooses its arrival rate $\lambda_P < \mu_P^{\max}$ independently of the secondary network. In the imperfect sensing case, $\mu_P < \mu_P^{\max}$, and hence, the SUs should limit their transmission power and/or transmission probabilities to limit the interference on the PU to ensure that its arrival rate λ_P be less than μ_P to avoid the instability of its queue.

It is straightforward to establish the following properties about the PU service rate μ_P given by Eq. (3.18).

Proposition 3.1. *The primary node service rate in the imperfect sensing case, as given by* Eq. (3.18) *satisfies*

(i) $0 \leq \mu_P \leq \mu_P^{\max}$.

(ii) $\lim_{a \to \infty} \mu_P = \mu_P^{\max}$.

(iii) $\lim_{a \to 0} \mu_P = \begin{cases} \mu_P^{\max} \left[1 - qP_e \right]^N, \textit{if } P_P > 0 \textit{ and } P_0 \to \infty \\ 0, \textit{ if } P_P \to 0. \end{cases}$

(iv) $\dfrac{\partial \mu_P}{\partial a} > 0$, *i.e., μ_P is strictly increasing with a.*

(v) $\lim_{q \to 0} \mu_P = \mu_P^{\max}$.

(vi) $\lim_{q \to 1} \mu_P = \mu_P^{\max} \left[1 - \dfrac{P_e}{a+1} \right]^N$.

(vii) $\dfrac{\partial \mu_P}{\partial q} < 0$, *i.e., μ_P is strictly decreasing with q.*

From Proposition 3.1, we can draw the following conclusions: property (i)

states that the effect of sensing errors at the SUs is the degradation of the service rate of the PU licensed node due to the interference from the SUs on the primary. Properties (ii),(iii) and (iv) reveal that unless $a \to \infty$ (i.e., either $P_P \to \infty$ or $P_0 \to 0$), the primary node cannot achieve its maximum service rate $\mu_P^{\max}$ that is achieved in the case of perfect sensing. Also, for finite P_P, which is the case of interest here, the SUs have a maximum power, possibly infinite if $\lambda_P < \mu_P^{\max}(1-qP_e)^N$, at which they can transmit without affecting the stability of the primary node. Moreover, even if the interference of the SUs is very high (case of $P_0 \to \infty$), the PU can still achieve a portion $(1-qP_e)^N$ of its maximum service rate $\mu_P^{\max}$. Finally, properties (v), (vi) and (vii) show that for fixed P_P and P_0, the secondary nodes can control their interference level on the primary user by adjusting their transmission probabilities q. This might be easier to implement than power control due to hardware complexity and non-linearity of the power amplifiers used for power control over wide range.

For $\lambda_P < \mu_P$ to be satisfied, we can solve for the minimum value of $a = \frac{\sigma^2_{S_PD_P} P_P r^{-\alpha}_{S_PD_P}}{\sigma_0^2 \beta_P P_0 r_0^{-\alpha}}$ and for the maximum value of q to calculate the maximum possible transmission power ($P_0^{\max}$) and the maximum possible transmission probability ($q_{\max}$) of the SUs while remaining "transparent" to the PU, i.e., without affecting its stability.

By using Eq. (3.18) and Proposition 3.1, we obtain

$$q < q_{\max} = \begin{cases} 1, & \text{if } \lambda_P < \mu_P^{\max}\left[1-\frac{P_e}{a+1}\right]^N \\ \left[1-\left(\frac{\lambda_P}{\mu_P^{\max}}\right)^{1/N}\right]\frac{a+1}{P_e}, & \text{if } \mu_P^{\max}\left[1-\frac{P_e}{a+1}\right]^N < \lambda_P < \mu_P^{\max}. \end{cases} \tag{3.20}$$

For fixed primary transmission power P_P, we can calculate the maximum transmis-

sion power allowed at the secondary nodes as

$$P_0 < P_0^{\max} = \begin{cases} \infty, & \text{if } \lambda_P < \mu_P^{\max}\left(1-qP_e\right)^N \\ \frac{\sigma_{S_PD_P}^2 P_P r_{S_PD_P}^{-\alpha}}{r_0^{-\alpha}\beta_P\sigma_0^2}\left[\frac{1-(\lambda_P/\mu_P^{\max})^{1/N}}{qP_e-1+(\lambda_P/\mu_P^{\max})^{1/N}}\right], & \text{if } \mu_P^{\max}\left(1-qP_e\right)^N < \lambda_P < \mu_P^{\max}. \end{cases} \tag{3.21}$$

From Eqs. (3.20) and (3.21), we conclude that for fixed P_P, if $\lambda_P < \mu_P^{\max}\left[1-\frac{P_e}{a+1}\right]^N$, the SUs can transmit at any desired chosen probability without affecting the stability of the primary, while they have to backoff to reduce their interference on the primary node if $\mu_P^{\max}\left[1-\frac{P_e}{a+1}\right]^N < \lambda_P < \mu_P^{\max}$. On the other hand, for fixed transmission probability q, if $\lambda_P < \mu_P^{\max}(1-qP_e)^N$, the SUs can transmit at any power without affecting the stability of the PU, while there exists a finite maximum allowed power if $\lambda_P > \mu_P^{\max}(1-qP_e)^N$. This can be understood by noting that $(1-qP_e)^N$ is the probability that none of the SUs transmit in a slot when the PU is busy, and hence, in addition to the simplicity of its decentralized operation, using random access as a multiple access protocol in the secondary network provides an additional protection to the primary. Note that, in practical situations, the transmission power of a node is also limited by the power amplifier used, but we ignore this aspect here.

3.4.3 Secondary Queues

A secondary node gets a packet served in the imperfect sensing case, either if the PU is idle with no false alarm occurring at that SU and that node transmits and is successful, or if the PU is busy with an incorrect detection of the PU occurring at that secondary node and the SU transmits and is successful.

Theorem 3.4. *The throughput of the jth secondary source node in the imperfect sensing case is*

$$\lambda_j = \left(1 - \frac{\lambda_P}{\mu_P}\right) \lambda_j^{(\text{P idle})} + \left(\frac{\lambda_P}{\mu_P}\right) \lambda_j^{(\text{P busy})}, \tag{3.22}$$

where $\lambda_j^{(\text{P idle})}$ and $\lambda_j^{(\text{P busy})}$ are the average throughput rates of the jth secondary node given that the primary node is idle, and busy respectively. and are given by Eqs. (3.28) *and* (3.29).

Proof: By the saturation assumption of the secondary queues, the average service rate of the primary node is independent of the states of the queues of the SUs (i.e. there is no queueing interactions) and hence, the probability that the primary node is idle = 1- Probability that primary node is busy = $1 - \frac{\lambda_P}{\mu_P}$.

The departure process at the jth secondary source node is given by

$$\begin{aligned} Y_j^t = & \sum_{\mathcal{F}^t \subseteq \{\mathcal{S}\setminus\{j\}\}} \sum_{\mathcal{T}^t \subseteq \{\mathcal{S}\setminus\mathcal{F}^t\}, j\in\mathcal{T}^t} \mathbb{1}\left\{\{Q_P^t = 0\} \bigcap \mathcal{F}^t \bigcap \mathcal{A}_{\mathcal{T}^t}^t \bigcap \overline{\mathcal{O}_{S_jD_j,\mathcal{F}^t,\mathcal{T}^t}^t}\right\} \\ & + \sum_{\mathcal{E}^t \subseteq \mathcal{S}, j\in\mathcal{E}^t} \sum_{\mathcal{T}^t \subseteq \mathcal{E}^t, j\in\mathcal{T}^t} \mathbb{1}\left\{\{Q_P^t \neq 0\} \bigcap \mathcal{E}^t \bigcap \mathcal{A}_{\mathcal{T}^t}^t \bigcap \overline{\mathcal{O}_{S_jD_j,\mathcal{E}^t,\mathcal{T}^t}^t}\right\}, \end{aligned} \tag{3.23}$$

where $\mathcal{S} = \{1, 2, ..., N\}$, $\mathcal{F}^t$ is the event that only the nodes in set $\mathcal{F}$ have a false alarm in slot t whenever the primary source node is idle, $\mathcal{E}^t$ is the event that only the nodes in set $\mathcal{E}$ have a misdetection of the primary node in slot t whenever the primary is busy, $\mathcal{A}_{\mathcal{T}^t}^t$ is the event that only nodes in set $\mathcal{T}^t$ transmit at time slot t. The event $\overline{\mathcal{O}_{S_jD_j,\mathcal{F}^t,\mathcal{T}^t}^t}$ is the event of no outage on the jth secondary source-destination link when the set $\mathcal{F}$ of nodes has false alarm and nodes in the set $\mathcal{T}$ of secondary nodes transmit simultaneously at time slot t, while the event $\overline{\mathcal{O}_{S_jD_j,\mathcal{E}^t,\mathcal{T}^t}^t}$ is the event of no outage on the jth secondary source-destination link when the set

$\mathcal{E}^t$ of nodes has misdetection of the activity of the primary node, and nodes in the set $\mathcal{T}^t$ of secondary nodes transmit simultaneously.

Since the process is stationary, we can drop the time indices.

The jth secondary node departure rate can be written as

$$\lambda_j = \mathbb{E}[Y_j^t] = \left(1 - \frac{\lambda_P}{\mu_P}\right) \lambda_j^{(\text{P idle})} + \left(\frac{\lambda_P}{\mu_P}\right) \lambda_j^{(\text{P busy})}. \tag{3.24}$$

If the primary node is idle, then the probability that a set $\mathcal{F} \subseteq \{\mathcal{S} \backslash \{j\}\}$ of secondary nodes has false alarms while all other secondary nodes do not is

$$\Pr[\mathcal{F} \mid \text{P is idle}] = \prod_{i \in \mathcal{F}} P_f^{(i)} \prod_{i \in \mathcal{S} \backslash \mathcal{F}} \left(1 - P_f^{(i)}\right), \tag{3.25}$$

then we can write

$$\lambda_j^{(\text{P idle})} = \sum_{\mathcal{F} \subseteq \mathcal{S} \backslash \{j\}} \Pr[\mathcal{F} \mid \text{P is idle}] \; \lambda_j^{(\text{P idle, } \mathcal{F})}, \tag{3.26}$$

where

$$\lambda_j^{(\text{P idle, } \mathcal{F})} = \sum_{\mathcal{T} \subseteq \mathcal{S} \backslash \mathcal{F},\, j \in \mathcal{T}} \left[\prod_{i \in \mathcal{T}} q_i \prod_{i \in \mathcal{T}^c} (1 - q_i)\right] \Pr\left[\frac{P_j |h_{S_j D_j}|^2 r_{S_j D_j}^{-\alpha}}{N_0 + \sum_{l \in \mathcal{T}} P_l |h_{S_l D_j}|^2 r_{S_l D_j}^{-\alpha}} > \beta_j\right], \tag{3.27}$$

and $\mathcal{T}^c = \mathcal{S} \setminus \{\mathcal{F} \bigcup \mathcal{T}\}$.

Hence

$$\lambda_j^{(\text{P idle})} = \sum_{\mathcal{F} \subseteq \mathcal{S} \backslash \{j\}} \left[\prod_{i \in \mathcal{F}} P_f^{(i)} \prod_{i \in \mathcal{S} \backslash \mathcal{F}} \left(1 - P_f^{(i)}\right)\right] \lambda_j^{(\text{P idle, } \mathcal{F})}. \tag{3.28}$$

Similarly

$$\lambda_j^{(\text{P busy})} = \sum_{\mathcal{E} \subseteq \mathcal{S},\, j \in \mathcal{E}} \left[\prod_{i \in \mathcal{E}} P_e^{(i)} \prod_{i \in \mathcal{S} \backslash \mathcal{E}} \left(1 - P_e^{(i)}\right)\right] \lambda_j^{(\text{P busy, } \mathcal{E})}, \tag{3.29}$$

where

$$\lambda_j^{(\text{P busy, } \mathcal{E})} = \sum_{\mathcal{T} \subseteq \mathcal{E},\, j \in \mathcal{T}} \left[\prod_{i \in \mathcal{T}} q_i \prod_{i \in \mathcal{E} \backslash \mathcal{T}} (1 - q_i)\right] P(\mathcal{T}), \tag{3.30}$$

and $P(\mathcal{T}) = \Pr\left[\frac{P_j|h_{S_jD_j}|^2 r_{S_jD_j}^{-\alpha}}{N_0+P_P|h_{S_PD_j}|^2 r_{S_PD_j}^{-\alpha}+\sum_{l\in\mathcal{T}} P_l|h_{S_lD_j}|^2 r_{S_lD_j}^{-\alpha}} > \beta_j\right]$. Finally, the throughput region of the secondary nodes is given by

$$\mathcal{L} = \bigcup_{(q_1,q_2,...,q_N)\in[0,1]^N} \Big\{(\lambda_1, \lambda_2, ..., \lambda_N)\Big\} \tag{3.31}$$

where λ_j is given by Eq. (3.24) for $j \in \mathcal{S} = \{1, 2, ..., N\}$. ■

Next, we specialize to the case of symmetric secondary cluster introduced in Section 3.2. In this case, $P_e^{(j)} = P_e$ and $P_f^{(j)} = P_f$ for all $j \in \{1, 2, ..., N\}$. The average throughput of the jth secondary node can be written as

$$\begin{aligned}\lambda_j =& \left(1-\frac{\lambda_P}{\mu_P}\right) q(1-P_f) \sum_{L=0}^{N-1} \binom{N-1}{l} (1-P_f)^L P_f^{N-1-L} \left[\sum_{k=0}^{L} \binom{L}{k} q^k (1-q)^{L-k} \lambda_j^{(\text{idle,k})}\right] \\ &+ \left(\frac{\lambda_P}{\mu_P}\right) qP_e \sum_{L=0}^{N-1} \binom{N-1}{L} P_e^L (1-P_e)^{N-1-L} \left[\sum_{k=0}^{L} \binom{L}{k} q^k (1-q)^{L-k} \lambda_j^{(\text{busy,k})}\right],\end{aligned} \tag{3.32}$$

where

$$\lambda_j^{(\text{idle,k})} = \Pr\left[\frac{P_0|h_{S_jD_j}|^2 r_j^{-\alpha}}{N_0 + \sum_{l=1}^{k} P_0|\tilde{h}_l|^2 r_j^{-\alpha}} > \beta\right] = \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 P_0 r_j^{-\alpha}}\right) \frac{1}{(1+\beta)^k}, \tag{3.33}$$

$$\begin{aligned}\lambda_j^{(\text{busy,k})} &= \Pr\left[\frac{P_0|h_{S_jD_j}|^2 r_j^{-\alpha}}{N_0 + \sum_{l=1}^{k} P_0|\tilde{h}_l|^2 r_j^{-\alpha} + P_P|h_{S_pD_j}|^2 r_j^{-\alpha}} > \beta\right] \\ &= \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 P_0 r_j^{-\alpha}}\right) \left(1 + \frac{P_P r_{S_PD_j}^{-\alpha} \beta \sigma_{S_PD_j}^2}{P_0 r_j^{-\alpha} \tilde{\sigma}^2}\right)^{-1} \frac{1}{(1+\beta)^k}.\end{aligned} \tag{3.34}$$

After some algebra, the throughput of the jth secondary source node is written as

$$\begin{aligned}\lambda_j =& \left(1-\frac{\lambda_P}{\mu_P}\right) \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 P_0 r_j^{-\alpha}}\right) q(1-P_f) \left[1 - q(1-P_f)\frac{\beta}{\beta+1}\right]^{N-1} \\ &+ \left(\frac{\lambda_P}{\mu_P}\right) \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 P_0 r_j^{-\alpha}}\right) qP_e \left(1 + \frac{P_P r_{S_PD_j}^{-\alpha} \beta \sigma_{S_PD_j}^2}{P_0 r_j^{-\alpha} \tilde{\sigma}^2}\right)^{-1} \left[1 - qP_e\frac{\beta}{\beta+1}\right]^{N-1}.\end{aligned} \tag{3.35}$$

The secondary nodes aim at maximizing their throughputs (for example sum throughput) by optimizing P_0 and q subject to the constraints given by Eqs. (3.20) and (3.21). That is by solving

$$\max_{P_0 \geq 0,\ q \geq 0} \sum_{j=1}^{N} \lambda_j$$

$$\text{s.t. Eqs. (3.20) and (3.21),}$$

$$\text{where } \lambda_j \text{ is given by Eq. (3.35).}$$

The objective function is however non-convex and the solution is hard to find in closed form. In Section 3.6, we present numerical solutions to that optimization problem and in the following, we present some intuitive properties of the solution based on the structure of the objective function. Clearly, the smaller the value of the false alarm probability P_f, the higher the fraction of idle slots in which the SUs access the channel without interference from the primary. The secondary throughput of node j (λ_j) decreases with P_f because of the smaller fraction of idle slots accessed by the SUs but also increases with P_f due to less interference between them. Such variation depends on other parameter values, but in general, if N is small, then the first effect can be significant while if N is large enough, then only the second effect becomes prevalent. Let $I = \frac{P_P r_{S_P D_j}^{-\alpha} \sigma_{S_P D_j}^{2\beta}}{P_0 r_j^{-\alpha} \tilde{\sigma}^2}$, which is proportional to the ratio of the interference of the PU on an SU to the SU's transmission power. If this term is large, then the PU highly interferes with the SU and the throughputs of the SUs when the PU is busy are largely reduced. In this case, the secondary throughput is dominated by the first term of Eq. (3.35). On the other hand, if I is small enough, then the second term of Eq. (3.35) might become significant and the interference from the PU

does not significantly reduce the throughputs of the SUs. In this case, λ_j increases with increasing P_e because of more opportunities for the SUs to transmit when the PU is busy despite the fact that the term $[1-qP_e\frac{\beta}{\beta+1}]^{N-1}$ decreases with P_e because of the secondary interference on each other. The variation of λ_j with $\frac{\lambda_P}{\mu_P^{\max}}$ is more subtle. On one hand, the fraction of idle slots $1-\frac{\lambda_P}{\mu_P}$ decreases with increasing $\frac{\lambda_P}{\mu_P^{\max}}$; but also $\frac{\lambda_P}{\mu_P}$ which represents the fraction of busy slots increases. Such variation is highly dependable on the other parameters. For small values of $\frac{\lambda_P}{\mu_P^{\max}}$, secondary nodes (by Eq. (3.21)) can transmit at their maximum power and get increasing throughputs due to the increase in the fraction of busy slots, especially when P_e is high until $\frac{\lambda_P}{\mu_P^{\max}}$ reaches a value at which transmission power P_0 and/or transmission probabilities q should decrease to limit the interference on the primary and thus secondary throughputs decrease as well. Finally, it should be noted that sensing errors might lead to higher SU throughputs compared with the perfect sensing case (in contrast with the PU where imperfect sensing always leads to lower maximum stable throughput). This can be explained by observing that incorrect sensing gives the secondary nodes more opportunities for transmission during the busy slots of the PU which might lead to a net increase in throughputs especially when $\frac{\lambda_P}{\mu_P^{\max}}$ is large. Such observations will become clearer in Section 3.6.

3.5 Relaying in the Perfect Sensing Case

Primary users would be willing to share their channel resources with secondary users if they benefit from such sharing. Hence, forcing the secondary nodes to relay

the primary node's unsuccessful packets would be the price that the SUs pay for accessing the PU's channel and the incentive for the PU to share his resources with the SUs. Moreover, by relaying the PU's packets, the SUs might benefit from the increase of the number of idle slots available for their transmissions. In this section, we propose and analyze a distributed cooperative protocol between the secondary and primary nodes. We restrict the analysis to the perfect sensing case[2]. We also restrict the analysis to the symmetric case as described previously. Similar properties of the results can be shown for the asymmetric case with arbitrary fading distributions as shown in Section 3.8.2.

3.5.1 Relaying Protocol

The relaying protocol achieves throughput gain with no channel state information (CSI) about $h_{S_k D_P}$ fading coefficients available at the SUs by using Distributed Orthogonal Space-Time Block Code (D-OSTBC). That is, each of the SUs that are able to successfully decode a PU's unsuccessful packet mimics an antenna in a regular Space-Time Code (STC) setting of a multiple-input single-output (MISO) channel. Such OSTBC schemes always exist for one dimensional signal constellations for any number of relaying nodes [65]. For that case, these OSTBC schemes achieve full diversity gain at coding rate =1 while ensuring simple decoding rule based on linear processing at the receiver.

Remarks:

[2]As will be seen, the imperfect sensing case involve solving for $N+1$ interacting queues which is, to date, an open problem for $N > 1$ [43].

1. Each of the relaying nodes must know which antenna it mimics in the underlying STC used. This can be achieved by either some coordination between the SUs, or by a prior node indexing and observing ACK/NACK packets generated by the SUs regarding the primary packet. Those ACK/NACK messages are assumed to be available to all nodes throughout the network.

2. If the packet length is not an integer multiple of the number of relaying nodes, the last block of symbols in the packet is relayed by a smaller number of nodes. However, such effect is typically small since for all practical situations the number of symbols per packet is much larger than the number of relaying nodes and thus, we ignore such "edge effects" in the sequel.

3. For two-dimensional constellations, it is shown in [66] that the rates of complex orthogonal spacetime block codes for more than two transmit antennas are upper-bounded by 3/4, while the rates of generalized complex orthogonal spacetime block codes for more than two transmit antennas are upper-bounded by 4/5. In this part, we mainly focus on the case of one dimensional constellation with code rate =1 in order to avoid both analytical and practical issues (such as synchronization problems) related to variable rate systems.

The relaying protocol works as follows. At every busy slot of the primary node, if one or more SUs are able to successfully decode the packet sent by the PU while the primary destination can not, then these SUs store this packet in a special queue (relaying queue) and send an ACK feedback to the primary, and the primary node releases the packet from its queue. We assume that this ACK messages will also be heard by all the SUs and thus the SUs which could not receive that packet will

abstain from transmission until that packet is successfully delivered to the primary destination and thus avoiding interfering with the PU's relayed packets. In the next available PU's idle slot, the SUs which were able to decode the PU packet will transmit it using D-OSTBC as described above. It should be noted that the primary packets are given priority for transmission, i.e., a secondary source node will not transmit its own packets unless it does not have any primary packets to relay and none of the other secondary nodes has any.

We illustrate the protocol operation by the following example for $N = 3$, where 'B' stands for Busy, 'I' stands for Idle, 'S' stands for Success and 'F' stands for Fail.

Time Slot	1	2	3	4	5	6	7	8
PU	B, S	B, F	B, S	B, F	I	I	B, S	I
SU1		S		F	B(P2),F	B(P2),S		I
SU2		F		S	I	I		B(P4),S
SU3		S		S	B(P2),F	B(P2),S		B(P4),S

In the first slot, the PU transmits a packet and is successful. The packet is then released from the system. In the second slot, the PU has a failed transmission while SU1 and SU3 were able to decode that packet. They then send an ACK to the PU transmitter which drops the packet from its queue. Meanwhile, SU2 receives that ACK and knows that packet 2 of the PU was not successful and that SU1 and SU3 were able to successfully decode it. Subsequently, SU2 will not transmit until receiving an ACK from the PU destination regarding packet 2. In slot 3, the PU has packet 3 successfully delivered to the PU destination. In slot 4, the PU has a

failed transmission, but both SU2 and SU3 are able to successfully decode it. SU2 and SU3 send an ACK to the PU. SU1 also receives that ACK. Then, once packet 2 is delivered, SU1 will not transmit until packet 4 is delivered. In slot 5, the PU becomes idle. SU2 knows that packet 2 has not been delivered yet and that SU1 and SU3 have it, then SU2 remains silent. Both SU1 and SU3 transmit packet 2 using D-OSTBC but the PU destination can not decode it. In slot 6, they retransmit it and are successful. In slot 7, the PU transmits and is successful. In slot 8, the PU is idle, and both SU2 and SU3 transmit packet 4 and are successful.

3.5.2 Protocol Analysis

We proceed by proving the main result for symmetric network with arbitrary fading distributions, and then we provide closed form expressions for the special case of Rayleigh fading. For the case of asymmetric network, similar results hold as proved in Section 3.8.2.

For a secondary source node to successfully decode a primary packet, the minimum required SNR value is β_P. Let P_c be the probability that one of the SUs is able to successfully decode the PU's packet (which is same for all secondary nodes by symmetry), then

$$P_c = \Pr\left[\frac{P_P|h_{S_PS_j}|^2 r_{S_PS_j}^{-\alpha}}{N_0} > \beta_P\right] = \Pr\left[|h_{S_PS_j}|^2 > \frac{\beta_P N_0}{P_P r^{-\alpha}}\right]. \tag{3.36}$$

Let M^t be a random variable denoting the number of secondary source nodes that successfully decoded a primary's packet in time slot t, then

$$\Pr[M^t = m] = \binom{N}{m} P_c^m (1-P_c)^{N-m},\ m = 0, 1, ..., N. \tag{3.37}$$

According to the above described protocol, the primary queue service process takes the form $Y_P^t = \mathbb{1}\left\{\overline{\mathcal{O}_{S_P D_P}^t} \bigcup \overline{\mathcal{O}_{S_P S_S}^t}\right\}$, where $\overline{\mathcal{O}_{S_P D_P}^t}$ and $\overline{\mathcal{O}_{S_P S_S}^t}$ denote the events of no outage on the primary source-primary destination link and the event that at least one secondary source node was able to successfully decode the packet, respectively. Clearly, the service process at the primary source queue is stationary. Hence, the success probability is given by

$$\mu_P = \mathbb{E}\left[Y_P^t\right] = \mu_P^{\max} + [1-(1-P_c)^N] - \mu_P^{\max}[1-(1-P_c)^N] = 1-(1-\mu_P^{\max})(1-P_c)^N, \tag{3.38}$$

which is strictly greater than $\mu_P^{\max}$.

It should be noted that stability of the relaying queues at the SUs (as will come clearer later in Eq. (3.45)) guarantees that the PU's packets that were successfully decoded at the secondary nodes will be eventually delivered to the primary destination because the SU's relaying queues, due to their stability, empty infinitely often.

Theorem 3.5. *Under the previously described relaying protocol, the stability condition of the system is*

$$\lambda_P < \frac{\mu_P P_s}{P_s + (1-\mu_P^{\max})P_c}, \tag{3.39}$$

where

$$P_s = \sum_{k=0}^{N-1} \binom{N-1}{k} P_c^k (1-P_c)^{N-1-k} \Pr\left[\sum_{i=1}^{k+1} |h_i|^2 > \frac{\beta_P N_0}{P_0 r_0^{-\alpha}}\right]. \tag{3.40}$$

Proof: Each SU, in addition to his own packets to be transmitted, has exogenous packet arrivals from the primary source node to be relayed in the subsequent idle

slots. The arrival process to the secondary source node from the primary is

$$X_{ext}^{P} = \mathbb{1}\left\{\{Q_P^t \neq 0\} \bigcap \mathcal{O}_{S_P D_P}^t \bigcap \overline{\mathcal{O}_{S_P S}^t}\right\}, \tag{3.41}$$

where $\mathcal{O}_{S_P D_P}^t$ is the outage event on the primary source-destination link and $\overline{\mathcal{O}_{S_P S}^t}$ is the event of no outage on the primary source-secondary source link at time slot t. The SNR per symbol of the relayed packet at the primary destination node given that k secondary nodes simultaneously transmit using D-OSTBC is given by

$$SNR = \frac{P_0 r_0^{-\alpha} \sum_{i=1}^{k} |h_i|^2}{N_0}. \tag{3.42}$$

Hence, the probability of no outage given $k+1$ nodes transmit in slot t is given by

$$\Pr[\overline{\mathcal{O}_{SD_P,k+1}^t}] = \Pr\left[\sum_{i=1}^{k+1} |h_i|^2 > \frac{\beta_P N_0}{P_0 r_0^{-\alpha}}\right]. \tag{3.43}$$

The service process of the primary packets queued at a secondary source node is given by

$$Y_{ext}^{P} = \sum_{k=0}^{N-1} \mathbb{1}\left\{\{Q_P^t = 0\} \bigcap \{\bar{M}^t = k\} \bigcap \overline{\mathcal{O}_{SD_P,k+1}^t}\right\}, \tag{3.44}$$

where $\bar{M}$ is a random variable denoting the number of other SUs that could decode the packet in service, and $\overline{\mathcal{O}_{SD_P,k+1}^t}$ is the event of no outage at the PU destination when $k+1$ SUs collaboratively transmit the relayed PU's packet.

The arrival and service processes X_{ext}^P and Y_{ext}^P are jointly stationary and hence by Loynes' theorem, we can get the condition of stability as

$$\lambda_{ext}^P = \mathbb{E}\left[X_{ext}^P\right] = \left(\frac{\lambda_P}{\mu_P}\right)(1-\mu_P^{\max})P_c < \mu_{ext}^P = \mathbb{E}\left[Y_{ext}^P\right] = \left(1-\frac{\lambda_P}{\mu_P}\right)P_s, \tag{3.45}$$

where P_s is as given by Eq. (3.40).

For stability of the primary queue, we should also have that $\lambda_P < \mu_P$ and Eq. (3.45)

satisfied, leading to

$$\lambda_P < \lambda_P^{\max} = \frac{\mu_P P_s}{P_s + (1 - \mu_P^{\max}) P_c}, \tag{3.46}$$

where $\lambda_P^{\max}$ is the maximum stable throughput rate of the primary queue. ■

Proposition 3.2. *The success probability P_s as given by* Eq. (3.40) *is strictly increasing with N. Moreover, as $N \to \infty$, $P_s \to 1$.*

Proof: Refer to Section 3.8.3. ■

Proposition 3.3. *The maximum possible arrival rate at the primary node that keeps the system stable as given by* Eq. (3.46) *is higher than in the case of no-relaying only if $\mu_P^{\max} < \frac{P_s[1-(1-P_c)^N]}{P_c}$.*

Proof: Follows immediately by setting

$$\mu_P^{\max} < \frac{\mu_P P_s}{P_s + (1 - \mu_P^{\max}) P_c},$$

and substituting μ_P from Eq. (3.38). ■

The term $\frac{1-(1-P_c)^N}{P_c}$ is always bounded between 1 and $1/P_c$ and is increasing with N. Hence, a sufficient condition for the condition in Proposition 3.3 to be satisfied is to have $\mu_P^{\max} < P_s$ which, by Proposition 3.2, is clearly satisfied for some N, possibly large, since P_s can be made arbitrarily close to 1 by increasing the number of secondary nodes. This attracts the attention that the more SUs the PU shares the channel with, the more benefit for the PU in terms of his stable throughput. It should be noted that increasing the transmission power P_0 of the SUs leads to satisfying the condition in Proposition 3.3 for a smaller number of secondary relaying nodes.

We also note that one node relaying ($N = 1$) always leads to higher primary stable throughput rate if

$$\frac{(1-(1-\mu_P^{\max})(1-P_c))P_s}{P_s+(1-\mu_P^{\max})P_c} > \mu_P^{\max}, \tag{3.47}$$

which can be simplified to

$$\mu_P^{\max} = \Pr\left[\frac{P_P|h_{S_PD_P}|^2 r_{S_PD_P}^{-\alpha}}{N_0} > \beta_P\right] < P_{\mathrm{s}} = \Pr\left[\frac{P_0|h_i|^2 r_0^{-\alpha}}{N_0} > \beta_P\right].$$

We then proceed to characterize the effect of relaying on the secondary nodes. As previously mentioned, secondary nodes might benefit from relaying the packets of the primary. This can be understood by noting that relaying the packets of the primary helps the primary's queue to become idle more often, that is, a larger number of idle slots will be available for the secondary nodes. However, a portion of that fraction is used for relaying the packets of the primary. If that fraction is smaller than the additional fraction available for the secondary by relaying, a net increase in throughput is achieved for the secondary nodes. It should be noted that even if the secondary nodes suffer from some reduction in throughput by relaying, they still achieve some non-zero throughput by accessing the resources of the primary and relaying can then be looked at as the price to pay for opportunistically accessing the channel. On the other hand, as previously shown, the primary always benefits from relaying when N is sufficiently large and this is the incentive to share his resources with more SUs.

The effect of the relaying protocol on the SUs can be found by first rewriting Eq. (3.11) as $\lambda_j = \left(1 - \frac{\lambda_P}{\mu_P^{\max}}\right)\tilde{\lambda}_j$. An SU transmits its own traffic in a slot only if the

slot is idle, the SU does not have any primary traffic to relay and no other nodes has any. Note that when an SU transmits, it behaves as in the perfect sensing case; thus we obtain the throughput of the jth secondary node as

$$\lambda_j^{\text{relaying}} = \left(1 - \frac{\lambda_P}{\mu_P}\right)\left(1 - \frac{\lambda_{ext}^P}{\mu_{ext}^P}\right)^N \tilde{\lambda_j}. \tag{3.48}$$

On one hand, the fraction of idle slots available to the secondary nodes is now $\left(1 - \frac{\lambda_P}{\mu_P}\right)$ which is larger than the fraction available to the secondary nodes with no relaying $\left(= 1 - \frac{\lambda_P}{\mu_P^{\max}}\right)$ since $\mu_P > \mu_P^{\max}$. However, the SU's throughput is reduced by a factor of $\left(1 - \frac{\lambda_{ext}^P}{\mu_{ext}^P}\right)^N$ due to the priority given to the relayed primary's packets instead of transmitting the SU's own packets. By increasing N, the term $\left(1 - \frac{\lambda_{ext}^P}{\mu_{ext}^P}\right)^N$ decreases due to the priority to transmit the packets of the primary which is the price for the secondary nodes to opportunistically access the channel. We make this precise in the following proposition:

Proposition 3.4. *The secondary nodes benefit from relaying only if*

$$\left(1 - \frac{\lambda_P}{\mu_P^{\max}}\right) < \left(1 - \frac{\lambda_P}{\mu_P}\right)\left(1 - \frac{\lambda_{ext}^P}{\mu_{ext}^P}\right)^N. \tag{3.49}$$

In particular, for a system with a single secondary node ($N = 1$), this condition is equivalent to

$$\frac{(1 - (1 - \mu_P^{\max})(1 - P_c))P_{\text{s}}}{P_{\text{s}} + (1 - \mu_P^{\max})P_c} > \mu_P^{\max} \Leftrightarrow \mu_P^{\max} < P_s.$$

Proof: Follows directly by rewriting Eq. (3.11) as $\lambda_j^{\text{no relaying}} = \left(1 - \frac{\lambda_P}{\mu_P^{\max}}\right)\tilde{\lambda_j}$ and using Eq. (3.48). The case for $N = 1$ follows after some algebra by substituting $N = 1$ in Eq. (3.38) and substituting in Eq. (3.49). ■

The condition in Proposition 3.4 for the secondary node to benefit from relaying in the case of $N = 1$ is identical to the condition in Eq. (3.47) for the primary node to benefit from relaying in case of one secondary node. This means that with one secondary node, either both the primary and secondary nodes benefit from relaying or none of them does.

For the imperfect sensing case, although having more SUs leads to more potential interference with the primary node, it also leads to more benefits of relaying as discussed above. Moreover, it leads to higher opportunities for cooperative channel sensing [67] leading to more accurate sensing results by reducing both false alarm and misdetection probabilities converging to the perfect sensing case discussed above. Quantifying the effect of sensing errors on our cooperative protocol require solving for $N + 1$ interacting queues, which is an open problem to date.

Finally, for the special case of Rayleigh fading, the different probabilities and throughputs are given by

$$\lambda_j = \left(1 - \frac{\lambda_P}{\mu_P}\right)\left(1 - \frac{\lambda_{ext}^P}{\mu_{ext}^P}\right)^N \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 r_j^{-\alpha} P_0}\right) q\left[1 - q\frac{\beta}{1+\beta}\right]^{N-1}, \tag{3.50}$$

$$P_c = \exp\left(\frac{-\beta_P N_0}{\sigma^2 P_P r^{-\alpha}}\right), \tag{3.51}$$

$$P_{\mathrm{s}} = \sum_{k=0}^{N-1} \binom{N-1}{k} \frac{P_c^k (1-P_c)^{N-1-k}}{k!} \Gamma\left(k+1, \frac{\beta_P N_0}{P_0 r_0^{-\alpha} \sigma_0^2}\right), \tag{3.52}$$

where $\Gamma(s, x)$ is the upper incomplete gamma function and can be represented by the integral $\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t}\, dt$.

For this special case, one node relaying ($N = 1$) always leads to a higher primary stable throughput rate if

$$\frac{(1-(1-\mu_P^{\max})(1-P_c))P_s}{P_s+(1-\mu_P^{\max})P_c} > \mu_P^{\max},$$

which can be simplified to

$$\mu_P^{\max} = \exp\left(\frac{-\beta_P N_0}{P_P \sigma^2_{S_P D_P} r^{-\alpha}_{S_P D_P}}\right) < P_{\rm s} = \Gamma\left(1, \frac{\beta_P N_0}{P_0 r_0^{-\alpha}\sigma_0^2}\right) = \exp\left(\frac{-\beta_P N_0}{P_0 \sigma_0^2 r_0^{-\alpha}}\right)$$

$$\Leftrightarrow P_0 r_0^{-\alpha}\sigma_0^2 > P_P r^{-\alpha}_{S_P D_P}\sigma^2_{S_P D_P}$$

$$\Leftrightarrow \mathbb{E}[\text{SNR on S-P link}] > \mathbb{E}[\text{SNR on P-P link}]. \tag{3.53}$$

In other words, assuming same transmission power for the PU and the SUs, one node relaying always helps both the primary and the secondary user (by Proposition 3.4) if the channel between secondary source and primary destination is on average better than the channel between the primary source and primary destination.

3.6 Numerical Results

In this section, we provide numerical results to illustrate the conclusions drawn analytically. The values of the parameters are chosen based on practical considerations, but also for the sake of clarity of presentation. Figures 3.2 and 3.3 illustrate the effect of erroneous sensing on the normalized maximum stable throughput of the primary node as given by Eq. (3.18). The term $\frac{\sigma^2_{S_P D_P} r^{-\alpha}_{S_P D_P} P_P}{\beta_P r_0^{-\alpha}\sigma_0^2}$ is fixed at value 10 dBW throughout Figs. 3.2 and 3.3. In Fig. 3.2, we plot the normalized maximum stable throughput of the primary node versus the SUs' transmission power. It shows that μ_P can severely drop from its perfect sensing value $\mu_P^{\max}$ even for small number of SUs and small values of qP_e, and shows that secondary nodes can effec-

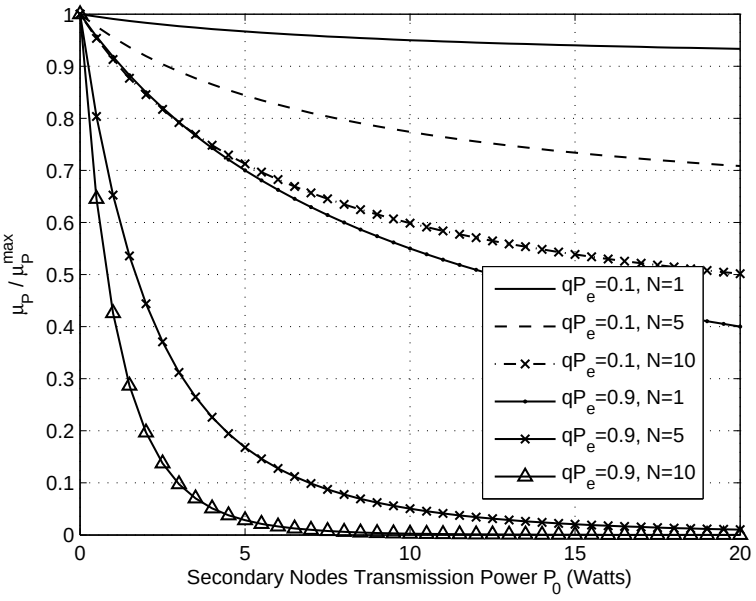

Figure 3.2: Effect of secondary transmission power on primary maximum stable throughput rate.

tively limit their interference on the primary by controlling their transmission power P_0, their channel access probability q or by enhancing the sensing performance to reduce P_e. Figure 3.3 shows the normalized maximum stable throughput rate at the PU versus the number of SUs N showing a similar effect. However, it should be noted that as shown in Fig. 3.2, $\lim_{a\to 0} \frac{\mu_P}{\mu_P^{\max}} = \lim_{P_0\to\infty} \frac{\mu_P}{\mu_P^{\max}} = (1 - qP_e)^N$, $\lim_{qP_e\to 1} \frac{\mu_P}{\mu_P^{\max}} = [1 - \frac{1}{a+1}]^N$ while $\lim_{N\to\infty} \frac{\mu_P}{\mu_P^{\max}} = 0$; meaning that for low enough primary arrival rates λ_P, controlling the transmission parameters of the secondary nodes is not as crucial as controlling the number of secondary transmissions in the system. This motivates the relaying protocol described earlier whose performance is illustrated in Figs. 3.4 and 3.5. Let $SNR = \frac{P_0 r_0^{-\alpha} \sigma_0^2}{\beta_P N_0}$. Figures 3.4 and 3.5 show

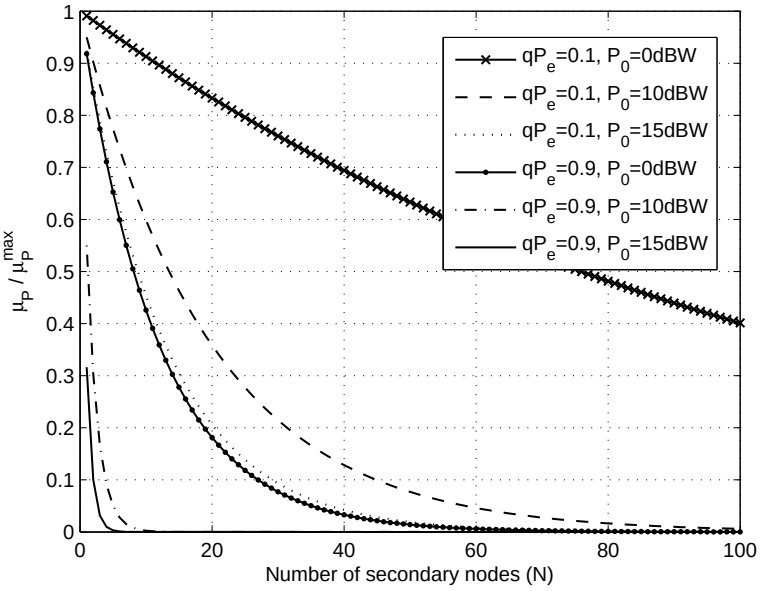

Figure 3.3: Effect of number of secondary nodes on primary maximum stable throughput rate.

the maximum stable throughput rate at the primary node ($\lambda_P^{\max}$) as given by Eq. (3.46) versus the number of secondary nodes for different SNR values for the case of Rayleigh fading as given by Eqs. (3.51) and (3.52). With no relaying, $\lambda_P^{\max} = \mu_P^{\max}$ and is shown by the horizontal line at $\mu_P^{\max} = 0.3$. It is clear that regardless of the values of the parameters, sufficiently large N always outperforms the non-relaying case and with higher SNR, a smaller number of secondary nodes is needed to outperform non-relaying. We also note that at SNR = 0 dB, even a single relaying node outperforms the non-relaying case. Figures 3.6 and 3.7 show the SUs' maximum throughput (optimized over q and P_0) versus the normalized average arrival rate at the primary node $\lambda_P/\mu_P^{\max}$ for different values of $I = \frac{P_P r_{S_P D_j}^{-\alpha} \beta \sigma_{S_P D_j}^2}{r_j^{-\alpha} \tilde{\sigma}^2}$ in both perfect and imperfect sensing cases as given by Eqs. (3.11) and (3.35), respectively. Note that for each value of $\lambda_P/\mu_P^{\max}$, the feasible set of (q, P_0) may be different to

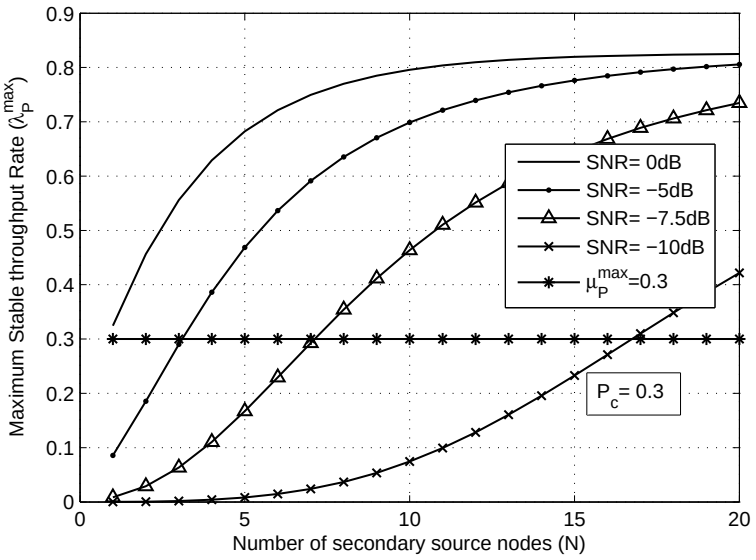

Figure 3.4: Effect of relaying on maximum stable throughput rate ($\lambda_P^{\max}$) for decoding probability $P_c = 0.3$.

ensure the protection of the primary node as in Eqs. (3.20) and (3.21). The value of the secondary threshold β is fixed at 10, $\frac{\beta N_0}{\tilde{\sigma}^2 r_j^{-\alpha}} = -5$ dBW,$\frac{\sigma^2_{S_P D_P} P_P r^{-\alpha}_{S_P D_P}}{\sigma_0^2 \beta_P r_0^{-\alpha}} = 0$ dBW and $P_f = 0.2$. We also impose a maximum possible value on P_0 equal to 10dBW which is a typical constraint imposed by the hardware. Figure 3.6 shows the secondary throughput for $I = 100$ which is the case where the primary node exerts high interference on the secondary nodes. In this case, perfect sensing leads to a higher throughput compared to imperfect sensing. Furthermore, the throughput λ_j decreases with increasing the error probability P_e because of the decrease of the fraction of the idle slots that are primary interference free in spite of the increase of the busy slots suffering from high primary interference which cannot balance the reduction of the relatively high throughput acquired in idle slots. Figure 3.7 shows the case of $I = 0.1$ which is the case of very low interference from the primary.

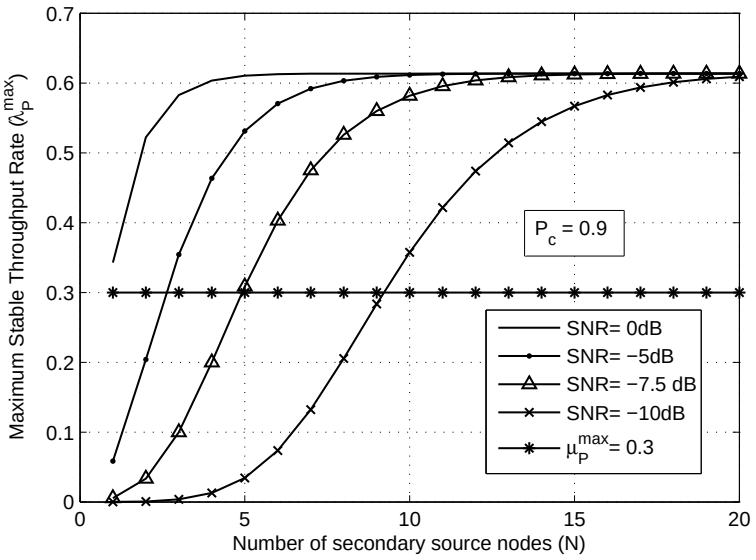

Figure 3.5: Effect of relaying on maximum stable throughput rate ($\lambda_P^{\max}$) for decoding probability $P_c = 0.9$.

In contrast with Fig. 3.6 in which the SUs' throughput decreases with increasing $\lambda_P/\mu_P^{\max}$, for some parameters values (for instance, $N = 1, P_e = 0.9$) secondary throughput increases with $\lambda_P/\mu_P^{\max}$. Moreover, except for $N = 1$, incorrect sensing leads to a higher throughput than perfect sensing, and increasing P_e leads to an increase in throughput. Hence, in this case, although increasing P_e might harm the primary node, secondary nodes benefit in terms of their throughputs. This is due to the increase of the opportunities at which the SUs access the channel as the fraction of busy slots suffering from low primary interference increases. This case is appealing to the secondary nodes if the primary arrival rate is low enough allowing them to increase P_e to the level which does not affect the primary node stability as discussed previously.

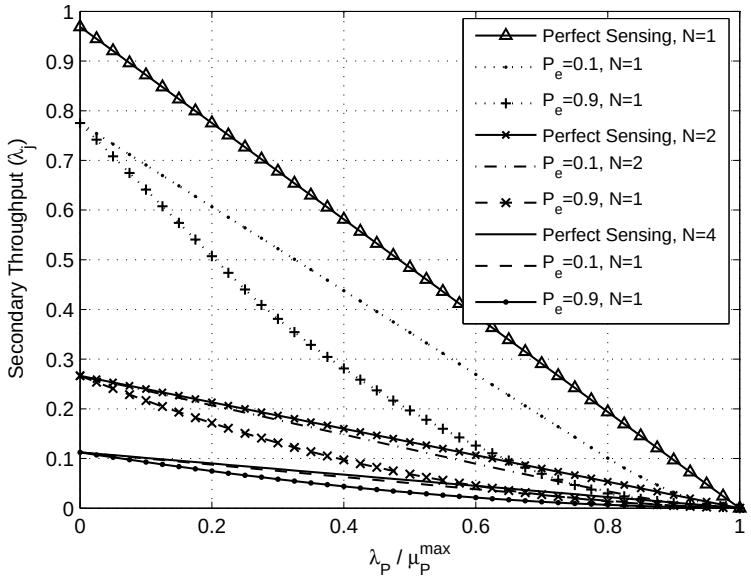

Figure 3.6: Secondary throughput versus primary normalized arrival rate for various values of P_e and N; $I = 100$ (case of high interference from the primary).

3.7 Summary and Conclusions

In this chapter, we studied the effect of the number of secondary nodes and their transmission parameters on the stable throughput of the primary user as well as on the secondary's throughputs in both perfect and imperfect sensing cases. It was shown that secondary transmission parameters (power and channel access probabilities) must depend on the arrival rate of the primary to ensure some protection to the primary. If the arrival rate at the PU is less than some calculated finite value, there is no need for controlling their parameters, otherwise, secondary nodes have to control their transmission parameters to limit their interference on the primary and avoid affecting its stability. The number of secondary users can be a benefit or a hindrance. If the secondary nodes do not relay the primary's unsuccessful packets, their

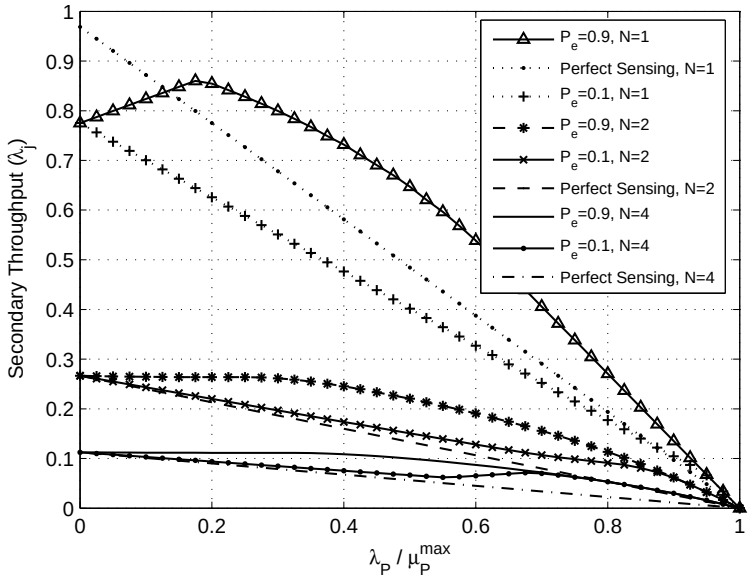

Figure 3.7: Secondary throughput versus primary normalized arrival rate for various values of P_e and N; $I = 0.1$ (case of very low interference from the primary).

presence is a harm for the primary as it reduces its maximum stable throughput. However, if the secondary nodes are forced to relay the primary's unsuccessful packets, then the primary always benefits from having many nodes relaying its packets. Secondary nodes might benefit from relaying by having access to a larger fraction of idle slots. This observation reveals that with appropriate relaying protocols, cognitive radio technology is appealing for licensed users to share their resources with other unlicensed users.

3.8 Appendix

3.8.1 Proof of Eqs. (3.8) and (3.10)

We use the following lemma in the proof.

Lemma 3.1. *Let $X_i \sim \exp(\theta_i)$ be independent random variables, then the probability density function of the sum $Z = \sum_{i=1}^{N} X_i$ is given by*

$$f_Z(z) = \sum_{i=1}^{N} \left(\prod_{\substack{j=1 \\ j \neq i}}^{N} \frac{\theta_j \theta_i}{\theta_j - \theta_i} \right) \exp(-\theta_i z). \tag{3.54}$$

If the random variables X_i are also identically distributed, i.e., $X_i \sim \exp(-\theta)$ for all i, then the sum $Z = \sum_{i=1}^{N} X_i$ has pdf given by the Erlang distribution

$$f_Z(z) = \theta \exp(-\theta z) \frac{(\theta z)^{N-1}}{(N-1)!}. \tag{3.55}$$

Proof: The proof follows by induction. Refer to [68] for details. ■

Hence, for the case of asymmetric configuration:

$$\begin{aligned}
&\Pr\left[\frac{P_{S_j} |h_{S_j D_j}|^2 r_{S_j D_j}^{-\alpha}}{N_0 + \sum_{k \in \mathcal{T}, k \neq j} P_k |h_{S_k D_j}|^2 r_{S_k D_j}^{-\alpha}} > \beta_j \right] = \Pr\left[|h_{S_j D_j}|^2 > \frac{\beta_j N_0}{P_j r_{S_j D_j}^{-\alpha}} + Z \right] \\
&= \int_0^{\infty} \Pr\left[|h_{S_j D_j}|^2 > \frac{\beta_j N_0}{P_j r_{S_j D_j}^{-\alpha}} + z \right] \sum_{\substack{k \in \mathcal{T} \\ k \neq j}} \prod_{l \neq k} \frac{\theta_l \theta_k}{\theta_l - \theta_k} \exp(-\theta_k z)\, dz \\
&= \exp\left(\frac{-N_0 \beta_j}{\sigma_{S_j D_j}^2 P_j r_{S_j D_j}^{-\alpha}} \right) \sum_{\substack{k \in \mathcal{T} \\ k \neq j}} \left(\prod_{l \neq k} \frac{\theta_l \theta_k}{\theta_l - \theta_k} \right) \frac{1}{\left(\theta_k + 1/\sigma_{S_j D_j}^2 \right)},
\end{aligned} \tag{3.56}$$

where $\theta_k = \frac{P_j r_{S_j D_j}^{-\alpha}}{P_k r_{S_k D_j}^{-\alpha} \sigma_{S_k D_j}^2 \beta_j}$.

For the case of symmetric network:

$$\begin{aligned}
&\Pr\left[\frac{P_0 |h_{S_j D_j}|^2 r_j^{-\alpha}}{N_0 + \sum_{m=1}^{k} P_0 |h_m|^2 r_j^{-\alpha}} > \beta \right] = \Pr\left[|h_{S_j D_j}|^2 > \frac{\beta N_0}{P_0 r_j^{-\alpha}} + \beta \sum_{l=1}^{k} |h_{S_l D_j}|^2 \right] \\
&= \int_0^{\infty} \Pr\left[|h_{S_j D_j}|^2 > \frac{\beta N_0}{P_0 r_j^{-\alpha}} + \beta z \right] \frac{\exp(-z/\tilde{\sigma}^2)(z/\tilde{\sigma}^2)^{k-1}}{\tilde{\sigma}^2 (k-1)!}\, dz \\
&= \exp\left(\frac{-\beta N_0}{\tilde{\sigma}^2 P_0 r_j^{-\alpha}} \right) \frac{1}{(1+\beta)^k},
\end{aligned} \tag{3.57}$$

where we have used that $\int_0^{\infty} x^{k-1} \exp(-x)\, dx = (k-1)!$, for k integer.

3.8.2 Proof of Relaying Protocol in the General Asymmetric Case

We use the following Lemma in the proof.

Lemma 3.2. *If $X_1, X_2, ..., X_N$ are positive random variables, then for every $c > 0$:*

$$\Pr\left[\sum_{i=1}^{k+1} X_i > c\right] > \Pr\left[\sum_{i=1}^{k} X_i > c\right],\ k \in \{1, 2, ..., N-1\} \tag{3.58}$$

Proof: Follows immediately by induction. ■

Let the flat fading coefficients between the primary source and the jth secondary source nodes be $h_{S_P S_j}$. The probability that the jth secondary node is able to successfully decode the primary's packet is then given by

$$P_c^{(j)} = \Pr\left[\frac{P_P |h_{S_P S_j}|^2 r_{S_P S_j}^{-\alpha}}{N_0} > \beta_P\right],\ j \in \{1, 2, ..., N\}. \tag{3.59}$$

Let $\mathcal{S} = \{1, 2, ..., N\}$ be the set of secondary nodes. The probability that some subset of the N secondary nodes is able to successfully decode the primary's packet is given by

$$P_c^{(N)} = 1 - \prod_{k \in \mathcal{S}} \left(1 - P_c^{(k)}\right), \tag{3.60}$$

where the superscript N denotes the number of secondary nodes in the system. According to the relaying protocol, the primary node is served either when the primary destination can successfully decode the packet or when the primary destination cannot, but one or more secondary source nodes can. Hence, the average service rate of the primary node is given by

$$\mu_P = \mu_P^{\max} + \left(1 - \mu_P^{\max}\right) P_c^{(N)} = P_c^{(N)} + \left(1 - P_c^{(N)}\right) \mu_P^{\max}, \tag{3.61}$$

which is clearly greater than $\mu_P^{\max}$.

If the relaying queues of the secondary nodes are stable (i.e. they empty infinitely often), then this ensures that the primary's unsuccessful packets which are successfully decoded by the secondary nodes eventually reach the primary destination.

The average arrival rate to the jth secondary node's relaying queue is given by

$$\lambda_j^{ext} = \left(\frac{\lambda_P}{\mu_P}\right)(1-\mu_P^{\max})\,P_c^{(j)}. \tag{3.62}$$

The average service rate of the jth secondary relaying queue is given by $\left(1-\frac{\lambda_P}{\mu_P}\right)P_s^{(j)}$, where

$$P_s^{(j)} = \sum_{T\subseteq\{1,2,..,N\}\setminus\{j\}}\left[\prod_{k\in T}P_c^{(k)}\prod_{k\in\{1,2,...,N\}\setminus\{T\bigcup j\}}\left(1-P_c^{(k)}\right)\right]P_j\left(T\right), \tag{3.63}$$

$$P_j\left(T\right) = \Pr\left[\frac{\sum_{i\in\{T\bigcup j\}}P_i|h_{S_iD_P}|^2 r_{S_iD_P}^{-\alpha}}{N_0} > \beta_P\right]. \tag{3.64}$$

For the stability of all the secondary nodes' relaying queues, we must have for all $j\in\{1,2,...,N\}$

$$\left(\frac{\lambda_P}{\mu_P}\right)(1-\mu_P^{\max})\,P_c^{(j)} < \left(1-\frac{\lambda_P}{\mu_P}\right)P_s^{(j)}, \tag{3.65}$$

which is equivalent to

$$\lambda_P < \min_{1\le j\le N}\left\{\frac{\mu_P P_s^{(j)}}{P_s^{(j)}+(1-\mu_P^{\max})\,P_c^{(j)}}\right\}. \tag{3.66}$$

Lemma 3.3. *$P_s^{(j)}$ as given in* Eq. (3.63) *is monotone increasing with N and converges to* 1 *as $N\to\infty$.*

Proof: The event $\left\{\frac{\sum_{i\in\{T\bigcup j\}}P_i|h_{S_iD_P}|^2 r_{S_iD_P}^{-\alpha}}{N_0} > \beta_P\right\}$ implies the event

$\left\{\frac{\sum_{i\in\{1,2,\ldots,N\}} P_i|h_{S_iD_P}|^2 r_{S_iD_P}^{-\alpha}}{N_0} > \beta_P\right\}$, hence

$$\Pr\left[\frac{\sum_{i\in\{T\bigcup j\}} P_i|h_{S_iD_P}|^2 r_{S_iD_P}^{-\alpha}}{N_0} > \beta_P\right] \leq \Pr\left[\frac{\sum_{i\in\{1,2,\ldots,N\}} P_i|h_{S_iD_P}|^2 r_{S_iD_P}^{-\alpha}}{N_0} > \beta_P\right]. \tag{3.67}$$

Using that $\sum_{T\subseteq\{1,2,..,N\}\backslash\{j\}}\left[\prod_{k\in T} P_c^{(k)} \prod_{k\in\{1,2,\ldots,N\}\backslash\{T\bigcup j\}}\left(1-P_c^{(k)}\right)\right] = 1$, we get that $P_s^{(j)}$ is upper bounded by $\Pr\left[\frac{\sum_{i\in\{1,2,\ldots,N\}} P_i|h_{S_iD_P}|^2 r_{S_iD_P}^{-\alpha}}{N_0} > \beta_P\right]$.

By Lemma 3.2, this bound converges to 1 as $N\to\infty$. We need to show that $P_s^{(j)}$ is monotone increasing with N and hence must converge to 1 as $N\to\infty$ by the monotone convergence theorem.

To show monotonicity, consider

$$P_s^{(j)}(N) = \sum_{T\subseteq\mathcal{S}\backslash\{j\}}\left[\prod_{k\in T} P_c^{(k)} \prod_{k\in\mathcal{S}\backslash\{T\bigcup j\}}\left(1-P_c^{(k)}\right)\right] P_j(T),$$

$$P_s^{(j)}(N+1) = \sum_{T\subseteq\mathcal{S}\bigcup\{N+1\}\backslash\{j\}}\left[\prod_{k\in T} P_c^{(k)} \prod_{k\in\mathcal{S}\bigcup\{N+1\}\backslash\{T\bigcup j\}}\left(1-P_c^{(k)}\right)\right] P_j(T). \tag{3.68}$$

The summation in $P_s^{(j)}(N+1)$ has twice as many terms as the summation in $P_s^{(j)}(N)$. Specifically, each set in $P_s^{(j)}(N)$ exists in $P_s^{(j)}(N+1)$ as well as the same set union the set $\{N+1\}$.

Let $\mathcal{M}$ be the set of all sets in $P_s^{(j)}(N)$, then:

$$P_s^{(j)}(N) = \sum_{m\in\mathcal{M}}\left[\prod_{k\in m} P_c^{(k)} \prod_{k\in\{1,2,\ldots,N\}\backslash\{m\bigcup j\}}\left(1-P_c^{(k)}\right)\right] P_j(m). \tag{3.69}$$

$$P_s^{(j)}(N+1) = \sum_{m\subseteq\mathcal{M}}\left[\prod_{k\in m} P_c^{(k)} \prod_{k\in\{1,2,\ldots,N+1\}\backslash\{m\bigcup j\}}\left(1-P_c^{(k)}\right)\right] P_j(m)$$
$$+\sum_{m\subseteq\mathcal{M}}\left[\prod_{k\in\{m\bigcup\{N+1\}\}} P_c^{(k)} \prod_{k\in\{1,2,\ldots,N\}\backslash\{m\bigcup j\}}\left(1-P_c^{(k)}\right)\right] P_j\left(m\bigcup\{N+1\}\right). \tag{3.70}$$

By using that $P_j\left(m\bigcup\{N+1\}\right) > P_j(m)$ since

$$\Pr\left[\frac{\sum_{i\in\{m\bigcup j\bigcup\{N+1\}\}} P_i|h_{S_iD_P}|^2 r_{S_iD_P}^{-\alpha}}{N_0} > \beta_P\right] > \Pr\left[\frac{\sum_{i\in\{m\bigcup j\}} P_i|h_{S_iD_P}|^2 r_{S_iD_P}^{-\alpha}}{N_0} > \beta_P\right], \tag{3.71}$$

we get

$$\begin{aligned} P_s^{(j)}(N+1) \geq \sum_{m\subseteq\mathcal{M}} \Bigg[& P_c^{(N+1)} \prod_{k\in m} P_c^{(k)} \prod_{k\in\{1,2,\ldots,N\}\setminus\{m\bigcup j\}} \left(1-P_c^{(k)}\right) + \\ & \prod_{k\in m} P_c^{(k)} \prod_{k\in\{1,2,\ldots,N\}\setminus\{m\bigcup j\}} \left(1-P_c^{(k)}\right)\left(1-P_c^{(N+1)}\right) \Bigg] P_j(m) \\ = \sum_{m\subseteq\mathcal{M}} & \left[\prod_{k\in m} P_c^{(k)} \prod_{k\in\{1,2,\ldots,N\}\setminus\{m\bigcup j\}} \left(1-P_c^{(k)}\right)\right] P_j(m) = P_s^{(j)}(N). \end{aligned} \tag{3.72}$$

■

Hence the sequence $P_s^{(j)}(N)$ is monotone increasing in N and upper bounded by 1, hence, converges to 1 as $N\to\infty$ for all $j\in\{1,2,...,N\}$.

If N is sufficiently large ($P_s^{(j)}\to 1$), the stability condition of the primary node in case of relaying given by Eq. (3.66) can be approximated by

$$\lambda_P < \frac{\mu_P}{1+(1-\mu_P^{\max})\max_{1\leq j\leq N}\{P_c^{(j)}\}}. \tag{3.73}$$

The maximum stable throughput rate in Eq. (3.73) is larger than $\mu_P^{\max}$ only if $\mu_P^{\max} < \frac{P_c^{(N)}}{\max_{1\leq j\leq N}\{P_c^{(j)}\}}$.

Finally, by noting that $P_c^{(N)} = 1-\prod_{k\in\{1,2,\ldots,N\}}\left(1-P_c^{(k)}\right) \geq \max_{1\leq j\leq N}\{P_c^{(j)}\}$, this condition is always satisfied for sufficiently large N and hence, relaying always improves the stable throughput of the primary for sufficiently large N.

3.8.3 Proof of Proposition 3.2

We proceed by showing that P_s is strictly increasing with N.

$$\begin{aligned}\psi(N) = P_s &= \sum_{k=0}^{N-1} \binom{N-1}{k} P_c^k (1-P_c)^{N-1-k} \Pr\left[\sum_{i=1}^{k+1} |h_i|^2 > c\right] \\ &= \sum_{k=0}^{N-1} \binom{N-1}{k} P_c^k (1-P_c)^{N-1-k} \phi(k). \end{aligned} \tag{3.74}$$

$$\begin{aligned}\psi(N+1)-\psi(N) &= \sum_{k=0}^{N} \binom{N}{k} P_c^k (1-P_c)^{N-k} \phi(k) - \sum_{k=0}^{N-1} \binom{N-1}{k} P_c^k (1-P_c)^{N-1-k} \phi(k) \\ &= P_c^N \phi(N) + \sum_{k=0}^{N-1} \binom{N-1}{k} P_c^k (1-P_c)^{N-1-k} \left[\frac{k-NP_c}{N-k}\right] \phi(k) \\ &= P_c^N \phi(N) + \sum_{k=0}^{\lfloor NP_c \rfloor} \binom{N-1}{k} P_c^k (1-P_c)^{N-1-k} \left[\frac{k-NP_c}{N-k}\right] \phi(k) \\ &\quad + \sum_{k=\lceil NP_c \rceil}^{N-1} \binom{N-1}{k} P_c^k (1-P_c)^{N-1-k} \left[\frac{k-NP_c}{N-k}\right] \phi(k) \\ &> P_c^N \phi(N) + \sum_{k=0}^{N-1} \binom{N-1}{k} P_c^k (1-P_c)^{N-1-k} \left[\frac{k-NP_c}{N-k}\right] \phi(\lceil NP_c \rceil) \\ &= P_c^N (\phi(N) - \phi(\lceil NP_c \rceil)) > 0, \end{aligned} \tag{3.75}$$

where we have used the fact that $\phi(N) > \phi(N-1) > ... > \phi(1)$ which follows by Lemma 3.2. Hence, P_s is a monotone increasing sequence with supremum equal to one and hence converges to one by the monotone convergence theorem.

Chapter 4: Enhanced Access Schemes for Cognitive Networks

4.1 Introduction

In this part, we propose and analyze low complexity access schemes at the SU that go beyond the traditional OSA. The proposed schemes [69] only require the knowledge of the statistics of the channels as well as the average arrival rate to the PU queue, which can be obtained through channel measurements. These schemes are capable of efficiently mitigating the negative effects of sensing errors occurring at the PHY layer. The analysis follows a cross-layer (PHY/MAC) approach taking into account the bursty nature of the traffic and the exact queue dynamics of both the PU and SU. Specifically, two low complexity MAC layer schemes are considered. In the first scheme, the SU accesses the channel at all slots with fixed probability p^* without sensing (sensing duration is exploited for data transmission), while in the second scheme, the SU accesses the channel when sensed to be idle (or busy) with fixed probabilities p_1^* (or p_2^*). Clearly, the first scheme has the advantage of simple implementation as well as offering the SU more time duration for data transmission but less PU protection since no sensing takes place. The second scheme offers some PU protection via sensing but less data transmission duration. We first study these schemes subject to a stability constraint on the PU queue and then

subject to an average delay constraint on the PU traffic. The analysis shows that if the receivers can successfully decode packets in presence of interference with high probability, then the first scheme is preferred for the SU since the sensing duration can be exploited for SU data transmission and hence higher SU throughput while the interference on the PU is handled by optimizing p^* and through the receiver decoding capability. However, if transmissions are likely to fail in presence of interference, sensing is crucial for PU protection and the second scheme outperforms the first. Therefore, in cognitive radio networks, the use of sophisticated receivers that can handle interference alleviates the need for complicated SU transmitters with strong sensing performance which is sometimes preferred in practice. For instance, if the secondary network is an uplink, the base station (receiver) can handle more complexity than the user equipments (transmitters) and thus the proposed access schemes are beneficial since sensing can be avoided at the expense of more complex receiver structure.

Effects of sensing errors on the network-layer performance for cognitive systems with bursty arrivals have been quantified in works such as [54, 61]. A simple enhanced cooperative MAC layer scheme for cognitive multiple access going beyond the traditional OSA has been proposed under perfect sensing in [56] but the analysis was restricted to a pure network-layer analysis without explicitly taking into account the PHY-layer parameters as done in our cross-layer (PHY/MAC) approach. Access schemes based on soft sensing have been analyzed in [70] and [71] with and without exploiting the primary feedback channel, respectively. However, the analysis was restricted to the simplified case of collision channels with no MPR capability which

is necessary to draw the main conclusion of this chapter.

The chapter is organized as follows. In Section 4.2, we describe the system model and the different access schemes considered in this part. In Section 4.3, we derive the stable throughput regions of the different access schemes in consideration, while we extend the problem to the case with an average delay constraint at the PU in Section 4.4. In Section 4.5, we present the numerical results and in Section 4.6, we conclude the chapter.

4.2 System Model

We consider a simple network consisting of one primary and one secondary link as shown in Fig. 4.1. Each user has a queue of infinite capacity to store incoming packets. Time is slotted with slot duration equal to T (seconds). For the PU, the slot duration is completely dedicated for data transmission while for the SU, the first τ seconds are dedicated for sensing and the remaining $T - \tau$ are for data transmission. The sensing duration τ is assumed to be fixed to yield some desired detection probability $P_d(\tau)$ and false alarm probability $P_f(\tau)$. In this part, we focus on the practical case where $P_d(\tau) >> P_f(\tau)$.

Packets arrive to each queue according to a stationary Bernoulli process with an average arrival rate of λ_i (packets/slot), $i \in \{1, 2\}$, where $i = 1$ designates the primary user and $i = 2$ the secondary. Each packet consists of B bits and hence the arrivals occur with average rates of $\lambda_i \frac{B}{T}$ (bits/sec), $i \in \{1, 2\}$. The primary and secondary nodes transmit at rates of R_p and R_s (bits/sec), respectively. In particular, these

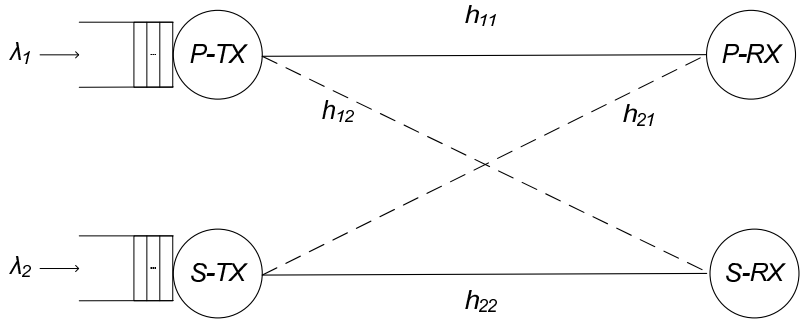

Figure 4.1: System Model

rates are chosen to be $R_p = \frac{B}{T}$ and $R_s = \frac{B}{T-\tau}$ so as to ensure one packet transmission per slot. Node i transmits at power P_i, $i \in \{1, 2\}$. Let $C_p^{(1)}$ (or $C_p^{(1,2)}$) be the capacity of the PU transmitter-receiver link in the absence (or presence) of SU interference. Similarly, let $C_s^{(2)}$ (or $C_s^{(1,2)}$) be the capacity of the SU transmitter-receiver link in the absence (or presence) of PU interference. Unlike the oversimplified collision model [72], we consider the possibility of simultaneous successful transmissions. We denote by $q_{i|\mathcal{A}}$ the probability that the ith node successfully decodes the packet when nodes in the set $\mathcal{A}$ transmit simultaneously. For a slow fading environment, which is what we focus on in this work, the success probabilities corresponding to transmission rates R_p and R_s are given by $q_{1|1} = \Pr\left[R_p < C_p^{(1)}\right]$, $q_{2|2} = \Pr\left[R_s < C_s^{(2)}\right]$, $q_{1|\{1,2\}} = \Pr\left[R_p < C_p^{(1,2)}\right]$ and $q_{2|\{1,2\}} = \Pr\left[R_s < C_s^{(1,2)}\right]$. Clearly, because of the effect of the interference, we have $q_{1|\{1,2\}} < q_{1|1}$ and $q_{2|\{1,2\}} < q_{2|2}$. We define the quantities $\gamma = 1 - \frac{q_{1|\{1,2\}}}{q_{1|1}} > 0$ and $\delta = 1 - \frac{q_{2|\{1,2\}}}{q_{2|2}} > 0$ to be used throughout this chapter.

After each transmission, the transmitter receives an instantaneous and error free Acknowledgment/Non-Acknowledgment (ACK/NACK) message from its receiver indicating whether the packet was successfully received or not. Failed transmissions

are retransmitted in subsequent slots.

We adopt the definition of queueing stability as previously defined in Section 2.3.

We study low complexity MAC layer access schemes at the SU that are capable of mitigating the effects of sensing errors occurring at the PHY-layer. These schemes leverage the knowledge of the channel statistics as well as the average arrival rate to the PU λ_1[1]. We compare these schemes through network-level performance metrics, namely, the stable throughput and the average delay, which, as we shall see, embodies a new perspective of the sensing-throughput tradeoff problem. These schemes trade performance against complexity. The goal of the SU is to maximize its stable throughput subject to a constraint on the performance of the PU (e.g., stability of the PU queue in Section 4.3, and average delay in Section 4.4).

We now present the different access schemes to be used throughout the chapter.

4.2.1 Fixed Access Scheme (FA)

In this scheme, the SU accesses the channel at all slots with the same fixed probability p without sensing (i.e., $\tau = 0$). The value of p is optimized to yield maximum stable throughput at the SU while satisfying a performance constraint at the PU (queueing stability in Section 4.3 and average delay in Section 4.4). This scheme is appropriate if the sensing process is difficult to implement or if sensing is to be avoided in order to allow a longer interval of data transmission and/or to save

[1] λ_1 can be known to the SU either by cooperation between the PU and the SU or by using a learning algorithm such as Baum-Welch algorithm to learn the PU traffic dynamics modeled as a Hidden Markov process.

processing energy of the sensed samples at the SU [28]. It should be noted that by avoiding sensing, the SU transmission rate R_s is equal to $\frac{B}{T}$ which leads to higher success probabilities at the SU ($\Pr[R_s < C_s^{(1)}]$ and $\Pr[R_s < C_s^{(1,2)}]$) compared to a scheme with sensing where the SU transmission rate is $R_s = \frac{B}{T-\tau}$. It will be shown later that this scheme is preferred if the channels are strong enough to support simultaneous transmissions. Under some conditions, this scheme can outperform other schemes with sensing.

4.2.2 Randomized Access Scheme (RA)

This is a scheme with sensing, where the secondary node upon detecting the primary to be idle (or busy), decides to transmit with probability p_1 (or p_2) respectively. These probabilities are optimized to yield maximum stable throughput at the SU. The advantage of this scheme is to exploit the knowledge of the channel statistics in optimizing p_1^* and p_2^*. This represents a simple MAC layer enhancement of performance for the same PHY-layer sensing errors.

For sake of comparison, in the numerical results in Section 4.5, we also consider the cases of traditional OSA where $p_1^* = 1$ and $p_2^* = 0$ with and without perfect sensing. We will see that using a good sensing technique is beneficial if simultaneous PU and SU transmissions are likely to fail; otherwise, a higher misdetection probability might be beneficial for the SU where misdetections of the PU offer more transmission opportunities while the interference on the PU can be handled through the capability of successfully decoding transmissions in presence of interference at

the receivers.

4.3 Stable Throughput Region Calculation

If sensing is perfect, the SU never interferes with the PU and the two queues are decoupled. However, in the presence of sensing errors, the service process of one queue depends on the state (empty or not) of the other. The two queues are hence "interacting". To circumvent this difficulty, we use the idea of stochastic dominance as in Chapter 2.

We proceed by defining the dominant systems for the randomized access case. The case of fixed access can be derived as a special case of it.

-First Dominant System (S_1)

This system is identical to the original system in terms of arrivals, transmissions and outcomes, except that, whenever the SU queue empties, the SU keeps transmitting dummy packets consisting of B bits each. Dummy packets do not contribute to the throughput, but cause interference to the PU. Hence, in S_1, the PU queue has a fixed service rate regardless of the state of the SU queue (no queue interaction).

Let $\mathcal{E}_t$ or $\mathcal{F}_t$ be the events of misdetection and false alarm at the SU transmitter in slot t when the PU is busy or idle respectively. The events $\mathcal{A}_t$ and $\mathcal{T}_t$ represent the events that the secondary node transmits in slot t, given that the slot is detected as idle or busy respectively. The service process of the primary node in (bits/sec)

is given by

$$Y_1^t = R_p \left[\mathbb{1}\Big\{ \mathcal{E}_t \bigcap \mathcal{A}_t \bigcap \overline{\mathcal{O}_{S_1D_1,I}} \Big\} + \mathbb{1}\Big\{ \overline{\mathcal{E}_t} \bigcap \mathcal{T}_t \bigcap \overline{\mathcal{O}_{S_1D_1,I}} \Big\} + \mathbb{1}\Big\{ \mathcal{E}_t \bigcap \overline{\mathcal{A}_t} \bigcap \overline{\mathcal{O}_{S_1D_1,NI}} \Big\} + \mathbb{1}\Big\{ \overline{\mathcal{E}_t} \bigcap \overline{\mathcal{T}_t} \bigcap \overline{\mathcal{O}_{S_1D_1,NI}} \Big\} \right], \tag{4.1}$$

where $\mathcal{O}_{S_1D_1,I}$ and $\mathcal{O}_{S_1D_1,NI}$ are the outage events on the PU transmitter-receiver channel in the presence or in the absence of SU interference respectively, while $\mathbb{1}\{\bullet\}$ is the indicator function and $\overline{\mathcal{U}}$ denotes the complement of the event $\mathcal{U}$. For a value x, we subsequently denote $(1-x)$ by $\bar{x}$.

The process $\{Y_1^t\}$ is stationary since it is a function of stationary events, and also independent of the arrival process to the PU $\{X_1^t\}$. Hence, by Loynes' theorem, we can get the stability condition of the primary queue in the first dominant system S_1 as $\lambda_1 \frac{B}{T} < \mathbb{E}[Y_1^t]$ which is equivalent to

$$\lambda_1 < \mu_1 = \left[q_{1|1} \left(P_d \overline{p_2} + \overline{P_d} \overline{p_1} \right) + q_{1|\{1,2\}} \left(p_2 P_d + p_1 \overline{P_d} \right) \right].$$

Equivalently

$$0 \leq \lambda_1 < \mu_1 = q_{1|1} \left[1 - \gamma \left(p_1 \left(1 - P_d \right) + p_2 P_d \right) \right]. \tag{4.2}$$

The service process in (bits/sec) of the secondary queue depends on the state of the primary queue and is given by

$$Y_2^t = R_s \Big(1 - \frac{\tau}{T}\Big) \Big[\mathbb{1}\Big\{ \{Q_1^t \neq 0\} \bigcap \mathcal{E}_t \bigcap \mathcal{A}_t \bigcap \overline{\mathcal{O}_{S_2D_2,I}} \Big\} + \mathbb{1}\Big\{ \{Q_1^t \neq 0\} \bigcap \overline{\mathcal{E}_t} \bigcap \mathcal{T}_t \bigcap \overline{\mathcal{O}_{S_2D_2,I}} \Big\} + \mathbb{1}\left\{ \{Q_1^t = 0\} \bigcap \mathcal{F}_t \bigcap \mathcal{T}_t \bigcap \overline{\mathcal{O}_{S_2D_2,NI}} \right\} + \mathbb{1}\left\{ \{Q_1^t = 0\} \bigcap \overline{\mathcal{F}_t} \bigcap \mathcal{A}_t \bigcap \overline{\mathcal{O}_{S_2D_2,NI}} \right\} \Big], \tag{4.3}$$

where $\mathcal{O}_{S_2D_2,I}$ and $\mathcal{O}_{S_2D_2,NI}$ are, respectively, the events of outage on the channel between the SU transmitter and receiver in the presence or in the absence of PU

interference. Using the result that a stable queue gives rise to stationary idle slots (see [41,48]), it follows that the processes $\mathbb{1}\{Q_1^t = 0\}$ and $\mathbb{1}\{Q_1^t \neq 0\}$ are stationary. Hence, the process Y_2^t is also stationary. Using Loynes' theorem, we obtain the stability condition of the secondary queue in the first dominant system as $\lambda_2 \frac{B}{T} < \mathbb{E}[Y_2^t]$ which is equivalent to

$$0 \leq \lambda_2 < \mu_2 = \left[\left(1 - \frac{\lambda_1}{\mu_1}\right) q_{2|2} \left[p_2 P_f + p_1(1 - P_f)\right] + \left(\frac{\lambda_1}{\mu_1}\right) q_{2|\{1,2\}} \left[p_2 P_d + (1 - P_d) p_1\right]\right]. \tag{4.4}$$

Equivalently

$$0 \leq \lambda_2 < \mu_2 = q_{2|2}[p_2 P_f + p_1(1 - P_f)]\left[1 - \left(\frac{\lambda_1}{\mu_1}\right)\left(1 - \frac{(p_2 P_d + p_1(1 - P_d)) q_{2|\{1,2\}}}{(p_2 P_f + p_1(1 - P_f)) q_{2|2}}\right)\right]. \tag{4.5}$$

-Second Dominant System S_2

In this system, the PU transmits dummy packets whenever it empties, while otherwise it is identical to the original system. The stability conditions for the queues in S_2 are given by

$$0 \leq \lambda_2 < \mu_2 = q_{2|\{1,2\}} \left[p_1(1 - P_d) + p_2 P_d\right], \tag{4.6}$$

$$0 \leq \lambda_1 < \mu_1 = q_{1|1} \left[1 - \gamma \left(\frac{\lambda_2}{\mu_2}\right) \left(p_2 P_d + p_1(1 - P_d)\right)\right]. \tag{4.7}$$

The case of traditional OSA can be directly obtained by setting $p_1 = 1, p_2 = 0$; while the case of fixed access can be obtained by setting $p_1 = p_2 = p$, and $\tau = 0$. Note that in this case $P_d(\tau = 0) = 0$ and $P_f(\tau = 0) = 1$.

Equations (4.2), (4.5), (4.6) and (4.7) give the stability region for fixed p_1 and p_2 that we denote by $\mathcal{S}(p_1, p_2)$. We seek to obtain the stability region for any (p_1, p_2),

that is, $\mathcal{S} = \bigcup_{(p_1,p_2)\in[0,1]^2} \mathcal{S}(p_1,p_2)$. This can be done by computing the equation of the boundary of the region and maximizing it over $(p_1,p_2) \in [0,1]^2$ for each of the two dominant systems S_1 and S_2. Then, taking the union of both optimized boundaries yields the "envelope" of the regions. It can be shown that S_2 always yields a region that is a subset of that given by S_1, hence we focus on obtaining the union over $(p_1,p_2) \in [0,1]^2$ of the stability region given by S_1.

4.3.1 Stability Region of the Fixed Access Scheme

Substituting $p_1 = p_2 = p$, $P_d = 0$ and $P_f = 1$ in Eqs. (4.2) and (4.5), yields the boundary of the stability region by solving the constrained optimization problem

$$\max_{0\le p\le 1} y = \widetilde{q_{2|2}} p \left[1 - \frac{\lambda_1 \widetilde{\delta}}{\widetilde{q_{1|1}}(1-\widetilde{\gamma}p)}\right],$$
$$\text{s.t. } 0 \le \lambda_1 < \widetilde{q_{1|1}}(1-\widetilde{\gamma}p). \tag{4.8}$$

Note that the tilde~notation is used to emphasize that the success probabilities of this scheme are larger than the success probabilities of the schemes with sensing due to the difference in transmission rates as discussed in Section 4.2.1.

By using the same approach as in proving Theorem 2.1, we obtain the optimum transmission probability p^* as

$$p^* = \begin{cases} 1, & 0 \le \lambda_1 \le \min\left\{\widetilde{q_{1|\{1,2\}}}, \frac{\widetilde{q_{1|1}}(1-\widetilde{\gamma})^2}{\widetilde{\delta}}\right\}, \\ \frac{1}{\widetilde{\gamma}}\left(1 - \sqrt{\frac{\lambda_1\widetilde{\delta}}{\widetilde{q_{1|1}}}}\right), & \frac{\widetilde{q_{1|1}}(1-\widetilde{\gamma})^2}{\widetilde{\delta}} \le \lambda_1 \le \widetilde{q_{1|1}}\widetilde{\delta}, \\ \frac{1}{\widetilde{\gamma}}\left(1 - \frac{\lambda_1}{\widetilde{q_{1|1}}}\right), & \max\left\{\widetilde{\delta}\widetilde{q_{1|1}}, \widetilde{q_{1|\{1,2\}}}\right\} \le \lambda_1 \le \widetilde{q_{1|1}}. \end{cases} \tag{4.9}$$

Note that the middle expression for p^* is valid only if $\widetilde{\gamma}+\widetilde{\delta} \geq 1$ to ensure that $\frac{\widetilde{q_{1|1}}(1-\widetilde{\gamma})^2}{\widetilde{\delta}} \leq \widetilde{q_{1|1}}\widetilde{\delta}$. Substituting into the objective function in Eq. (4.8), we obtain the boundary of the stability region. We differentiate between two cases.

4.3.1.1 Case of $\widetilde{\gamma}+\widetilde{\delta} > 1$

In this case, $\widetilde{q_{1|\{1,2\}}} \geq \frac{\widetilde{q_{1|1}}(1-\widetilde{\gamma})^2}{\widetilde{\delta}} \Leftrightarrow \widetilde{\delta} \leq (1-\widetilde{\gamma})$. The stability region consists of three parts and is bounded by a convex curve. The boundary of the stability region is given by

$$\lambda_2 = \begin{cases} \widetilde{q_{2|2}}\left[1-\frac{\lambda_1\widetilde{\delta}}{\widetilde{q_{1|1}}(1-\widetilde{\gamma})}\right] & \text{, for } 0 \leq \lambda_1 \leq \frac{\widetilde{q_{1|1}}(1-\widetilde{\gamma})^2}{\widetilde{\delta}}, \\ \frac{\widetilde{q_{2|2}}}{\widetilde{\gamma}}\left(1-\sqrt{\frac{\lambda_1\widetilde{\delta}}{\widetilde{q_{1|1}}}}\right)^2, & \text{for } \frac{\widetilde{q_{1|1}}(1-\widetilde{\gamma})^2}{\widetilde{\delta}} \leq \lambda_1 \leq \widetilde{q_{1|1}}\widetilde{\delta}, \\ \frac{\widetilde{q_{2|2}}(1-\widetilde{\delta})}{\widetilde{\gamma}}\left(1-\frac{\lambda_1}{\widetilde{q_{1|1}}}\right), & \text{for } \widetilde{\delta}\widetilde{q_{1|1}} \leq \lambda_1 \leq \widetilde{q_{1|1}}. \end{cases} \tag{4.10}$$

4.3.1.2 Case of $\widetilde{\gamma}+\widetilde{\delta} \leq 1$

In this case, the stability region consists of two linear parts. The boundary of the stability region is given by

$$\lambda_2 = \begin{cases} \widetilde{q_{2|2}}\left[1-\frac{\lambda_1\widetilde{\delta}}{\widetilde{q_{1|1}}(1-\widetilde{\gamma})}\right] & \text{, for } 0 \leq \lambda_1 \leq \widetilde{q_{1|1}}(1-\widetilde{\gamma}), \\ \frac{\widetilde{q_{2|2}}(1-\widetilde{\delta})}{\widetilde{\gamma}}\left(1-\frac{\lambda_1}{\widetilde{q_{1|1}}}\right), & \text{for } \widetilde{q_{1|1}}(1-\widetilde{\gamma}) \leq \lambda_1 \leq \widetilde{q_{1|1}}. \end{cases} \tag{4.11}$$

4.3.2 Stability Region of the Randomized Access Scheme

The stability region in this case can be found by solving the optimization problem in which we maximize μ_2 in Eq. (4.5) subject to the constraint on λ_1 in

Eq. (4.2). This can be written as

$$\max_{(p_1,p_2)\in[0,1]^2} y = q_{2|2}\left[p_2P_f + p_1(1-P_f)\right]\left[1-\left(\frac{\lambda_1}{\mu_1}\right)\left(1-\frac{(p_2P_d+p_1(1-P_d))q_{2|\{1,2\}}}{(p_2P_f+p_1(1-P_f))q_{2|2}}\right)\right]$$

$$\text{s.t. } 0 \le \lambda_1 \le \mu_1 = q_{1|1}\left[1-\gamma\left(p_1\overline{P_d}+p_2P_d\right)\right]. \tag{4.12}$$

The solution of this problem depends on whether $\gamma+\delta \le 1$ or $\gamma+\delta > 1$, and is given by

4.3.2.1 Case of $\gamma+\delta \le 1 \Leftrightarrow \frac{q_{1|\{1,2\}}}{q_{1|1}} + \frac{q_{2|\{1,2\}}}{q_{2|2}} \ge 1$

$$\begin{aligned} p_1^* &= \max\left[0, \min\left\{1, \frac{1}{\gamma(1-P_d)}\left(1-\frac{\lambda_1}{q_{1|1}}\right)\right\}\right], \\ p_2^* &= \max\left[0, \min\left\{1, \frac{1}{P_d}\left[\frac{1}{\gamma}\left(1-\frac{\lambda_1}{q_{1|1}}\right)-(1-P_d)\right]\right\}\right]. \end{aligned} \tag{4.13}$$

The boundary of the stability region is given by

$$\lambda_2 = \begin{cases} q_{2|2}\left[1-\frac{\lambda_1\delta}{q_{1|1}(1-\gamma)}\right], & \text{for } 0 \le \lambda_1 \le q_{1|1}(1-\gamma), \\ \frac{q_{2|2}(1-\delta)}{\gamma}\left(1-\frac{\lambda_1}{q_{1|1}}\right), & \text{for } q_{1|1}(1-\gamma) \le \lambda_1 \le q_{1|1}. \end{cases} \tag{4.14}$$

4.3.2.2 Case of $\gamma+\delta > 1 \Leftrightarrow \frac{q_{1|\{1,2\}}}{q_{1|1}} + \frac{q_{2|\{1,2\}}}{q_{2|2}} < 1$

In this case p_1^* and p_2^* are given by

$$p_1^* = \max\left[0, \min\left\{1, \frac{1}{\gamma\overline{P_d}}\left(1-\frac{\lambda_1}{q_{1|1}}\right), \frac{1}{\gamma\overline{P_d}}\left(1-\sqrt{\frac{\lambda_1}{q_{1|1}}\left(1-\frac{\overline{P_d}(1-\delta)}{1-P_f}\right)}\right)\right\}\right], \tag{4.15}$$

$$p_2^* = \max\left[0, \min\left\{1, \frac{1}{P_d}\left[\frac{1}{\gamma}\left(1-\frac{\lambda_1}{q_{1|1}}\right)-\overline{P_d}\right], \frac{1}{P_d}\left[\frac{1}{\gamma}\left(1-\sqrt{\frac{\lambda_1 C}{q_{1|1}P_f}}\right)-\overline{P_d}\right]\right\}\right], \tag{4.16}$$

where $C = P_f - (1-\delta)P_d + \gamma(P_d - P_f)$.

The boundary of the stability region can then be obtained by substituting p_1^* and p_2^* in μ_2 from Eq. (4.5) with $0 \leq \lambda_1 \leq q_{1|1}$.

Proof: Refer to Section 4.7. ■

In the first case where $\gamma + \delta \leq 1 \Leftrightarrow \frac{q_{1|\{1,2\}}}{q_{1|1}} + \frac{q_{2|\{1,2\}}}{q_{2|2}} \geq 1$, the channels can support simultaneous transmissions with high success probability. As will be shown in Section 4.5, in this case, the boundary of the stability region is a polyhedron that might contain all or part of the stability region of the perfect sensing. Hence, in this case, sensing errors, under optimized randomized access might lead to stable rates that are not achievable with traditional OSA with perfect sensing. Moreover, the stability region of the fixed access is a superset of that of the randomized access. This means that, whenever this condition is satisfied, no sensing with optimal fixed access is better than sensing with optimal randomized access. This is the result of avoiding the time overhead needed for sensing, thereby leading to a higher success probabilities and more data transmission duration for the SU. We also note from Eq. (4.14) that, with randomized access, any sensor (i.e. any P_f and P_d) leads to the same stability region since optimal randomized access mitigates the effects of sensing errors. On the other hand, if $\gamma + \delta > 1$, the channels are not likely to support simultaneous transmissions successfully. In this case, the stability region is bounded by a convex curve. The stability region of the OSA with perfect sensing is a superset of that of the randomized access which strictly contains that of the fixed access (except for very small and very large values of λ_1). These observations are discussed further in Section 4.5.

4.4 Effect of Average Delay Constraint at the PU

So far, we considered the queueing stability problem, where the secondary node aims at maximizing its stable throughput while guaranteeing the stability of the primary queue. However, a stable queue may yet achieve long delay values if the average arrival rate is close to the maximum stable throughput rate of the queue. In this section, we revisit the problem with an emphasis on the average delay of the primary node, that is, we study the problem where the SU aims at maximizing its throughput while guaranteeing an average delay constraint at the PU.

The calculation of the average delay in interacting queues is known to be extremely thorny. In [73], the average delay of two asymmetric interacting queues in a random access setting with no multipacket reception capability was computed by mapping the problem to a Riemann-Hilbert boundary value problem. For the special case of two symmetric queues, where the arrival processes and the channels are statistically identical, the average delay has been found in [51] for a collision channel and in [43] for a channel with multipacket reception capability. However, the assumption of symmetric queues and channels is not appropriate in a cognitive setting. In order to bypass the difficulty in the cognitive radio setting, a fluid queue approximation was used in [74] to approximately characterize the delay behavior of the SUs. Alternatively, in order to exactly characterize the PU delay behavior, we consider the special case where the secondary queue is saturated, i.e, never empties, while the primary queue empties infinitely often. This represents a worst-case scenario and hence a lower bound on performance for the PU compared with a

non-saturated SU queue. The secondary node aims at finding p_1^* and p_2^* to maximize its throughput while guaranteeing an average delay constraint to the primary. The problem is formulated as a quasiconcave program which can be readily solved using known algorithms [75].

4.4.1 Problem Formulation

Since the arrival process to the primary queue is a Bernoulli process, the primary queue evolves as a Geo/Geo/1 queue. In every "busy" slot of the PU, the SU transmits with probability p_1 if it senses the channel to be idle, which occurs with probability $1 - P_d$, and with probability p_2 if it senses the channel to be busy which occurs with probability P_d. Let X be the time (in slots) required to serve one primary packet. Then, X is geometrically distributed with parameter $\mu_1 = q_{1|1}\left[(1-P_d)(1-p_1) + P_d(1-p_2)\right] + q_{1|\{1,2\}}\left[(1-P_d)p_1 + P_d p_2\right]$. In vector notation, we write $\mu_1 = q_{1|1}\left[1 - \mathbf{a^T x}\right] + q_{1|\{1,2\}}\mathbf{a^T x} = q_{1|1}\left[1 - \gamma \mathbf{a^T x}\right]$, where $\mathbf{a} = [1-P_d,\ P_d]^T$, $\mathbf{x} = [p_1,\ p_2]^T$ and $(\bullet)^T$ denotes vector transposition. The average delay at the primary is then given by [76]

$$\mathbb{E}\left[D\right] = \frac{1-\lambda_1}{\mu_1 - \lambda_1} = \frac{1-\lambda_1}{q_{1|1}\left[1 - \gamma \mathbf{a^T x}\right] - \lambda_1}, \tag{4.17}$$

where $\lambda_1 < \mu_1$ to ensure the stability of the primary queue.

The SU throughput in (packets/slot) is given by Eq. (4.5), which can be written in

vector notation as

$$\mu_2 = \left[\left(1-\frac{\lambda_1}{\mu_1}\right) q_{2|2}\left[p_2 P_f + p_1(1-P_f)\right] + \left(\frac{\lambda_1}{\mu_1}\right) q_{2|\{1,2\}}\left[p_2 P_d + (1-P_d)p_1\right]\right]$$
$$= \frac{q_{2|2}q_{1|1}\mathbf{b^T x} - q_{2|2}q_{1|1}\gamma \mathbf{x^T b a^T x} - q_{2|2}\lambda_1 \mathbf{b^T x} + q_{2|\{1,2\}}\lambda_1 \mathbf{a^T x}}{q_{1|1}\left[1-\gamma \mathbf{a^T x}\right]}, \tag{4.18}$$

where $\mathbf{b} = [1-P_f,\ P_f]^T$. The minimum average delay at the PU is given by $D_{\min} = \frac{1-\lambda_1}{q_{1|1}-\lambda_1}$. If the average delay constraint at the PU is D_0, the optimization problem can be written as

$$\max_{\mathbf{0}\preceq\mathbf{x}\preceq\mathbf{1}} \mu_2 = \frac{q_{2|2}q_{1|1}\mathbf{b^T x} - q_{2|2}q_{1|1}\gamma \mathbf{x^T b a^T x} - q_{2|2}\lambda_1 \mathbf{b^T x} + q_{2|\{1,2\}}\lambda_1 \mathbf{a^T x}}{q_{1|1}\left[1-\gamma \mathbf{a^T x}\right]}$$
$$\text{s.t. } \frac{1-\lambda_1}{q_{1|1}\left[1-\gamma \mathbf{a^T x}\right]-\lambda_1} \le D_0$$
$$\lambda_1 < q_{1|1}\left[1-\gamma \mathbf{a^T x}\right] \tag{4.19}$$

where $\preceq$ denotes componentwise $\le$ operation, and the second constraint is the stability constraint. Note that as $D_0 \to \infty$, the first constraint is always satisfied and only the stability constraint is of importance.

By rewriting the first constraint as $(1-\lambda_1) \le D_0\left(q_{1|1}\left[1-\gamma \mathbf{a^T x}\right]-\lambda_1\right)$, it is clear that the stability constraint is implicitly implied since $0 \le \lambda_1 < q_{1|1} < 1$, thus we only focus on the delay constraint. Note that the constraint in this form is affine and hence convex.

4.4.2 Solution of the Optimization Problem

First we denote the objective function in Eq. (4.19) by $f_0(\mathbf{x})$. The domain of $f_0(\mathbf{x})$ is the compact set $\mathbf{dom} f_0(\mathbf{x}) = [0,1]^2$ which is a convex set. In order to show that the problem in hand is a quasiconcave one, we use the following lemma.

Lemma 4.1. *The objective function in* Eq. (4.19) *is of the form* $f_0(\mathbf{x}) = \frac{p(\mathbf{x})}{q(\mathbf{x})}$ *where* $p(\mathbf{x})$ *is concave and* ≥ 0 *over* $\mathbf{dom} f_0(\mathbf{x})$ *while* $q(\mathbf{x})$ *is affine and* > 0 *and hence convex over* $\mathbf{dom} f_0(\mathbf{x})$.

Proof: We have $q(\mathbf{x}) = q_{1|1}\left[1 - \gamma \mathbf{a^T x}\right]$ which satisfies $0 < q_{1|1}(1-\gamma) \leq q(\mathbf{x}) \leq q_{1|1}$ and is hence positive. Since it is affine, it is also convex. As the secondary throughput, by definition, is non-negative, we have $p(\mathbf{x}) \geq 0$. To show concavity of $p(\mathbf{x})$, we show that its Hessian matrix is a negative semidefinite matrix over $\mathbf{dom} f_0(\mathbf{x})$. The Hessian matrix of $p(\mathbf{x})$ is given by

$$\begin{aligned} \nabla^2 p(\mathbf{x}) &= -\gamma q_{1|1} q_{2|2} \left(\mathbf{b a^T} + \mathbf{a b^T}\right) \\ &= -\gamma q_{1|1} q_{2|2} \begin{bmatrix} 2\overline{P_f}\ \overline{P_d} & P_f\overline{P_d} + P_d\overline{P_f} \\ P_f\overline{P_d} + P_d\overline{P_f} & 2P_fP_d \end{bmatrix}. \end{aligned} \tag{4.20}$$

It is straightforward to show that $\mathbf{x^T} \nabla^2 p(\mathbf{x}) \mathbf{x} \leq 0$ for all $\mathbf{x} \in \mathbf{dom} f_0(\mathbf{x})$; hence, the Hessian matrix is negative semidefinite over $\mathbf{dom} f_0(\mathbf{x})$ and hence $p(\mathbf{x})$ is concave over its domain. ■

We now prove the quasiconcavity of the problem.

Lemma 4.2. *The optimization problem given in* Eq. (4.19) *is a quasiconcave optimization problem.*

Proof: As discussed above, the constraint is affine and hence convex. Therefore, we only need to show that the objective function in Eq. (4.19) is quasiconcave [75]. For $\epsilon \in \mathbb{R}$, define S_ϵ to be the ϵ-superlevel set of $f_0(\mathbf{x})$ which is given by $S_\epsilon = \{\mathbf{x} \in \mathbf{dom} f_0(\mathbf{x}) \mid f_0(\mathbf{x}) \geq \epsilon\}$. Since $p(\mathbf{x}) \geq 0$ and $q(\mathbf{x}) > 0$,

in order to show quasiconcavity of $f_0(\mathbf{x})$, it suffices to show that S_ϵ are convex sets for all $\epsilon \in \mathbb{R}$ [75]. If $\epsilon < 0$, then by the non-negativity of $f_0(\mathbf{x})$, we have $S_\epsilon = \{\mathbf{x} \in \mathbf{dom} f_0(\mathbf{x}) \mid f_0(\mathbf{x}) \geq \epsilon\} = \mathbf{dom} f_0(\mathbf{x})$ which is a convex set. If $\epsilon \geq 0$, then $p(\mathbf{x}) - \epsilon q(\mathbf{x})$ is a concave function and hence, $S_\epsilon = \{\mathbf{x} \in \mathbf{dom} f_0(\mathbf{x}) \mid f_0(\mathbf{x}) \geq \epsilon\} = \{\mathbf{x} \in \mathbf{dom} f_0(\mathbf{x}) \mid p(\mathbf{x}) - \epsilon q(\mathbf{x}) \geq 0\}$ is a convex set since the superlevel sets of concave functions are convex. ■

The problem of maximizing a quasiconcave function over a convex set under convex constraints, as the problem at hand, can be efficiently solved by converting the problem into a set of convex feasibility problems and using the bisection method [75].

4.5 Discussion of the Results

The equations derived so far can be applied to almost any channel model. However, we need a specific model with formulas for the capacities $C_p^{(1)}$, $C_s^{(2)}$, $C_p^{(1,2)}$ and $C_s^{(1,2)}$ in order to compute the different success probabilities $q_{1|1}$, $q_{2|2}$, $q_{1|\{1,2\}}$ and $q_{2|\{1,2\}}$ in terms of the transmission powers, the channel fading statistics, and the sensing duration τ. Thermal noise w_i at the ith receiver ,$i \in \{1, 2\}$, is assumed to be additive complex Gaussian noise with distribution $w_i \sim \mathcal{CN}(0, N_0)$ and independent between the two receivers. The channel gain h_{ij} on the $(i-j)$th link has distribution $\mathcal{CN}(0, \sigma_{ij}^2)$ and the channel gains are independent between the different links. We assume narrowband transmissions on a bandwidth of W (Hz) such that the fading is flat. Fading is assumed fixed over the slot duration and is i.i.d. between slots.

We assume that the receivers have perfect channel state information (CSIR). Without coordination between the PU and the SU nodes, the PU (or SU) receiver does not know the codebook used by the SU (or PU) transmitter. Hence, we assume that the receivers treat the interference as noise.[2] The PU and SU transmitters use random Gaussian codebooks for their transmissions. Assuming that the number of bits per packet is large (i.e., large value of B), and assuming that the receivers treat the interference as noise, the capacities (in bits/sec) of the different links for the schemes with sensing (the RA and OSA schemes) can be approximated by (ignoring the error due to finite block length)

$$C_p^{(1)} = W \log_2 \left(1 + \frac{P_1|h_{11}|^2}{N_0 W}\right), \tag{4.21}$$

$$C_s^{(2)} = W \left(1 - \frac{\tau}{T}\right) \log_2 \left(1 + \frac{P_2|h_{22}|^2}{N_0 W}\right), \tag{4.22}$$

$$C_p^{(1,2)} = W \log_2 \left(1 + \frac{P_1|h_{11}|^2}{N_0 W + P_2|h_{21}|^2}\right), \tag{4.23}$$

$$C_s^{(1,2)} = W \left(1 - \frac{\tau}{T}\right) \log_2 \left(1 + \frac{P_2|h_{22}|^2}{N_0 W + P_1|h_{12}|^2}\right). \tag{4.24}$$

The corresponding capacities for the fixed access scheme can be directly obtained by substituting $\tau = 0$ in Eqs. (4.21)-(4.24). Note that the capacity $C_p^{(1,2)}$ is the capacity of the PU source-destination channel in the presence of SU interference. The first τ seconds during which the SU senses the channel are interference free while the remaining $T - \tau$ seconds incur SU interference. Since the sensing duration is typically small (i.e., $\tau << T$), we use the lower bound on the capacity $C_p^{(1,2)}$ as given in Eq. (4.23) (which corresponds to a lower bound on $q_{1|\{1,2\}}$) assuming that

[2]It has also been shown that if the codebooks are known but the interference is undecodable at the receivers, treating the interference as noise does not incur any loss in rate [77].

the interference occupies the whole slot. The corresponding success probabilities are given by

$$
\begin{aligned}
q_{1|1} &= \Pr\left[R_p < W\log_2\left(1+\frac{P_1|h_{11}|^2}{N_0W}\right)\right] = \exp\left(\frac{-C_1}{\sigma_{11}^2}\right),\\
q_{2|2} &= \Pr\left[R_s < W\left(1-\frac{\tau}{T}\right)\log_2\left(1+\frac{P_2|h_{22}|^2}{N_0W}\right)\right] = \exp\left(\frac{-C_2}{\sigma_{22}^2}\right),\\
q_{1|\{1,2\}} &= \Pr\left[R_p < W\log_2\left(1+\frac{P_1|h_{11}|^2}{N_0W+P_2|h_{21}|^2}\right)\right] = \exp\left(\frac{-C_1}{\sigma_{11}^2}\right)\left[\frac{\sigma_{11}^2}{\sigma_{11}^2+C_3\sigma_{21}^2}\right],\\
q_{2|\{1,2\}} &= \Pr\left[R_s < W\left(1-\frac{\tau}{T}\right)\log_2\left(1+\frac{P_2|h_{22}|^2}{N_0W+P_1|h_{12}|^2}\right)\right] = \exp\left(\frac{-C_2}{\sigma_{22}^2}\right)\left[\frac{\sigma_{22}^2}{\sigma_{22}^2+C_4\sigma_{12}^2}\right],
\end{aligned}
\tag{4.25}
$$

where $C_1 = \frac{\left(2^{R_p/W}-1\right)N_0W}{P_1}$, $C_2 = \frac{\left(2^{\overline{R_s}/W}-1\right)N_0W}{P_2}$, $C_3 = \frac{\left(2^{R_p/W}-1\right)P_2}{P_1}$, $C_4 = \frac{\left(2^{\overline{R_s}/W}-1\right)P_1}{P_2}$ and $\overline{R_s} = \frac{B}{T-\tau}$.

The success probabilities for the fixed access case $\widetilde{q_{1|1}}$, $\widetilde{q_{2|2}}$, $\widetilde{q_{1|\{1,2\}}}$ and $\widetilde{q_{2|\{1,2\}}}$ can be obtained by substituting $\tau = 0$ in Eqs. (4.25) and using that $\overline{R_s} = \frac{B}{T}$.

In our numerical results, we choose the values of the parameters based on practical considerations to illustrate the main conclusions of this chapter. Typical values of the outdoor delay spread are of order of microseconds, hence we use $W =$ 400 KHz to ensure flat fading. We assume that the slot duration T is equal to 10 msec, and for the schemes with sensing, we choose $\tau = 1$ msec (i.e., $\frac{\tau}{T} = 0.1$). The number of bits per packet is chosen to be $B = 2500$ bits. Define $SNR_1 = \frac{P_1}{N_0W}$ and $SNR_2 = \frac{P_2}{N_0W}$, where N_0 is the white Gaussian noise power spectral density. We use $SNR_1 = 3$ dB and $SNR_2 = 1$ dB. Unless stated otherwise, the false alarm and detection probabilities for the schemes with sensing are equal to $P_f = 0.15$ and $P_d = 0.9$. The variances of the fading coefficients are $\sigma_{11} = \sigma_{22} = \sigma_{12} = 1$. We

consider cases of $\sigma_{21} = 1$ and $\sigma_{21} = 5$, i.e., cases of weak and strong interference from the SU on the PU which correspond to large ($\gamma + \delta < 1$) and small ($\gamma + \delta > 1$) probabilities of successful simultaneous transmissions, respectively.

Figure 4.2 shows the stable throughput regions of the schemes under consideration for the case of $\gamma + \delta > 1$ (strong SU interference where $\sigma_{21} = 5$). This is the case where, in general, sensing is needed for a better SU throughput although it represents an overhead for the SU. The randomized access scheme leads to better performance for the SU at the expense of more complexity. It should also be noted that only if the average arrival rate to the PU is small does fixed access outperform all other schemes since no PU protection is needed, and sensing duration can be exploited for data transmission. Although the traditional OSA scheme is simple to implement, the stable rates it achieves are strictly less than those achieved by the more complicated randomized access scheme if λ_1 is less than some value. This means that the traditional OSA scheme has good performance only if λ_1 is sufficiently large; otherwise, it represents a waste of transmission opportunities for the SU and an overprotective policy for the PU. From Fig. 4.3, we note that a better sensor leads to a higher SU stable throughput for the randomized access scheme. Note also that the stable throughput region of the fixed access is independent of the sensor since no sensing takes place in this scheme. Figure 4.4 shows the stable throughput region for the case of $\gamma + \delta < 1$ ($\sigma_{21} = 1$). Unlike the case where $\gamma + \delta > 1$, the fixed access scheme outperforms the randomized access for all values of λ_1, meaning that schemes with no sensing are preferred as they give the SU more duration for data transmission while still protecting the PU receiver. In this case,

the traditional OSA represents a too conservative approach since the SU can afford to be more aggressive in transmission. Finally, we note that for $\gamma + \delta > 1$ as shown in Fig. 4.2, perfect sensing with OSA is mostly better than all other schemes at the cost of high complexity to achieve perfect sensing performance. On the other hand, for $\gamma + \delta < 1$ as shown in Fig. 4.4, fixed and randomized access schemes are better since sensing errors lead to more transmission opportunities for the SU and hence to an increase in throughput.

Figures 4.5 and 4.6 show the SU throughput versus the PU delay tolerance for $\lambda_1 = 0.3$ for the cases of $\gamma + \delta < 1$ and $\gamma + \delta > 1$ respectively. Clearly, the larger the value of D_0 is, the higher the SU throughput becomes, while for $D_0 \leq D_{\min}$, the SU throughput is zero as expected. For large D_0, the delay constraint becomes ineffective and the problem reduces to the stability problem considered in Section 4.3 where the SU throughput values for various schemes coincide with those in Figs. 4.2 and 4.4, respectively, for $\lambda_1 = 0.3$. For the case of $\gamma + \delta < 1$, we note that unlike the case of maximizing the SU stable throughput subject to PU queueing stability where the fixed access scheme outperforms the randomized access, we see from Fig. 4.5 that for values of D_0 slightly larger than $D_{\min}$, the randomized access scheme outperforms the fixed access. This can be explained by noting that in this case, high PU protection is required to avoid violating the PU delay constraint and hence, schemes with sensing are needed to limit the interference on the PU. This also illustrates the fundamental difference between the delay and the stability problems as previously mentioned in Section 4.4.

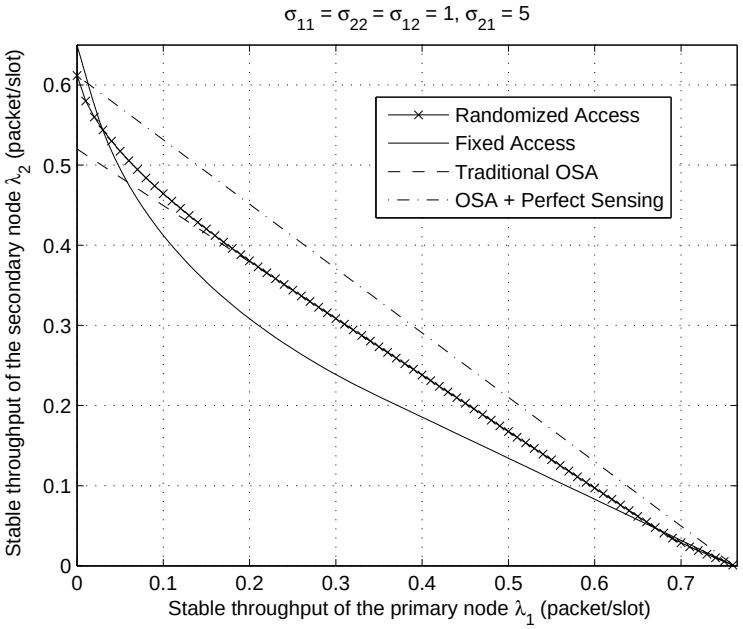

Figure 4.2: Stable throughput region for the case with weak MPR capability.

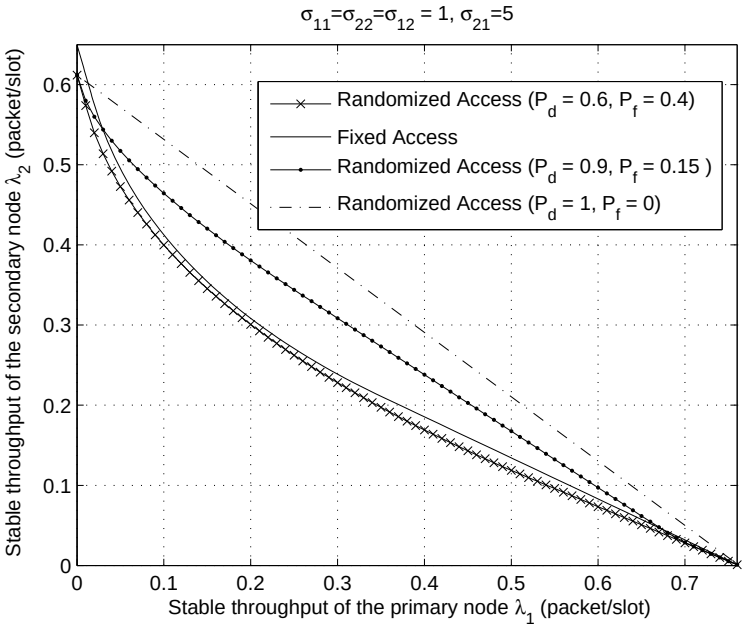

Figure 4.3: Effect of P_f and P_d on the stable throughput region.

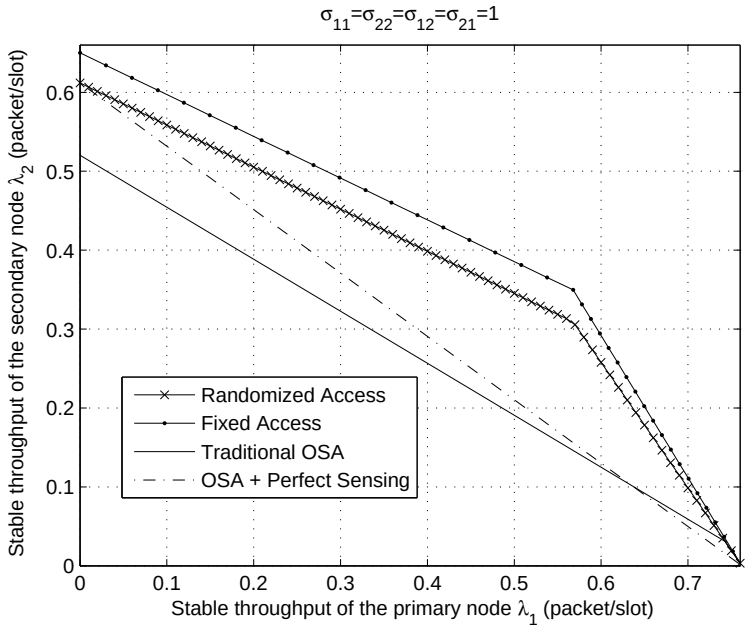

Figure 4.4: Stable throughput region for the case with strong MPR capability.

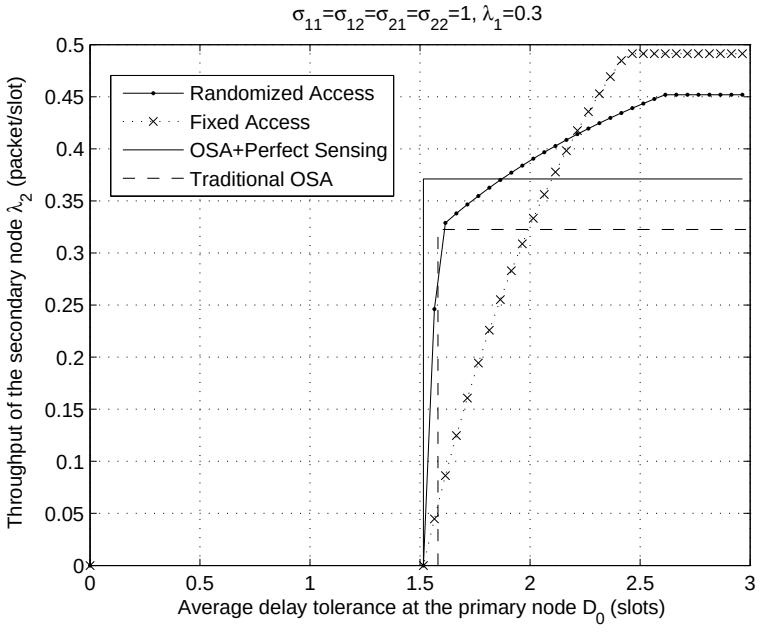

Figure 4.5: Maximum SU throughput versus PU delay tolerance - strong MPR.

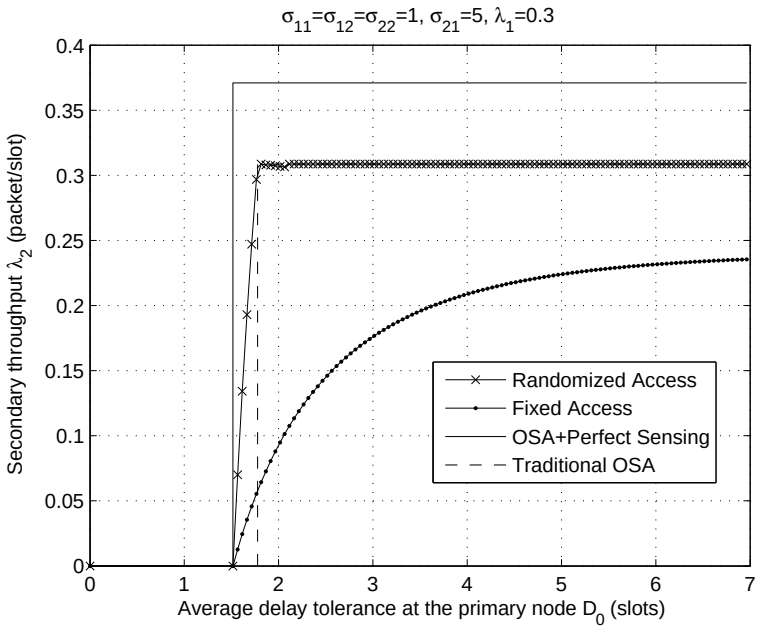

Figure 4.6: Maximum SU throughput versus PU delay tolerance - weak MPR.

4.6 Summary and Conclusions

We studied several MAC layer protocols at the SU that trade complexity of implementation for performance. A cross-layer (PHY/MAC) approach was followed in the analysis. The proposed schemes exploit the SU knowledge of the statistics of the channels as well as the average arrival rate to the PU to enhance the SU throughput. It was shown that, with a stability constraint at the PU, if the receivers are capable to carry simultaneous transmissions successfully or if the average arrival rate at the PU is small; access schemes with no sensing are preferable since they allow more data transmission duration for the SU while the interference on the PU is handled through the decoding capability of the receivers. On the other hand, if simultaneous transmissions of the PU and SU are likely to fail, schemes with sensing

are essential for PU protection. For the case of an average delay constraint at the PU, it is shown that if simultaneous transmissions are likely to fail, schemes with sensing are always preferred. For the case of high success probability of simultaneous transmissions, schemes with sensing are preferable for small delay tolerance; while for large delay tolerance at the PU, schemes with no sensing are preferable.

4.7 Appendix: Solution of the Optimization Problem in Eq. (4.12)

Define the linear transformations

$$\begin{aligned} \phi &= [p_2 P_f + p_1(1 - P_f)] \,, \\ \psi &= [p_2 P_d + p_1(1 - P_d)] \,. \end{aligned} \tag{4.26}$$

For the case of interest where $P_f < P_d$, the region $\mathcal{D}_1 = \{p_1, p_2 | 0 \leq p_1 \leq 1, 0 \leq p_2 \leq 1\}$ is mapped under this transformation to the region $\mathcal{D}_2$ as shown in Fig. 4.7.

The region $\mathcal{D}_2$ is a parallelogram with sides as given in Table 4.1.

Side	Equation	Properties
0A	$\psi = \frac{(1-P_d)}{(1-P_f)}\phi$	$p_2 = 0$
0C	$\psi = \frac{P_d}{P_f}\phi$	$p_1 = 0$
AB	$\psi = 1 + \frac{P_d}{P_f}(\phi - 1)$	$p_1 = 1$
CB	$\psi = 1 + \frac{(1-P_d)}{(1-P_f)}(\phi - 1)$	$p_2 = 1$

Table 4.1: Mapping of feasibility region under the linear transformation.

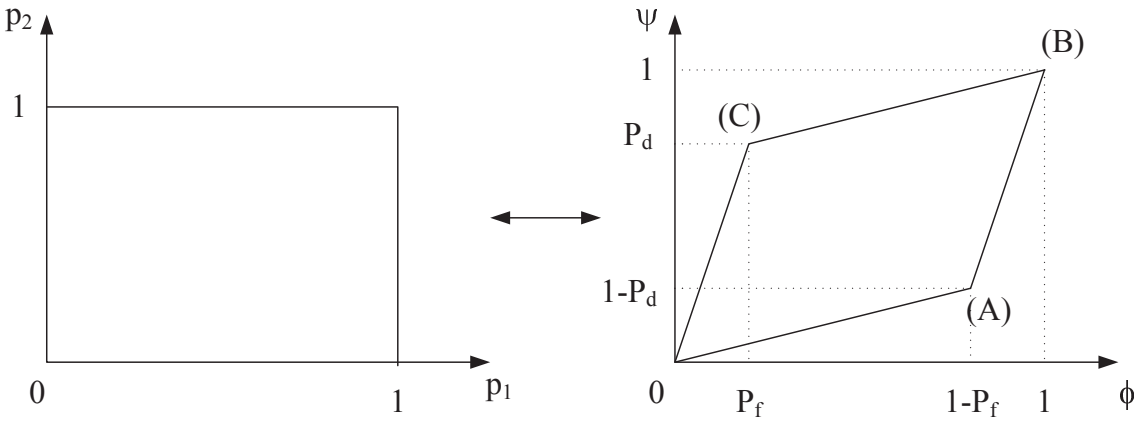

Figure 4.7: Linear Transformation

The optimization problem in Eq. (4.12) can then be written as

$$\max_{(\phi,\psi)\in\mathcal{D}_2} y = q_{2|2}\phi\left[1 - \frac{\lambda_1}{q_{1|1}(1-\gamma\psi)}\left(1 - \frac{\psi(1-\delta)}{\phi}\right)\right]$$

$$\text{s.t. } \lambda_1 \leq q_{1|1}(1-\gamma\psi). \tag{4.27}$$

It can be shown that

$$\frac{\partial y}{\partial \phi} = 1 - \frac{\lambda_1}{q_{1|1}(1-\gamma\psi)}, \tag{4.28}$$

$$\frac{\partial y}{\partial \psi} = \frac{(-\lambda_1)}{q_{1|1}}\left[\frac{\gamma\phi - (1-\delta)}{(1-\gamma\psi)^2}\right]. \tag{4.29}$$

By Eq. (4.28) and the constraint in Eq. (4.27), it is clear that $\frac{\partial y}{\partial \phi} \geq 0$ for every ψ. On the other hand, $\frac{\partial y}{\partial \psi} \geq 0 \Leftrightarrow \phi \leq \frac{(1-\delta)}{\gamma}$.

We differentiate between two cases.

(A) $\gamma + \delta \leq 1 \Leftrightarrow \frac{(1-\delta)}{\gamma} \geq 1$

In this case, both $\frac{\partial y}{\partial \psi}$ and $\frac{\partial y}{\partial \phi}$ are positive, that is y increases with increasing ϕ or ψ. Hence, choosing $\phi = 1$ and $\psi = 1$ is best. However, this may not be feasible for some values of λ_1 because of the constraint in Eq. (4.27) since $\psi \leq \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right)$. Hence, we consider three subcases.

(1) $0 \leq \lambda_1 \leq q_{1|1}\,(1-\gamma) \Leftrightarrow \psi = \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right) \geq 1$

The optimum solution is to choose $\psi = \phi = 1$, and the optimum transmission probabilities are $p_1^* = p_2^* = 1$.

(2) $q_{1|1}\,(1-\gamma) \leq \lambda_1 \leq q_{1|1}\,[1-\gamma(1-P_d)] \Leftrightarrow (1-P_d) \leq \psi = \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right) \leq 1$

The optimum solution is to choose $\psi^* = \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right)$. An optimum point is the intersection of the horizontal line $\psi^* = \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right)$ with the segment $\overline{AB}$. In this case, the optimum transmission probabilities are given by $p_1^* = 1$ and by substituting in ψ^*, we obtain $p_2^* = \frac{1}{P_d}\left[\frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right) - (1-P_d)\right]$.

It should be noted that we only provided one optimal solution (which is not unique in this case). This can be seen by noting that whenever the constraint in Eq. (4.27) $\psi \leq \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right)$ is satisfied with equality, the objective function is independent of ϕ and the optimal choice for ϕ is not unique. Therefore, the values of p_1^* and p_2^* are not unique.

(3) $q_{1|1}\,[1-\gamma(1-P_d)] \leq \lambda_1 \leq q_{1|1} \Leftrightarrow 0 \leq \psi = \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right) \leq (1-P_d)$

The optimum solution is to choose $\psi^* = \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right)$. An optimum point is the intersection of the horizontal line $\psi^* = \frac{1}{\gamma}\left(1 - \frac{\lambda_1}{q_{1|1}}\right)$ with the segment $\overline{0A}$. In this case, the optimum transmission probabilities are given by $p_2^* = 0$ and by substituting in ψ^*, $p_1^* = \frac{1}{\gamma(1-P_d)}\left(1 - \frac{\lambda_1}{q_{1|1}}\right)$.

Finally the three subcases can be combined as in Eq. (4.13).

(B) $\gamma + \delta > 1 \Leftrightarrow \frac{(1-\delta)}{\gamma} < 1$. In this case, $\frac{\partial y}{\partial \psi}$ can be greater, less than or equal to zero. However, $\frac{\partial y}{\partial \phi} \geq 0$ for all ψ. Hence, for the case of $P_f < P_d$, as shown in Fig. 4.7, the optimum point (ϕ^*, ψ^*) lies on either the segment $\overline{0A}$ or $\overline{AB}$, that is, for

optimality, either $p_1^* = 1$ or $p_2^* = 0$.

For $p_2^* = 0$, $\psi = p_1(1 - P_d)$ and $\phi = p_1(1 - P_f)$. By substituting in the objective function in Eq. (4.27) and solving $\frac{dy}{dp_1} = 0$, we obtain

$$p_1 = \frac{1}{\gamma(1 - P_d)} \left(1 - \sqrt{\frac{\lambda_1}{q_{1|1}} \left(1 - \frac{(1 - P_d)(1 - \delta)}{1 - P_f} \right)} \right). \tag{4.30}$$

Moreover, when $p_2 = 0$, we have $\frac{d^2y}{dp_1^2} = \frac{-2\gamma\lambda_1 q_{2|2}(1-P_f)(1-P_d)\left(1-\frac{(1-P_d)(1-\delta)}{(1-P_f)}\right)}{q_{1|1}(1-\gamma(1-P_d)p_1)^3} < 0$, and hence y is concave and p_1 in Eq. (4.30) is indeed a maximizer. By substituting $p_2 = 0$ into the constraint in Eq. (4.27), we get that

$$p_1 \leq \frac{1}{\gamma(1 - P_d)} \left(1 - \frac{\lambda_1}{q_{1|1}} \right). \tag{4.31}$$

Since $0 \leq p_1 \leq 1$, and by concavity of y, the optimal p_1 is given by p_1^* as in Eq. (4.15).

Similarly, for the case of $p_1 = 1$, we have that $\phi = (1 - P_f) + p_2 P_f$ and $\psi = (1 - P_d) + p_2 P_d$. Substituting into the objective function in Eq. (4.27) and solving for p_2, we obtain

$$p_2 = \frac{1}{P_d} \left[\frac{1}{\gamma} \left(1 - \sqrt{\frac{\lambda_1 C}{q_{1|1} P_f}} \right) - (1 - P_d) \right], \tag{4.32}$$

where $C = P_f + \gamma(P_d - P_f) - (1 - \delta)P_d$. Note that for the case of interest where $\gamma + \delta \geq 1$, we have $C \geq 0$.

Substituting $p_1 = 1$ into the constraint in Eq. (4.27), we get

$$p_2 \leq \frac{1}{P_d} \left[\frac{1}{\gamma} \left(1 - \frac{\lambda_1}{q_{1|1}} \right) - (1 - P_d) \right]. \tag{4.33}$$

Since $0 \leq p_2 \leq 1$, we obtain p_2^* as in Eq. (4.16).

Chapter 5: Cognitive Relaying with Network Coding for Multicasting Networks

5.1 Introduction

Cooperative communications was shown to be effective in combating multipath fading over wireless channels due to the induced spatial diversity. Much work has been done to analyze cooperative diversity at the physical layer based on information theoretic considerations [17, 78], and at the network layer [21].

Little work has been done to incorporate network coding with cooperative diversity. In [79] different protocols combining deterministic network coding and cooperative diversity are proposed and it is shown that, for a single source - single relay system with two destinations, the use of network coding at the relay increases the stable throughput of the source. In our work [80], we propose a network-level relaying protocol in which the relay uses the periods of silence of the source to forward the source's unsuccessful packets and hence avoiding allocating any explicit channel resources to the relay. Furthermore, the relay performs random linear network coding on the packets it has in queue. It is shown that, compared with ARQ or protocols based solely on network coding [81], the stable throughput for the source

increases by relaying and further enhancement can be achieved by using network coding at the relay.

The chapter is organized as follows. We discuss the system model in Section 5.2 and introduce various protocols in Section 5.3. In Section 5.4, we evaluate the maximum stable throughput rate of different protocols and quantify the improvements due to cooperation and network coding. In Section 5.5, we present the numerical results and in Section 5.6 we conclude the chapter.

5.2 System Model

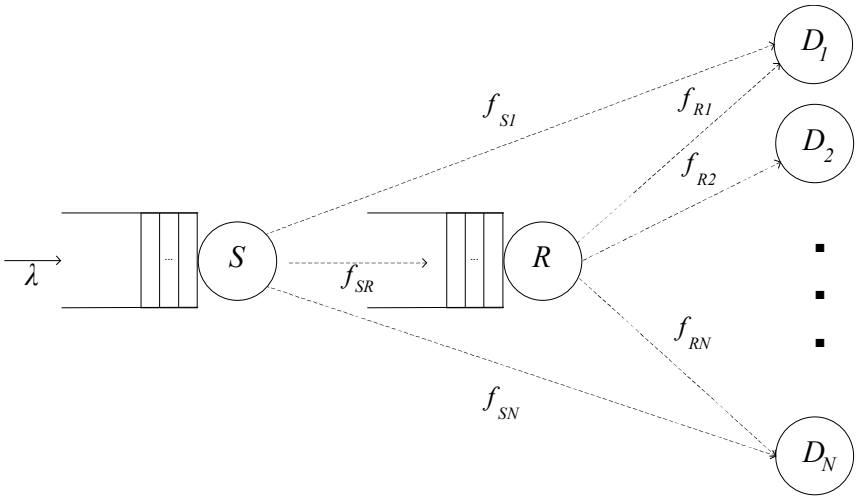

Figure 5.1: System Model.

We consider one source node transmitting packets to each of N receivers with the aid of a relay as shown in Fig. 5.1. We consider a slotted synchronous system in which one slot duration is equal to one packet transmission duration. Packets are independently generated (or received) according to a Bernoulli process with average rate λ and are addressed to each of the N receivers. Noise at the receivers

and at the relay is assumed to be i.i.d. additive white complex Gaussian random process with zero mean and variance N_0. All links are subject to i.i.d. flat fading with coefficients h_{ij}. After each transmission, an instantaneous and error-free acknowledgment/non-acknowledgement (ACK/NACK) messages are fedback to the corresponding transmitter (source or relay). The ACK/NACK messages sent to the source are also heard by the relay. Throughput the chapter we designate the source and relay nodes by the subscripts S and R respectively and the jth destination node by the subscript $j \in \{1, 2, ..., N\}$. We adopt the SNR threshold model for reception in which a node $j \in \{1, 2, ..., N\}$ can successfully decode a packet transmitted by node $i \in \{S, R\}$ if the SNR at node j exceeds some threshold β. This can be expressed in terms of success probabilities f_{ij} over $(i - j)$ link

$$f_{ij} = \Pr\left[SNR_{ij} > \beta\right] = \Pr\left[\frac{|h_{ij}|^2 P}{N_0} > \beta\right], \tag{5.1}$$

where P_i is the transmission power of node $i \in \{S, R\}$. We adopt the definition of queueing stability as in Section 2.3 and make use of Loynes' Theorem. We use the stable throughput of the source node as the performance metric to compare several transmission protocols with and without relaying and network coding.

5.3 Network Protocols

In this section, we present different transmission protocols for multicasting networks with and without relaying and network coding. We compare their stable throughput performance in Section 5.5.

5.3.1 Retransmission Policy (ARQ)

This is the ordinary retransmission protocol where neither relaying nor network coding takes place. The source node transmits a new packet only if the previous packet has been successfully received by all receivers (i.e., an ACK message has been received from all the N destinations), otherwise the same packet is retransmitted.

5.3.2 Random Linear Network Coding

The source node buffers the incoming packets in its queue. When K packets are accumulated, the source node transmits one random linear combination of these packets at a time until they are all successfully decoded by all the N receivers. Then, the source node transmits a new combination and so on. The relay does not assist the source in transmitting its packets.

5.3.3 Cognitive Relaying

The source node transmits its traffic with the help of the cognitive relay without performing network coding on the packets. At each time slot, if the transmitted packet is successfully received by all N destinations or by the relay; it is released from the source queue; otherwise it is kept in the source queue for retransmission in the following time slot. At the beginning of every time slot, the relay senses the channel. If the source does not have any traffic to send (idle), which happens infinitely often[1], the relay accesses the channel during these idle time slots to transmit

[1]This is due to the stability of the source queue. Refer to Section 2.3 for details.

the packets in its queue to the destinations that could not receive them. This way, no explicit channel resources are accorded to the relay. We assume that sensing at the relay is perfect and hence the queues at the source and the relay nodes do not interact. A packet is released from the relay's queue if it is successfully decoded by all the destinations.

5.3.4 Cognitive Relaying with Network Coding at the Relay

Similar to the cognitive relaying protocol but the relay transmits linear combinations of the packets it has in queue while the source does not perform any network coding. Note that we only consider the case of fixed K, that is the relay only transmits if it has at least K packets in queue.

5.4 Stable Throughput Analysis

In this section , we compute the maximum stable throughput rate at the source queue for each of the protocols described in Section 5.3. By the stationarity of the fading and noise processes, it is straightforward to establish the stationarity of the service processes at the source and relay nodes. Since the arrival process at the source queue is independent from the service process at the source queue, by Loynes' theorem, the stability condition of the source queue is that the average arrival rate at the source node λ should be less than the average service rate of the queue. The case with relaying is different and will be considered later.

5.4.1 Retransmission Policy (ARQ)

The stability condition of the source queue is given by

$$\lambda < \mu_{ARQ} = \left(1 + \sum_{t=1}^{\infty}\left(1 - \prod_{i=1}^{N}\left[1 - (1 - f_{Si})^{t}\right]\right)\right)^{-1}. \tag{5.2}$$

Proof: Let N_i be the number of transmissions required until the ith receiver successfully decode a packet transmitted by the source node, then

$$\Pr[N_i = k] = f_{Si}\left(1 - f_{Si}\right)^{k-1}, k = 1, 2, 3, ... \text{ and } i \in \{1, 2, ..., N\}. \tag{5.3}$$

Let T denote the number of source transmissions until all the receivers successfully decode the packet, then

$$T = \max_{i=1,2,...,N} N_i. \tag{5.4}$$

By independence of the channel fading processes

$$\Pr[T \leq t] = \prod_{i=1}^{N} \Pr\left[T_i \leq t\right] = \prod_{i=1}^{N}\sum_{r=1}^{t} f_{S_i}\left(1 - f_{S_i}\right)^{r-1}. \tag{5.5}$$

Since the random variable T is non-negative, the expected value of T is given by

$$\mathbb{E}[T] = \sum_{t=0}^{\infty} \Pr[T > t] = 1 + \sum_{t=1}^{\infty}\left(1 - \prod_{i=1}^{N}\sum_{r=1}^{t} f_{Si}\left(1 - f_{Si}\right)^{r-1}\right). \tag{5.6}$$

Hence, the stability condition is given by

$$\lambda < \frac{1}{\mathbb{E}[T]} = \left(1 + \sum_{t=1}^{\infty}\left(1 - \prod_{i=1}^{n}\sum_{r=1}^{t} f_{Si}\left(1 - f_{Si}\right)^{r-1}\right)\right)^{-1}. \tag{5.7}$$

■

5.4.2 Random Linear Network Coding

In [81] it has been shown that that the queueing stability condition of a queue transmitting network-coded packets is given by $\lambda < \mu_{NC} = \frac{K}{\mathbb{E}[T_{NC}]}$, where

$$\mathbb{E}[T_{NC}] = \sum_{l=K}^{\infty} \frac{1}{q^{l-K}} \left(1 - \frac{1}{q^K}\right) \prod_{k=l+1-K}^{l-1} \left(1 - \frac{1}{q^k}\right) \left(l + \sum_{t=l}^{\infty} \left(1 - \prod_{i=1}^{N} \sum_{r=l}^{t} \binom{r-1}{l-1} (1 - f_{Si})^{r-l} f_{Si}^{l}\right)\right), \tag{5.8}$$

where q is the field size used for network coding ($GF(q)$) and K is the coding block size (number of packets over which network coding is performed).

5.4.3 Cognitive Relaying

In this case, two queues are involved: the source queue, and the relay queue where the source's packets to be relayed are stored. The system is stable if both queues are stable. In the following we analyze each of the two queues separately.

5.4.3.1 Source Queue

The arrival process at the source node is stationary by assumption. The service process is also stationary as it depends only on the source-destination channels and the source-relay channel which are subject to i.i.d. (and hence stationary) fading. Moreover, the service at the source node is independent of the arrivals and thus they are jointly stationary and we can apply Loynes' theorem for the source queue. In other words, the source queue is stable if $\lambda < \mu$, where

$$\mu = \frac{1}{1 + \sum_{t=1}^{\infty} (1 - f_{SR})^t \left[1 - \prod_{i=1}^{N} \left[1 - (1 - f_{Si})^t\right]\right]}. \tag{5.9}$$

Note that this protocol reduces to the retransmission protocol $f_{SR} = 0$ as expected.

Proof: The source queue is served if the packet is successfully delivered to the relay or to all N destinations. Let T be the number of time slots needed to serve the source queue, then

$$T = \min\{T_L, T_D\}, \tag{5.10}$$

where T_L is the number of time slots needed to deliver the packet to the relay and T_D is the number of time slots needed to deliver the packet to all N destinations.

$$\Pr\left[T_L = k\right] = f_{SR}\left(1 - f_{SR}\right)^{k-1}, k = 1, 2, 3, ... \tag{5.11}$$

$$T_D = \max_{i=1,2,...,N} T_{Di}, \tag{5.12}$$

where T_{Di} represents the number of time slots needed to deliver the packet to the ith destination. we have that

$$\Pr[T_{Di} = r] = f_{Si}\left(1 - f_{Si}\right)^{r-1}, r = 1, 2, 3, ... \tag{5.13}$$

$$\Pr[T_D \leq t] = \prod_{i=1}^{N} \sum_{r=1}^{t} f_{Si}\left(1 - f_{Si}\right)^{r-1}, \tag{5.14}$$

$$\Pr[T > t] = \Pr[T_D > t]\Pr[T_L > t] = \left(1 - f_{SR}\right)^{t} \left(1 - \prod_{i=1}^{N} \sum_{r=1}^{t} f_{Si}\left(1 - f_{Si}\right)^{r-1}\right), \tag{5.15}$$

$$\mathbb{E}[T] = 1 + \sum_{t=1}^{\infty} (1 - f_{SR})^{t} \left(1 - \prod_{i=1}^{N} \sum_{r=1}^{t} f_{Si}\left(1 - f_{Si}\right)^{r-1}\right), \tag{5.16}$$

$$\mu = \frac{1}{\mathbb{E}[T]} = \left(1 + \sum_{t=1}^{\infty} (1 - f_{SR})^{t} \left(1 - \prod_{i=1}^{N} \sum_{r=1}^{t} f_{Si}(1 - f_{Si})^{r-1}\right)\right)^{-1}. \tag{5.17}$$

By Loynes' theorem, the proof is complete. ■

5.4.3.2 Relay Queue

At the time when the relay starts transmitting the packets it has in queue, there are 2^N possible states of the N destinations regarding the state of success of packet reception. One of them is that all destinations received the packet which is uninteresting since in that case, the relay has no role in packet delivery. Thus, we need to consider each of the other $2^N - 1$ cases separately. According to our assumptions, the relay receives ACK messages from the destinations that already received the packet, so the exact state of the N destinations is known at the relay. Each state of the N destinations is identified by a set $\mathcal{S}$ whose elements are the nodes that already received the packet while source was transmitting and a set $\mathcal{F} = \mathcal{S}^c = \{1, 2, ..., N\} \setminus \mathcal{S}$, representing nodes that failed to receive that packet and the relay has to forward the packets to them. The arrival and service processes at the relay are stationary as they are functions of stationary processes which are the fading processes. However, they are not independent as in the case of the source queue. The reason is that if the m-th packet takes a longtime to reach the relay, which means that it has a long inter-arrival time, then it is more likely to be successfully delivered to a larger number of destinations during the source transmissions and hence, it will get served faster at the relay. However, the arrival and service processes are still jointly stationary since the m-th and the $(m+1)$-st packets are subject to the same transmission conditions which depend on the i.i.d. fading processes. Thus, the arrival and service processes are jointly stationary and hence Loynes' theorem can still be applied at the relay node as a necessary and

sufficient condition for stability. The stability condition for the relay node is then given by $\lambda_R < \mu_R$, where

$$\lambda_R = \left(\frac{\lambda}{\mu}\right) f_{SR} \left(1 - \prod_{i=1}^{N} f_{Si}\right), \tag{5.18}$$

$$\mu_R = \left(1 - \frac{\lambda}{\mu}\right) / \left(\mathbb{E}[T_R] - \frac{\lambda}{\mu}\right), \tag{5.19}$$

where $\mathbb{E}[T_R]$ is the average number of time slots needed for the relay to deliver the packet to all destinations that failed to receive it if the relay transmits continuously, not just during idle time slots of the source and is given by

$$\mathbb{E}[T_R] = \sum_{\text{all states } S,F} \left[\sum_{t=0}^{\infty} \left[1 - \prod_{i \in F} \left(1 - (1 - f_{Ri})^t\right) \right] \sum_{m=1}^{\infty} \left[Q(1-Q)^{m-1} \prod_{i \in F} (1 - f_{Si})^m \prod_{j \in S} \left[1 - (1 - f_{Sj})^m\right] \right] \right], \tag{5.20}$$

where $Q = f_{SR}\left(1 - \prod_{i=1}^{N} f_{Si}\right)$.

Proof: Let $\mathbb{E}[T_R]$ be the average number of time slots needed for the relay to deliver the packet to all destinations that failed to receive the packet if we allow the relay to transmit continuously and not to be confined to the idle time slots. Let v_1,v_2,... be a sequence of random variables representing the number of successive time slots in which the source is busy, possibly of length zero if no arrivals occur. It is clear that this sequence forms an i.i.d sequence of random variables. In a given time slot we have Pr[an arrival occurs] $= 1 -$ Pr[no arrivals] $= \lambda$. The source queue is a Geo/G/1 queue. Let $\rho = \frac{\lambda}{\mu}$, then using the result in [76]

$$\mathbb{E}[\text{Busy period}] = \frac{1}{\mu - \lambda}, \tag{5.21}$$

$$\mathbb{E}[v] = \lambda \mathbb{E}[\text{Busy period}] = \frac{\rho}{1 - \rho}. \tag{5.22}$$

Let T be the total number of time slots needed for the relay to get served, including those in which the source will be transmitting. Then since between T_R time slots of

the relay we have $(T_R - 1)$ source busy periods, possibly of length zero, we have

$$T = T_R + \sum_{i=1}^{T_R - 1} v_i, \tag{5.23}$$

$$\mathbb{E}[T] = \mathbb{E}[T_R] + (\mathbb{E}[T_R] - 1)\,\mathbb{E}[v] = \frac{\mathbb{E}[T_R] - \rho}{1 - \rho}, \tag{5.24}$$

$$\mu_R = \frac{1}{\mathbb{E}[T]} = \frac{1 - \rho}{\mathbb{E}[T_R] - \rho}. \tag{5.25}$$

We then prove the equation for $\mathbb{E}[T_R]$. Let T^* be the number of time slots until the relay has an arrival, then

$$\Pr[T^* = m] = f_{SR}\left(1 - \prod_{i=1}^{n} f_{Si}\right)\left[1 - f_{SR}\left(1 - \prod_{i=1}^{n} f_{Si}\right)\right]^{m-1}. \tag{5.26}$$

The destinations are at a certain state before the relay starts transmission, where the state is described by the set $\mathcal{F}$ of destinations that failed to receive the packet. Then $T_R = \max_{i \in \mathcal{F}} T_{R,i}$, where $T_{R,i}$ is the time for the relay to successfully deliver a packet to the i-th node, which is geometrically distributed with parameter f_{Ri}. We have that

$$\Pr\left[T_R \le t\right] = \prod_{i \in \mathcal{F}} \sum_{k=1}^{t} f_{Ri}\left(1 - f_{Ri}\right)^{k-1} = \prod_{i \in \mathcal{F}} \left[1 - \left(1 - f_{Ri}\right)^{t}\right], \tag{5.27}$$

$$\mathbb{E}[T_R|\text{ state } \mathcal{S}, \mathcal{F}] = \sum_{t=0}^{\infty}\left[1 - \prod_{i \in \mathcal{F}}\left[1 - \left(1 - f_{Ri}\right)^{t}\right]\right]. \tag{5.28}$$

Thus

$$\mathbb{E}[T_R] = \sum_{\text{all states}\mathcal{S},\mathcal{F}} \sum_{t=0}^{\infty}\left[1 - \prod_{i \in \mathcal{F}}\left[1 - \left(1 - f_{Ri}\right)^{t}\right]\right] P\left(\mathcal{S}, \mathcal{F}\right), \tag{5.29}$$

where $P\left(\mathcal{S}, \mathcal{F}\right)$ is the probability that the destinations are in state $(\mathcal{S}, \mathcal{F})$ when the

relay starts transmission.

$$P(\mathcal{S},\mathcal{F}) = \sum_{m=1}^{\infty} \Pr(\mathcal{S},\mathcal{F}|T^*=m)\Pr[T^*=m] \tag{5.30}$$

$$= \sum_{m=1}^{\infty} \left[\prod_{i\in\mathcal{F}} (1-f_{Si})^m \prod_{i\in\mathcal{S}} [1-(1-f_{Si})^m]\right] \Pr[T^*=m]. \tag{5.31}$$

Hence, by direct substitution, Eq. (5.20) directly follows. ■

For system stability, both source and relay queues should be stable. Hence, both $\lambda < \mu$ and $\lambda_R < \mu_R$ must be satisfied. Substituting λ_R from Eq. (5.18) in Eq. (5.19), we get the stability condition for the source queue as

$$\lambda < \min\left(\mu, \mu\frac{(1+Q\mathbb{E}[T_R]) - \sqrt{(1+Q\mathbb{E}[T_R])^2 - 4Q}}{2Q}\right), \tag{5.32}$$

where $Q = f_{SR}\left(1-\prod_{i=1}^{N} f_{Si}\right)$ and $\mathbb{E}[T_R]$ is given by Eq. (5.20).

5.4.4 Cognitive Relaying with Network Coding at the Relay

It is clear that the arrival and service processes of the source node as well as the arrival process to the relay are identical to the case of relaying with no network coding where the rates are given by Eqs. (5.9) and (5.18). However, the service process at the relay node is different.

The stability condition for this protocol is given by

$$\lambda < \min\left(\mu, \mu\frac{(K+Q\mathbb{E}[T_R]) - \sqrt{(K+Q\mathbb{E}[T_R])^2 - 4KQ}}{2Q}\right), \tag{5.33}$$

where $Q = f_{SR}\left(1-\prod_{i=1}^{N} f_{Si}\right)$, μ is as given by Eq. (5.9), q is the field size of network coding used at the relay, K is the coding block size. The sets F_i and S_i

describe the state of reception of the ith packet at the N destinations when the relay starts transmission.

$$\mathbb{E}[T_R] = \sum_{\text{all states } F_1,..,F_K} \mathbb{E}[T_R|\text{State } F_1, F_2, ...F_K]\Pr[\text{State } F_1, F_2, ..., F_K], \quad (5.34)$$

$$\mathbb{E}[T_R|\text{ State } F_1, F_2, ...F_K] =$$
$$\sum_{l=K}^{\infty} \frac{1}{q^{l-K}}\left(1-\frac{1}{q^K}\right)\prod_{k=l+1-K}^{l-1}\left(1-\frac{1}{q^k}\right)\left[l+\sum_{t=l}^{\infty}\left(1-\prod_{i\in\bigcup_{j=1}^{K}F_j}\sum_{r=l}^{t}\binom{r-1}{l-1}(1-f_{Ri})^{r-l}f_{Ri}^{l}\right)\right], \quad (5.35)$$

$$\Pr[\text{State } F_1, F_2, ..., F_K] = \prod_{l=1}^{K}\left[\sum_{m=1}^{\infty}\left[Q(1-Q)^{m-1}\prod_{i\in F_l}(1-f_{Si})^m\prod_{j\in S_l}[1-(1-f_{Sj})^m]\right]\right]. \quad (5.36)$$

Proof: For every packet of the K packets, we have a corresponding state at the N destinations, i.e. for every packet $j \in \{1, 2, ..., K\}$, we have a set F_j of destinations that could not receive that packet. For the relay to serve the K packets, all the destinations in the union of the sets F_j (i.e. $\bigcup_{j=1}^{K} F_j$) must be able to successfully decode them. Thus, given a certain state $(F_1, F_2, ..., F_K)$ of the K packets and using the result of the case of network coding in Section 5.4.2, we get $\mathbb{E}[T_R|$ State $F_1, F_2, ...F_K]$ as in Eq. (5.35). The rest of the proof follows along the same lines as the case of relaying with no network coding. The factor K is introduced by noting that $\mathbb{E}[T_R]$ is the time required to decode a batch of K packets. Hence, the time required to decode one packet is $\frac{\mathbb{E}[T_R]}{K}$. ■

5.5 Numerical Results

In Fig. 5.2, we compare the maximum stable throughput rate achieved at the source node for the first three protocols as given by Eqs. (5.2), (5.8) and (5.33). For clarity of presentation, we consider a symmetric configuration in which all the source-destinations links have the same success probability denoted by f_{SD} and all relay-destinations links have the same success probability denoted by f_{RD}. It is clear that relaying leads to a significant increase in the stable throughput rate of the source node compared with ARQ and NC where the relay does not assist the source in forwarding its traffic and thereby losing the advantage of spatial diversity. It should be noted that relaying only helps whenever the success probability of the relay-destination channels is larger than that of the source-destination channels.

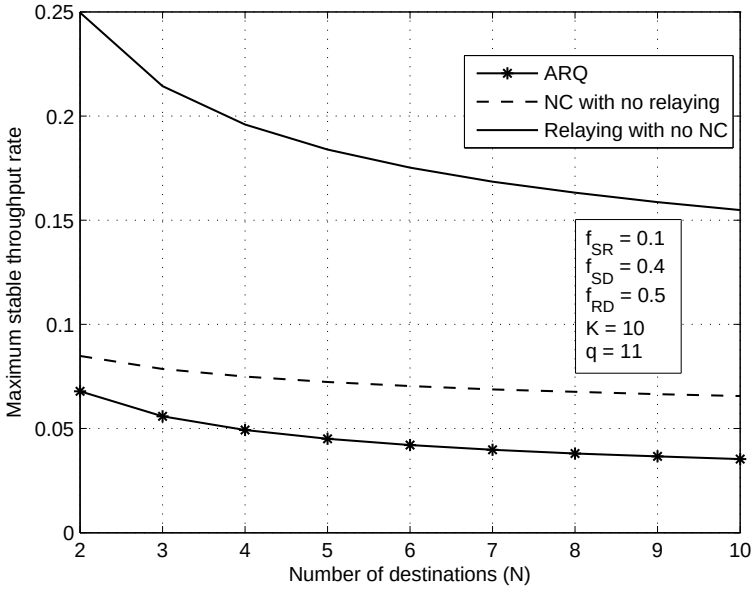

Figure 5.2: Stable throughput rates of protocols (A),(B) and (C).

In Figs. 5.3 and 5.4, we compare the maximum stable throughput rates of relaying with and without network coding as given, respectively, by Eqs. (5.32) and (5.33), where K and q are the respectively the coding block size and the field size of the network coding scheme used at the relay. It is clear that random linear network coding at the relay can increase the maximum stable throughput of the source by increasing q or K, and it becomes more advantageous than relaying without network coding as the number of destinations gets larger. It should also be noted that for small values of K and q, relaying without network coding can outperform relaying with network coding. This is due to the fact that for small values of K and q, the probability of choosing the coefficients to be all zeros $\left(= \frac{1}{q^K}\right)$ or previously chosen coefficients is high, which degrades the performance of network coding.

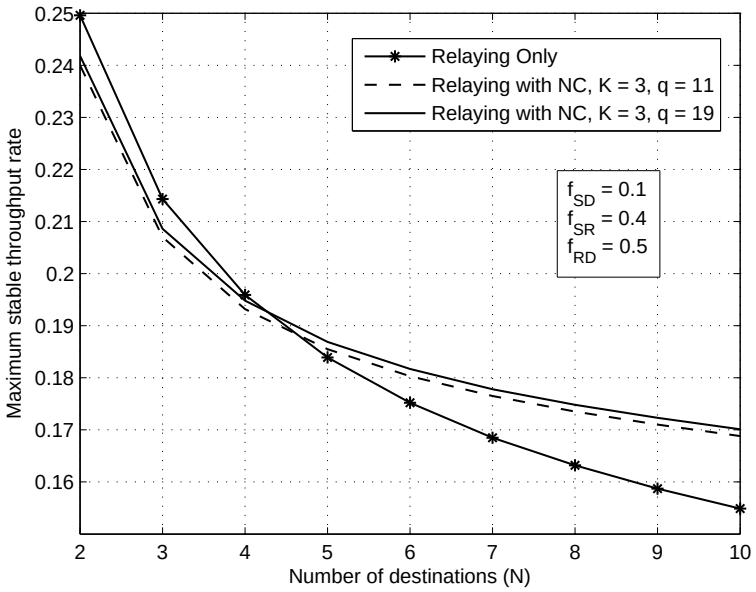

Figure 5.3: Effect of network coding at the Relay for various values of field size q.

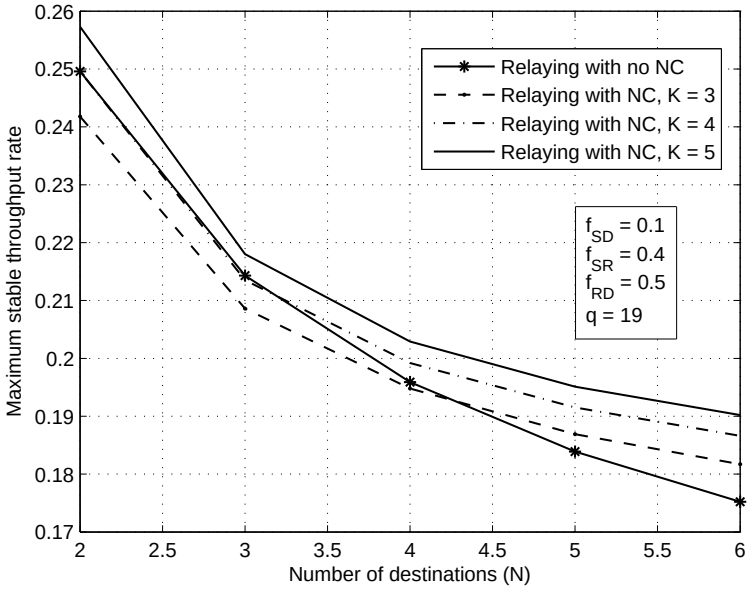

Figure 5.4: Effect of using network coding at the relay for various values of coding block size K.

5.6 Summary and Conclusions

In this work, we proposed and analyzed a (PHY/MAC) protocol for wireless multicasting networks that exploits cognitive relaying, as well as an enhanced protocol that combines both benefits of cognitive relaying and network coding at the relay. Cognitive relaying allows the relay to exploit the idle time slots of the source node and hence avoiding allocating any explicit resources to the relay. Our analysis showed that relaying can substantially increase the maximum stable throughput rate of the source node and further throughput gains can be achieved by using random linear network coding at the relay.

5.7 Appendix (A): Case of symmetric network

In this section, we provide expressions for the symmetric case where $f_{Si} = f_{SD}$ and $f_{Ri} = f_{RD}$ for all $i \in \{1, 2, ..., N\}$.

5.7.1 Retransmission Policy (ARQ)

Using the Binomial theorem, it is straightforward to establish that

$$\mu_{ARQ} = \left[1 - \sum_{r=1}^{N} \binom{N}{r} (-1)^r \left(\frac{(1-f_{SD})^r}{1-(1-f_{SD})^r}\right)\right]^{-1}. \tag{5.37}$$

5.7.2 Random Linear Network Coding

In this case, Eq. (5.8) can be written as

$$\mathbb{E}[T_{NC}] = \sum_{l=K}^{\infty} \frac{1}{q^{l-K}} \left(1 - \frac{1}{q^K}\right) \prod_{k=l+1-K}^{l-1} \left(1 - \frac{1}{q^k}\right) \left(l + \sum_{t=l}^{\infty} \left(1 - \left(\sum_{r=l}^{t} \binom{r-1}{l-1} (1-f_{SD})^{r-l} f_{SD}^{l}\right)^N\right)\right). \tag{5.38}$$

5.7.3 Cognitive Relaying

Using the Binomial Theorem, the average service rate of the source queue can be written as

$$\mu = \left[1 - \sum_{r=1}^{N} \binom{N}{r} (-1)^r \left(\frac{(1-f_{SR})(1-f_{SD})^r}{1-(1-f_{SR})(1-f_{SD})^r}\right)\right]^{-1}. \tag{5.39}$$

Let $Q = f_{SR}\left(1 - f_{SD}^N\right)$ and denote by $|F|$ the cardinality of the set F. Clearly in the symmetric case, summing over all possible states as in Eq. (5.20) is equivalent to summing over all cardinalities of the sets that failed to receive the packet. After

some algebra, it can be shown that

$$\mathbb{E}[T_R] = \sum_{|F|=1}^{N} \sum_{r=0}^{|S|} \sum_{v=1}^{|F|} \binom{|F|}{v} \binom{|S|}{r} \frac{(-1)^{r+v+1} Q(1-f_{SD})^{|F|+r}}{\left(1-(1-Q)\left(1-f_{SD}\right)^{(|F|+r)}\right)\left(1-(1-f_{RD})^{v}\right)}, \tag{5.40}$$

where $|S| = N - |F|$.

5.7.4 Cognitive Relaying with Network Coding at the Relay

The expressions in the symmetric case take the following forms.

$$\mathbb{E}[T_R| \text{ State } F_1, F_2, ...F_K] =$$
$$\sum_{l=K}^{\infty} \frac{1}{q^{l-K}} \left(1-\frac{1}{q^K}\right) \prod_{k=l+1-K}^{l-1} \left(1-\frac{1}{q^k}\right) \left[l + \sum_{t=l}^{\infty} \left(1 - \left(\sum_{r=l}^{t} \binom{r-1}{l-1} (1-f_{RD})^{r-l} f_{RD}^{l} \right)^{\left|\bigcup_{j=1}^{K} F_j\right|} \right) \right], \tag{5.41}$$

$$\Pr[\text{State } F_1, F_2, ..., F_K] = \prod_{v=1}^{K} \sum_{r=0}^{|S_v|} \binom{|S_v|}{r} \frac{(-1)^r Q (1-f_{SD})^{(r+|F_v|)}}{1-(1-Q)(1-f_{SD})^{(|F_v|+r)}}, \tag{5.42}$$

$$\mathbb{E}[T_R] = \sum_{|F_1|=1}^{N} \sum_{|F_2|=1}^{N} ... \sum_{|F_K|=1}^{N} \binom{N}{|F_1|} \binom{N}{|F_2|} ... \binom{N}{|F_K|} \times$$
$$\mathbb{E}[T_R|\text{State } F_1, F_2, ...F_K] \Pr[\text{State } F_1, F_2, ..., F_K].$$

5.8 Appendix (B): Reducing the computations of the relaying with network coding protocol

As mentioned in Appendix (A), the expression for $\mathbb{E}[T_R|$ State $F_1, F_2, ..., F_K]$ in that case only depends on the cardinality of the union of the sets $F_1, ..., F_K$ given by $\left|\bigcup_{j=1}^{K} F_j\right|$ and not on the cardinalities of the individual sets. Let L_i be the

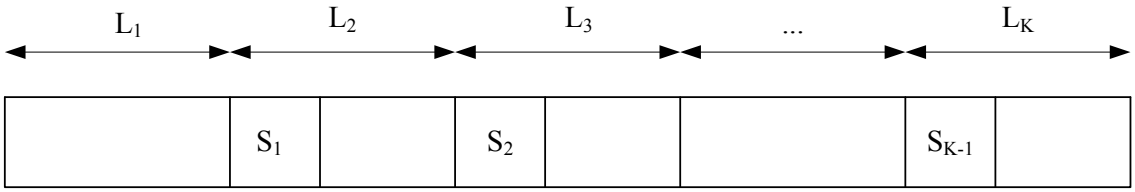

Figure 5.5: Illustration of sets $F_1, F_2, ..., F_K$.

cardinality of the set F_i, $i \in \{1, 2, ..., K\}$. Given $L_1, L_2, ..., L_K$, the cardinality of the union $\left|\bigcup_{j=1}^{K} F_j\right|$ lies between $\max(L_1, L_2, ..., L_K)$ and $\min(N, L_1 + L_2 + ... + L_K)$. By knowing the number of sets that lead to a union of a particular cardinality, we can sum over the possible cardinalities of the union of the sets which are N^K rather than summing over all possible set cardinalities which are $(2^N - 1)^K$,and thereby largely reducing the computations.

The contribution of this part is summarized in the following Lemma.

Lemma 5.1. *Given cardinalities* $0 \leq L_i \leq N$ *of the sets* F_i, $i \in \{1, 2, ..., K\}$, *the number of sets of all the possible sets with those cardinalities that has union set of cardinality* R *is*

$$\binom{N}{L_1} \sum_{s_1, s_2, ..., s_{K-2}} \binom{N-L_1}{s_1} \binom{L_1}{L_2 - s_1} \binom{N - L_1 - \sum_{i=1}^{K-2} s_i}{R - L_1 - \sum_{i=1}^{K-2} s_i} \binom{L_1 + \sum_{i=1}^{K-2} s_i}{L_K + L_1 + \sum_{i=1}^{K-2} s_i - R} \times \prod_{j=1}^{K-3} \binom{N - L_1 - \sum_{i=1}^{j} s_i}{s_{j+1}} \binom{L_1 + \sum_{i=1}^{j} s_i}{L_{j+2} - s_{j+1}},$$

where $\max(L_1, L_2, ..., L_K) \leq R \leq \min(N, L_1 + L_2 + ... + L_K)$, $\max(0, L_2 - L_1) \leq s_1 \leq \min(N - L_1, L_2)$ *and*

$$\max\left(0, L_{j+2} - L_1 - \sum_{i=1}^{j} s_i\right) \leq s_{j+1} \leq \min\left(N - L_1 - \sum_{i=1}^{j} s_i, L_{j+2}\right); 1 \leq j \leq K-4,$$

$$\max\left(0, L_{K-1}-L_1-\sum_{i=1}^{K-3}s_i, R-L_1-L_K-\sum_{i=1}^{K-3}s_i\right)\leq s_{K-2}$$
$$\leq \min\left(L_{K-1}, N-L_1-\sum_{i=1}^{K-3}s_i, R-L_1-\sum_{i=1}^{K-3}s_i\right).$$

Proof: Let S_j be the elements in the set F_{j+1} that do not belong to any of the sets $F_1, F_2, ..., F_j$, i.e., the innovation of the set. This means that for any set F_{j+1}, the elements $F_{j+1} \setminus S_j$ are repeated elements found in the previous sets. Suppose we order the sets as in Fig. 5.5 where for every set F_j, we divide the elements into the innovation S_{j-1} and the repeated elements. Denote by s_i the cardinality of the set S_i. Then, for the first set, we have $\binom{N}{L_1}$ possibilities. For the next set F_2, the innovation has $\binom{N-L_1}{s_1}$ possibilities and the repeated part has $\binom{L_1}{L_2-s_1}$ possibilities. The third set F_3 has $\binom{N-L_1-s_1}{s_2}$ possible innovations and $\binom{L+s_1}{L_3-s_2}$ possible repeated parts. Following the same arguments, we obtain the expression in Lemma 5.1. Note that for the last set, in order to have cardinality of the union equal to R, we should have R distinct elements per set which enforces that $s_{K-1} = R-L_1-s_1-...-s_{K-2} = R - L_1 - \sum_{i=1}^{K-2} s_i$. The total number of sets is obtained by summing over all the possible values of $s_1, s_2, ..., s_{K-2}$. The limits on the values of $s_1, s_2, ..., s_{K-2}$ can be easily obtained by expanding the binomial coefficients and setting the factorial terms to be non-negative. ■

Chapter 6: Opportunistic Access in Network-Coded Spectrum

6.1 Introduction

In this part, we study how an SU can exploit the structure of the PUs' idle and busy periods induced by batch processing systems such as network coding, to reliably learn the spectrum characteristics of the PUs and effectively mitigate spectrum sensing errors. We mainly focus on the effect of the spectrum predictability gain by considering perfect channels and show that it leads to throughput gains for both the PUs and the SU even if there is no spectrum availability gain (i.e., same fraction of idle slots available to the SU). The throughput achieved with only spectrum predictability gain is the worst-case throughput of the SU where spectrum availability gain exists as well.

Network coding has been studied for spectrum sensing purposes mainly in two directions. First, network coding can help efficiently disseminate control information among the SUs for collaborative spectrum sensing [82]. Second, the correlation among PU spectrum states due to network coding can be used by the SUs to track multiple PU channels (by assuming a busy slot will be more likely followed by another busy slot) and to identify an idle channel [83]. However, the model in [83] assumes that the SU has perfect sensing capability and can correctly distinguish

an idle slot from a busy one on the channel it chooses to sense. In this work, we study the practical case with possible spectrum sensing errors. While providing some degree of protection to the PUs, the SU pursues two objectives: (i) quickest detection of an idle slot and (ii) average throughput maximization. Since we are interested in the spectrum predictability gain, we consider perfect PU channels such that there is no spectrum availability gain (i.e., the fraction of the idle slot is the same independent of coding block size K).

For the quickest detection problem, we apply the Cumulative Summation (CUSUM) algorithm [84, 85] if the PUs' spectrum dynamics are unknown at the SU, and identify the potential benefit of network coding for spectrum predictability. Then, we use the Viterbi algorithm [86] to optimize the spectrum sensing performance if the PUs' spectrum dynamics are known. For the throughput maximization problem, we show that the spectrum predictability due to network coding applied at the PUs can actually improve the SU throughput (and the gain increases with K), even when the spectrum utilization remains the same. Our results show that the benefit of using network coding by the PUs is not limited to possible PUs' throughput gain, but also, if properly exploited, improves the spectrum sensing accuracy and increases the throughput for the SU by mitigating possible sensing errors. Hence, for the PUs, network coding represents a "self-protective" scheme against SU interference caused by sensing errors. This way, the overall spectrum efficiency of a cognitive radio network can be significantly improved.

To complement our analytical formulations and simulation studies, we evaluate our approach with testbed measurements. Our testbed results corroborate the

feasibility of our approach to mitigate spectrum sensing errors and to improve the SU throughput while protecting the PU transmissions.

The rest of the chapter is organized as follows. In Section 6.2, we introduce the system model for spectrum sensing on a PU channel. Section 6.3 and Section 6.4 study spectrum predictability gain under quickest idle slot detection and throughput maximization problems, respectively. We validate the results with testbed experiments in Section 6.5. Section 6.6 concludes the chapter.

6.2 System Model

We consider a cognitive radio network consisting of M primary users (PUs) and one secondary user (SU). The SU tries to detect the idle periods on the PUs' channels for opportunistic access. The M PUs occupy M orthogonal channels and their spectrum states (Idle/Busy) are assumed to evolve independently. Time is slotted with slot duration equal to one packet transmission duration. At the beginning of every slot, the SU chooses a channel to sense and then senses the spectrum (e.g., with an energy detector [16]) to detect whether that PU channel is idle or not. Based on channel sensing results (subject to detection errors) and possible prior knowledge about the traffic statistics of the PU, the SU decides on whether to transmit or not. We consider two goals for the SU, namely quickest detection of an idle slot, and throughput maximization, while guaranteeing some level of protection to the PUs' transmissions.

6.2.1 Network Coding and Traffic Model

Each PU receives (or generates) packets according to a stationary process and buffers them until K packets are accumulated in its queue. The PU server then codes each block of K packets linearly and transmits K coded packets. The PU and its receiver(s) agree on a set of linearly independent coding coefficients such that K successful transmissions are needed at a receiver to decode a block of K packets (alternatively, random network coding with sufficiently large field size is used). The state of each PU channel (idle/busy) is assumed to be fixed over a slot duration and varies between slots according to a Markov chain that models the correlation between the PU states. For PU systems without network coding, the busy/idle periods have been shown to follow a two-state Markov chain (shown in Fig. 6.1) through channel measurement [87] and this model has been widely used in spectrum sensing (e.g., [34, 35, 88]). Our model can be viewed as a natural generalization of this two-state model to network coding. For simplicity, we consider the case of a perfect PU channel that requires K transmissions (in K slots) to deliver K network-coded packets. In this case, the busy periods on the PU channels are multiple of K slots and the idle periods are on average K times longer which introduces K steps memory to the PU states. The PU state evolution over perfect channels can then be modeled by the Markov chain in Fig. 6.2 where λ and ν controls the level of correlation between the idle and busy states, respectively.

From the structure of the Markov chain, we expect that with larger values of K, the SU can more reliably detect the idle slots on the PU channels through

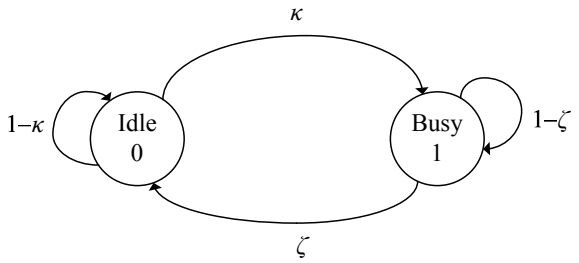

Figure 6.1: PU spectrum dynamics without network coding.

better tracking of the PUs' states, since it is more likely that a busy slot is followed by another busy slot due to the block transmission of coded packets. On the other hand, if $K = 1$, the busy/idle sequence, as modeled by the Markov chain in Fig. 6.1, involves less structure and in particular, for $\kappa = \zeta = 0.5$, it forms an i.i.d. sequence and hence there is no memory between the PU states that the SU can use to track the sequence.

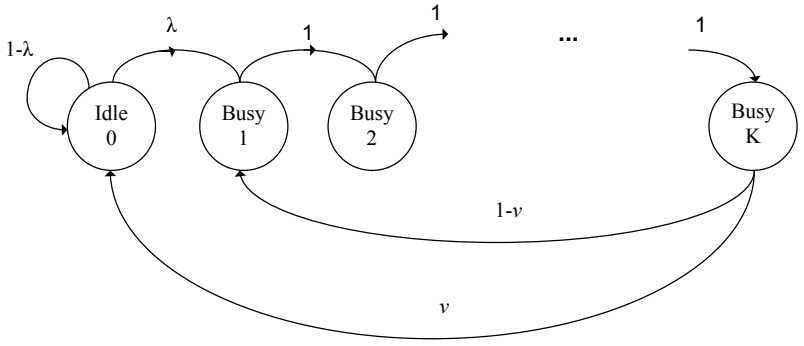

Figure 6.2: PU spectrum dynamics with network coding over perfect channels.

The idle state of the Markov chain in Fig. 6.2 has stationary probability $\pi_0 = \frac{\nu}{\nu+K\lambda}$ and the channel utilization is $u = 1 - \pi_0 = \frac{K\lambda}{\nu+K\lambda}$. We choose λ and ν to approximate the real behavior of a queue using network coding over multicast perfect channels. Through extensive simulations, it was found that choosing $\nu = 1 - \lambda$ well approximates the idle/busy evolution of network-coded transmissions for the

parameter values considered in this chapter. We seek to compare between two systems: the first evolves according to the Markov chain in Fig. 6.2 with $K = 1$, that is where the idle and busy periods can be of any length; and the second is for general K where once the PU becomes busy, it remains busy for K slots. Since we aim at studying the gain due to spectrum predictability, the channel utilization for all K should be the same. This can be achieved by choosing $\lambda = \frac{u}{u+K(1-u)}$. This generates the same fraction of idle and busy slots with busy periods a multiple of K slots. Although the spectrum availability to the SU is the same, we will show that the spectrum predictability gain of the PU spectrum achieved by network coding can be leveraged by the SU to improve the throughput over the case when the PU does not use network coding.

6.2.2 Channel Sensing Model

At the beginning of each slot, the SU chooses one of the M PU channels to sense and decides whether or not to transmit based on the current and possibly previous observations. Although our approach applies with any channel sensing scheme, here we specify the channel model and channel sensing scheme to be used for performance evaluation. At each slot, the SU observes n samples $\{Y_i\}_{i=1}^n$ which constitute the observations in that slot. Under hypothesis 1 (H_1), which corresponds to a busy slot, the observations $\{Y_i\}_{i=1}^n$ are i.i.d. $\sim \mathcal{N}(0, P + \sigma^2)$ and under hypothesis 0 (H_0), which corresponds to an idle slot, the observations $\{Y_i\}_{i=1}^n$ are i.i.d. $\sim \mathcal{N}(0, \sigma^2)$, where P denotes the average PU power received at the SU and σ^2 is

the noise variance. For simulations, we will use the Gaussian model of observations, while for testbed experiments in Section 6.5 we will use empirical distributions derived from real radio transmissions. We assume that the SU makes a hard decision based on energy detection [16]. The energy detector is used here for its simplicity as it does not require knowledge of the signal structures of the PUs' signals[1]. However, the SU is assumed to know the coding block size K used by the PUs. Since the block size is just one number and is fixed, it can be easily known at the SU either through some prior knowledge about the PU system or prior to the system operation through mutual communications, with negligible overhead, between the PU and the SU. One can also try to infer this number by observing PU's transmissions over a sufficient time interval as a part of learning the Markov chain structure [89]. For $Y_i \sim \mathcal{N}(0, \Sigma^2)$ (where $\Sigma^2 = \sigma^2$ under hypothesis H_0 and $\Sigma^2 = \sigma^2 + P$ under hypothesis H_1), the energy of the observed signals, $Z = \sum_{i=1}^{n} Y_i^2$, has probability density function given by

$$f_Z(z) = \frac{z^{(n/2)-1} e^{-z/(2\Sigma^2)}}{2^{n/2}\Sigma^n \Gamma(n/2)},\ z \geq 0, \tag{6.1}$$

where $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x}\, dx$ is the gamma function. For a given detection threshold τ, the misdetection (p_M) and false alarm (p_F) probabilities are given by

$$p_M = \Pr\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \leq \tau | H_1\right] = \frac{\gamma(\frac{n}{2}, \frac{n\tau}{2(\sigma^2+P)})}{\Gamma(n/2)}; p_F = \Pr\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \geq \tau | H_0\right] = \frac{\Gamma(\frac{n}{2}, \frac{n\tau}{2\sigma^2})}{\Gamma(n/2)}, \tag{6.2}$$

[1]By PU signal structure we mean the structure of the PU symbols such as the modulation used, the FEC used and other PHY-layer parameters required to decode the PU signal.

where $\gamma(n,x) = \int_0^x t^{n-1}e^{-t}\,dt$ and $\Gamma(n,x) = \int_x^\infty t^{n-1}e^{-t}\,dt$ are the lower and upper incomplete gamma functions, respectively. The detection threshold τ can be chosen to yield a tolerable misdetection probability for PU protection p_M, and then the false alarm probability p_F can be determined accordingly. We will show that the SU can mitigate channel sensing errors by tracking the PU spectrum dynamics, in contrast to the classical memoryless sensing strategies that cannot exploit the possible correlation of the PU states across slots. We will show in the next sections that this spectrum predictability gain becomes more significant as the coding block size increases.

We adopt the commonly used assumption that the PUs and the SU are perfectly synchronized. This can be achieved at the SU by overhearing the control signals of the PU whose channel is to be sensed. Despite being outside the scope of this work, we expect that synchronization with a PU using network coding is easier since there are less frequent and more regular transitions of the PU states. With network coding, the time scale of the change of the PU states is increased by K. This conjecture is left for future investigations.

6.3 SU Objective 1: Quickest Detection of an Idle Slot

The first objective for the SU is to minimize the expected time to detect an idle slot given that the initial state of the chain is randomly chosen according to its stationary distribution. Throughout this section, we restrict ourselves to the case of a single PU (i.e., $M = 1$) and show the benefit of using network coding at the PU

(i.e., larger K). It should be noted that the procedure described in this section can be applied to the case of multiple PUs if the SU first chooses a channel to sense, runs the procedure described here on the chosen channel until detecting an idle slot and then repeats the cycle. We distinguish two cases depending on whether or not the SU knows the PU traffic dynamics, namely the parameters of the PU Markov chain in Fig. 6.2.

6.3.1 Unknown PU Spectrum Dynamics

First, we assume that the SU does not know the parameters of the Markov chain for the underlying PU spectrum dynamics. The Cumulative Summation (CUSUM) algorithm is the optimal scheme to detect a single change in distribution based on sequential observations [84, 85] when the distribution of the change point is not known. After a change from H_1 (busy state) to H_0 (idle state), the CUSUM algorithm minimizes the expected delay to capture the change (detecting an idle slot) under some constraint on the average duration between times of detecting H_1 as H_0 (misdetection). If that duration is large, the PU is highly protected but this implies that the expected duration for detecting an idle slot is also large, and vice versa. The CUSUM algorithm works by accumulating energy over consecutive slots in contrast to memoryless energy detectors that are restricted to a single slot.

Under the assumption that the expected time to detect a change is much shorter than the time scale for the change to occur (and hence a single change in distribution occurs), the CUSUM algorithm was applied to spectrum sensing with

a single change in [90, 91], and quickest detection was formulated in the framework of POMDPs in [88], which requires the knowledge of the PU traffic dynamics at the SU. However, due to the quick dynamics of the PU spectrum, the assumption of a single change is no longer valid in our case since multiple transitions between idle and busy states might occur before a decision could be made. In general, the quickest change detector in the case of multiple changes is unknown. Therefore, we apply the CUSUM algorithm without any optimality claim, and show the performance gain it achieves. The CUSUM algorithm only requires the knowledge of the distribution of the observations under both hypotheses and does not need the knowledge of the traffic parameters of the PU, which may take a long time to be learned at the SU.

Define f_0 and f_1 as the probability density functions of one sample Y under hypotheses H_0 and H_1, respectively. The CUSUM algorithm is given as

Algorithm 1 CUSUM Algorithm

1: Initialize $k = 0$ and $S_0 = 0$.

2: Increment k and compute the log-likelihood ratio (LLR) $Z_k = \sum_{i=1}^{n} \log\left[\frac{f_0(Y_{i,k})}{f_1(Y_{i,k})}\right]$ for each slot k, where $Y_{i,k}$ is the ith observation in the kth slot and n is the number of samples per slot. The LLR is given by $\frac{n}{2}\log\left(1+\frac{P}{\sigma^2}\right) - \frac{P}{2\sigma^2(P+\sigma^2)}\left(\sum_{i=1}^{n} Y_{i,k}^2\right)$ for i.i.d. Gaussian observations.

3: Compute $S_k = \max\{0, S_{k-1} + Z_k\}$. If $S_k > h$ for some threshold h, a decision that hypothesis H_0 exists (i.e. idle slot) is made and the algorithm stops; else, go to Step 2.

The rationale behind this algorithm is that $\mathbb{E}_{H_0}[\text{LLR}] = D(f_0||f_1) > 0$ while

$\mathbb{E}_{H_1}[\text{LLR}] = -D(f_1||f_0) < 0$, where $D(f_0||f_1)$ is the Kullback-Leibler distance between the densities f_0 and f_1. Hence, the sequence $\{S_k\}$ will have positive drift under H_0 and negative drift under H_1. Define X_m as the true state of the system in slot m and T_d as the time until a correct decision is made. Then, T_d is given by $T_d = \inf\{m : S_m > h \text{ and } X_m = 0 \text{ (“Idle”)}\}$. The threshold h needs to be chosen such that $\mathbb{E}[T_d]$ is small (i.e., h is not too large) and misdetection rate[2] $\widetilde{p}_M$ is small (i.e., h is not too small).[3] By varying the threshold h and computing the corresponding $\mathbb{E}[T_d]$ and $\widetilde{p}_M$, we provide simulation results for $\mathbb{E}[T_d]$ versus $\widetilde{p}_M$ for the CUSUM algorithm. We expect that the performance of the CUSUM gets better with K, since multiple consecutive changes lead to energy accumulation and hence the test statistics hits the threshold faster.

In numerical results, we focus on low SNR PU signal observed at the SU. Unless stated otherwise, the average received signal power at the SU is $P = 1$ and the noise variance is $\sigma^2 = 1$ (i.e., SNR=0 dB). The sensor at the SU uses $n = 10$ channel sensing samples per slot with $p_M = 0.2$ which corresponds to a false alarm probability $p_F = 0.2618$. We fix the channel utilization to $u = 0.5$ by properly selecting λ. This way, we can evaluate the spectrum predictability gain of NC by separating the possible gain of extending the SU's spectrum availability. This models the worst-case throughput of the SU where spectrum availability gain exists

[2]Defined as the inverse of the average duration between misdetections.

[3]The exact calculation of $\mathbb{E}[T_d]$ and $\widetilde{p}_M$ as a function of the threshold h involves solving Fredholm integral equation which is a difficult task, even for simple cases [92]. For Markovian evolution of the idle and busy slots, these quantities can be obtained by Monte-Carlo simulations [92].

as well.

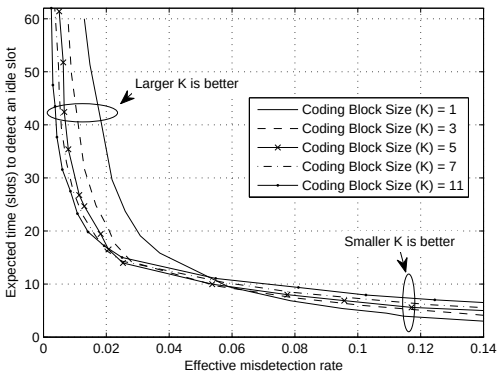

Figure 6.3: Quickest detection performance of the CUSUM algorithm.

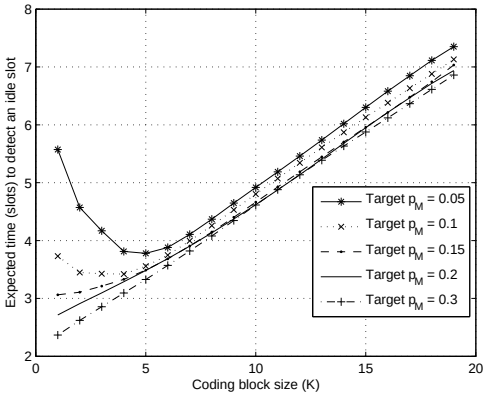

Figure 6.4: Quickest detection performance of the Viterbi algorithm.

Figure 6.3 shows the expected time to detect an idle slot for the CUSUM algorithm, where we vary the threshold h and compute $\mathbb{E}[T_d]$ and the corresponding $\widetilde{p}_M$. For the MC shown in Fig. 6.2 and under perfect sensing assumption ($p_M = p_F = 0$), the expected time needed to capture an idle slot is

$$\mathbb{E}[T_d] = \frac{K\lambda}{K\lambda + \nu}\left[\frac{K}{2} + \frac{2-\nu}{2\nu}\right], \tag{6.3}$$

which increases with K for the same u and ν. Equation (6.3) follows from the first passage time of the MC in Fig. 6.2; and it reveals that with perfect sensing, due to the long busy periods (of length K), the SU needs longer time to detect an idle slot as K increases, thus a small value of K is preferred for the SU. If sensing is imperfect, there are two opposite effects taking place when K increases. The spectrum predictability improves but in the meantime it takes longer until the MC changes to the idle state. If the target $\widetilde{p}_M$ is large (i.e., low PU protection), the SU does not need reliable channel sensing and a small value of K is still better

for the SU, since the gain associated with spectrum predictability is not needed. For small target $\widetilde{p}_M$, the spectrum predictability gain by increasing the value of K compensates for the increase in $\mathbb{E}[T_d]$; therefore, a larger K is preferred. These opposite effects are shown in Fig. 6.3, where a low (or high) $\widetilde{p}_M$ corresponds to a high (or low) threshold. In that case, the benefit of increasing K is due to the energy accumulation over correlated slots in the CUSUM algorithm and thus the test statistic can hit the required threshold faster.

6.3.2 Known PU Spectrum Dynamics

In this part, we assume that, in addition to the distribution of the observations under different PU states, the SU knows the parameters of the Markov chain in Fig. 6.2 for the underlying PU spectrum dynamics. This knowledge can be acquired offline from measurements through learning algorithms such as the Baum-Welch algorithm [93, 94]. Subsequently, we assume that the parameters of the Hidden Markov Process (HMP) are perfectly known to the SU. Hidden Markov Models (HMMs) have been used to learn and track the spectrum availability in [35, 87, 95]. However, all these works focused on the two-state PU dynamics as in Fig. 6.1, that cannot be used to address the spectrum availability/predictability gains due to network coding. After learning the parameters of the PU Markov chain, the SU tracks the state of the PU spectrum using the Viterbi algorithm [86], which is optimal for estimating the most likely PU state sequence given all the previous observations. From the estimated PU state sequence, the SU can estimate the current PU state.

Protection to the PU is guaranteed by using a spectrum sensor with misdetection probability p_M equal to the target collision probability p_C as in [34]. Note that by such a choice the actual collision probability $\hat{p}_C$ satisfies $\hat{p}_C \leq p_M$, if the SU uses memoryless spectrum sensing, and $\hat{p}_C < p_M$, if the SU leverages the memory through the Viterbi algorithm.

In the following, we illustrate the performance of the Viterbi algorithm for quickest detection and compare it with the CUSUM algorithm. Figure 6.4 shows the expected time for detecting an idle slot as a function of the coding block size K. As shown in Fig. 6.4, for small target misdetection probabilities (e.g., $p_M = 0.05$), it is not necessarily true that a larger K value is preferred (as in CUSUM algorithm). For instance, the case with $p_M = 0.05$ has the smallest value of $\mathbb{E}[T_d]$ at $K = 5$. This can be explained by the fact that the benefit of the correlation of the PU states beyond $K = 5$ does not compensate for the increase in $\mathbb{E}[T_d]$ due to the increase in the duration of the busy period. For the CUSUM algorithm (Fig. 6.3) with $p_M = 0.05$, the minimum value of $\mathbb{E}[T_d]$ is 11 and achieved at $K = 5$, while for $p_M = 0.1$ the minimum value of $\mathbb{E}[T_d]$ is 5 and achieved at $K = 1$. For the same values of K, the Viterbi algorithm achieves a lower value of $\mathbb{E}[T_d]$. This can be seen in Fig. 6.4, where for $p_M = 0.05$ and $K = 5$ we have $\mathbb{E}[T_d] = 3.75$, whereas for $p_M = 0.1$ and $K = 1$ we have $\mathbb{E}[T_d] = 3.7$. These expected time values are lower than those achieved by the CUSUM algorithm. Moreover, the Viterbi algorithm achieves a lower value for the minimum $\mathbb{E}[T_d]$ than the CUSUM algorithm for the same misdetection probability, i.e., for $p_M = 0.05$ and $p_M = 0.1$ the Viterbi algorithm achieves the minimum values of $\mathbb{E}[T_d] = 3.75$ and $\mathbb{E}[T_d] = 3.4$ at $K = 5$ and $K = 1$, respectively. This shows the

advantage of exploiting the knowledge of the PU spectrum dynamics at the SU to detect idle slots on the PU channels.

6.4 SU Objective 2: Average Throughput Maximization

The second objective for the SU is to maximize the average throughput, namely the average rate of detecting idle slots on the PUs' channels. We first study the learning phase, then we focus on the tracking phase, where the SU actively tracks the channels for opportunistic access. We consider both the optimal POMDP solution for a single PU in Section 6.4.2, a suboptimal greedy policy for a single PU in Section 6.4.3 and then we extend to the case of multiple PUs in Section 6.4.4.

6.4.1 Learning Phase

For the learning phase, we focus on learning one particular PU channel; but the procedure is repeated for all the channels until the SU has knowledge about the traffic parameters of all the PUs. The SU learns the Markov chain parameters given the coding block size K used by the PU. Define N as the number of slots over which the SU observes the PU chain for learning and $\mathcal{S} = \{0, 1, ..., K\}$ as the state space of the PU Markov chain. The true sequence of states $s_1^N = \{s_t \in \mathcal{S} | t = 1, 2, ..., N\}$ is hidden to the SU but a sequence of corresponding sensing outcomes $y_1^N = \{y_t \in \mathcal{Y} | t = 1, 2, ..., N\}$ is available to the SU, where $\mathcal{Y} =$ {"Idle", "Busy"}.

Given only the observation sequence, the Baum-Welch algorithm generates a sequence of parameter estimates of non-decreasing likelihood values for the Hidden

Markov Process (HMP). Define $\hat{\eta}_r = (\hat{\pi}_r, \hat{A}_r)$ as the estimate of the parameters of the Markov chain at the rth iteration of the algorithm, where $\hat{\pi}_r$ is the estimated initial distribution of the chain and $\hat{A}_r = [\hat{a}_{ij}]_r$ is the estimated state transition matrix. The algorithm starts with an initial guess $\hat{\eta}_0 = (\hat{\pi}_0, \hat{A}_0)$ and then updates the parameter estimates by maximizing the likelihood given the observation sequence $\{y_1^N\}$. The rth iteration starts with an estimate $\hat{\eta}_{r-1}$ and estimates a new parameter set $\hat{\eta}_r$ according to

$$\hat{\eta}_r = \arg\max_{\bar{\eta_r}} \sum_{s_1^N} P_{\hat{\eta}_{r-1}}\left(s_1^N | y_1^N\right) \ln\left[P_{\bar{\eta_r}}\left(s_1^N, y_1^N\right)\right], \tag{6.4}$$

where $P_{\hat{\eta}_{r-1}}\left(s_1^N | y_1^N\right)$ is the probability of the state sequence s_1^N given the observation sequence y_1^N under model estimate $\hat{\eta}_{r-1}$, and $\bar{\eta_r}$ is the set of the feasible parameters at the rth iteration. The algorithm terminates when a convergence criterion is satisfied, e.g., when $\ln P_{\hat{\eta}_r}(y_1^N) - \ln P_{\hat{\eta}_{r-1}}(y_1^N) < \varepsilon$ for a given threshold ε. Note that although the Baum-Welch algorithm is guaranteed to converge, it might converge to a local optimum and therefore different initial guesses may be needed.

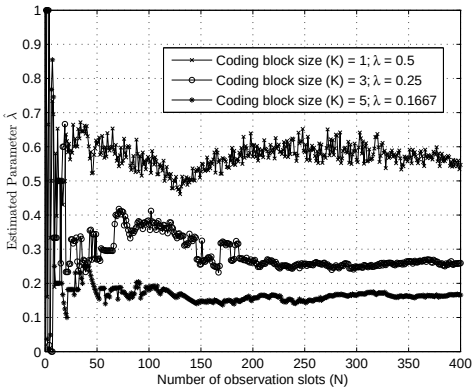

Figure 6.5: Estimated parameter $\hat{\lambda}$ vs. N.

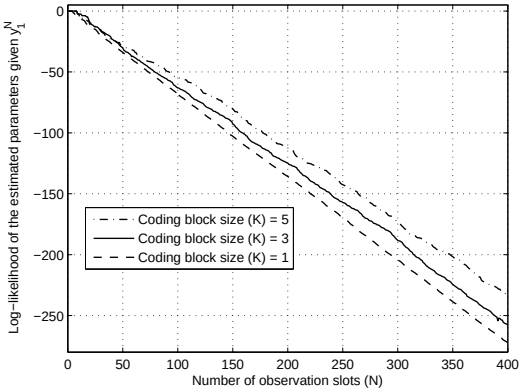

Figure 6.6: Log-likelihood of $\hat{\eta}$ vs. N.

We give an example in Fig. 6.5 for estimating the value of λ with threshold

$\varepsilon = 10^{-4}$. Figure 6.5 shows that a higher value of K leads to a better estimate of λ for the same number of observations. For $K = 1$ and $K = 3$, the estimate $\hat{\lambda}$ does not exactly match the true values of λ, which are 0.5 and 0.25, respectively, for observation vector lengths of up to 400 slots, while it matches the true value of λ for $K = 5$. Figure 6.6 shows the log-likelihood of the estimated parameters given the observation sequence y_1^N for different number of observation slots. For the same number of observations, the Baum-Welch algorithm achieves higher log-likelihoods for higher values of K and this shows that the PU spectrum structure due to network coding improves the estimate of the PU Markov chain parameters, which we use in the next section for better tracking of the PU spectrum.

6.4.2 Tracking Phase - Single PU - Optimal Policy

We first consider a single PU whose Markov chain evolves as in Fig. 6.2. We assume that the SU perfectly knows the parameters of the PU Markov chain. At the beginning of each slot, the PU Markov chain makes a state transition, then the SU senses the channel, updates its belief vector (to be defined shortly) and then chooses between two actions: to transmit or to remain silent. With rewards incurred for different SU actions and different PU states, the optimal access policy can be found through a POMDP formulation. The SU maintains a belief vector about the state of the Markov chain for the PU spectrum. Each component of the belief vector represents the conditional probability that the PU Markov chain is in a certain state given the decision and observation history. The belief vector is updated based on

the sensing outcome and the feedback received upon transmission. Note that we can incorporate the case of time varying channels modeled as finite state chains (e.g., Gilbert-Elliott) with imperfect channel state information at the SU by adding the channel state as states to the belief vectors at the expense of more complexity in computing the optimal access policy.

The existence of an optimal policy for infinite-horizon POMDP problems is undecidable [96], hence we focus on numerical solutions obtained by discretizing the belief space. The problem of maximizing the SU throughput subject to some PU protection constraint (e.g., misdetection probability p_M) leads to a constrained POMDP. For constrained POMDPs with discretized belief space, randomized policies may be needed for optimality, while there always exists an optimal deterministic policy if they are unconstrained [Theorem 6.2.10 in [97]], [98]. In order to bypass the difficulties encountered in solving the constrained POMDP problem, we formulate the problem as an unconstrained POMDP where the PU is protected by using a reward function at the SU that is a weighted sum of the PU and SU throughputs. By adjusting the weight, the PU is supported with different throughput values. We discuss this weighted throughput formulation in the reward part.

As typically assumed in similar models (e.g., [33,34]), we assume that the PU's traffic dynamics are independent of the actions taken at the SU and we consider the extended Markov chain structure in Fig. 6.2 to represent network coding effects. We denote by $X_t \in \{0, 1, ..., K\}$ the state of the PU in time slot t.

6.4.2.1 Actions

Two actions are possible at the SU in each slot t, to remain silent ($A_t = 0$) or to transmit ($A_t = 1$).

6.4.2.2 Rewards

When the PU spectrum is in state X_t and action A_t is taken by the SU, the reward is given by

$$R(X_t, A_t) = \begin{cases} 0, & \text{if } X_t = 0,\ A_t = 0, \\ wr_{P,1}, & \text{if } X_t \neq 0,\ A_t = 0, \\ (1-w)r_{S,1}, & \text{if } X_t = 0,\ A_t = 1, \\ wr_{P,2} + (1-w)r_{S,2}, & \text{if } X_t \neq 0,\ A_t = 1, \end{cases} \tag{6.5}$$

where $r_{P,1}$ and $r_{S,1}$ are the rates achieved by the PU and SU, respectively, when they do not interfere, and $r_{P,2}$ and $r_{S,2}$ are the rates achieved by the PU and SU, respectively, when they interfere with each other. The weight w represents the relative importance of the PU throughput and is used for the PU protection [99]. Choosing $w = 1$ gives full priority to the PU throughput (i.e., full PU protection), while $w = 0$ favors the SU throughput (i.e., no PU protection). By varying w between 0 and 1, we can reach different degrees of PU protection corresponding to different PU throughputs. In the following, we assume $r_{P,2} = 0$ and $r_{S,2} = 0$, that is, when both the PU and the SU transmit simultaneously, a collision occurs and both packets are lost. However, a similar approach can be taken for the more

general case with multi-packet reception capability. Also, for our packet based model, $r_{P,1} = r_{S,1} = 1$ (coded) packet/slot.

Note that if the PU codes over K packets, a lost coded packet (due to collision) may prevent the PU receiver from decoding the entire block of coded packets. This issue can be overcome if the PU codes over $K-1$ packets, transmits this block, and then at the Kth transmission, either retransmits one of these coded packets if a collision occurred, or else transmits the left uncoded packet. As we will see, the optimal access policy at the SU guarantees $K-1$ collision-free transmissions out of K PU transmissions, and hence at least $K-1$ PU coded packets can be successfully delivered during every PU busy period. This validates Eq. (6.5) where the PU has one packet delivered whenever it is busy and no collision occurs.

6.4.2.3 Spectrum Sensing

Although some level of PU protection can be achieved solely by adjusting w in the reward function, spectrum sensing leads to better inference of the PU spectrum state and consequently better SU throughput for the same level of PU protection. We assume that the spectrum sensing scheme has a misdetection probability p_M, which corresponds to some false alarm probability p_F.

6.4.2.4 Channel Feedback from SU receiver

If the SU chooses to transmit, an error-free feedback message is sent from the SU receiver to the SU transmitter indicating whether the packet was successfully

received or not. This feedback message also reveals about the idle/busy state of PU spectrum Specifically, under the collision channel model considered, if an ACK (or NACK) is received at the end of a slot, the SU learns that the PU was idle (or busy) during that slot. This feedback is then used by the SU for updating the belief in the following slot.

6.4.2.5 Observations and Belief Vector

Since the true state of the PU spectrum cannot be exactly observed from channel sensing results because of possible sensing errors, the SU maintains a belief about the state of the PU. For POMDP problems, the belief is a sufficient statistic for deciding on the action given all the past observations and actions [100]. Given the spectrum sensing observation and the transmission feedback in each slot, the SU updates its belief regarding the state of the PU. We denote by $\mathbf{\Lambda}_t$ the $(K+1) \times 1$ belief vector of the PU state in time slot t, where the mth component $\mathbf{\Lambda}_t(m)$ denotes the belief in time slot t that the Markov chain of the PU (Fig. 6.2) is in state m, where $0 \leq m \leq K$. Note that the first component of the belief vector is $\mathbf{\Lambda}_t(0)$.

(a) Under the action (0): The SU chooses not to transmit. No channel feedback is observed and the belief in the following slot is updated solely based on the channel sensing outcome in that slot. That is, the observations are either "Busy" or "Idle". Given a belief vector $\mathbf{\Lambda}_t$, the probability of observing the outcome "Busy" or "Idle"

in slot $t+1$ are given by

$$\Pr\left[\text{Busy}|\mathbf{\Lambda}_t\right] = \sum_{m=0}^{K} \Pr\left[X_t = m|\mathbf{\Lambda}_t\right] \Pr\left[\text{Busy}|\mathbf{\Lambda}_t, X_t = m\right] = \mathbf{\Lambda}_t(0)[\lambda(1-p_M)+(1-\lambda)p_F] + \mathbf{\Lambda}_t(K)\left[\lambda(1-p_M)+(1-\lambda)p_F\right] + (1-p_M)\left[\sum_{m=1}^{K-1} \mathbf{\Lambda}_t(m)\right], \tag{6.6}$$

$$\Pr\left[\text{Idle}|\mathbf{\Lambda}_t\right] = \mathbf{\Lambda}_t(0)\left[\lambda p_M + (1-\lambda)(1-p_F)\right] + \mathbf{\Lambda}_t(K)\left[\lambda p_M + (1-\lambda)(1-p_F)\right] + p_M\left[\sum_{m=1}^{K-1} \mathbf{\Lambda}_t(m)\right]. \tag{6.7}$$

The belief update under action $A_t = 0$ and observation $O(A_t) = \text{Idle}$ is given by

$$\begin{aligned}\mathbf{\Lambda}_{t+1}(m) &= \Pr\left[X_{t+1} = m|A_t = 0, \mathbf{\Lambda}_t, O(A_t) = \text{Idle}\right] \\ &= \frac{\Pr\left[O(A_t) = \text{Idle}|A_t = 0, \mathbf{\Lambda}_t, X_{t+1} = m\right] \Pr\left[X_{t+1} = m|A_t = 0, \mathbf{\Lambda}_t\right]}{\Pr\left[O(A_t) = \text{Idle}|A_t = 0, \mathbf{\Lambda}_t\right]} \\ &= \frac{\left[(1-p_F)\mathbb{1}[m=0] + p_M\mathbb{1}[m\neq 0]\right]\Gamma_m}{\sum_{m=0}^{K}\left[(1-p_F)\mathbb{1}[m=0] + p_M\mathbb{1}[m\neq 0]\right]\Gamma_m},\end{aligned} \tag{6.8}$$

where $\mathbb{1}[\bullet]$ is the indicator function and

$$\Gamma_m = \sum_{i=1}^{K-1} \mathbf{\Lambda}_t(i)\mathbb{1}[m = i+1] + \mathbb{1}[m=0]\left[\mathbf{\Lambda}_t(0)(1-\lambda) + \mathbf{\Lambda}_t(K)(1-\lambda)\right] + \mathbb{1}[m=1]\left[\mathbf{\Lambda}_t(0)\lambda + \mathbf{\Lambda}_t(K)\lambda\right]. \tag{6.9}$$

Under observation $O(A_t) = \text{Busy}$, the belief is updated according to

$$\begin{aligned}\mathbf{\Lambda}_{t+1}(m) &= \Pr\left[X_{t+1} = m|A_t = 0, \mathbf{\Lambda}_t, O(A_t) = \text{Busy}\right] \\ &= \frac{\Pr\left[O(A_t) = \text{Busy}|A_t = 0, \mathbf{\Lambda}_t, X_{t+1} = m\right] \Pr\left[X_{t+1} = m|A_t = 0, \mathbf{\Lambda}_t\right]}{\Pr\left[O(A_t) = \text{Busy}|A_t = 0, \mathbf{\Lambda}_t\right]} \\ &= \frac{\left[p_F\mathbb{1}[m=0] + (1-p_M)\mathbb{1}[m\neq 0]\right]\Gamma_m}{\sum_{m=0}^{K}\left[p_F\mathbb{1}[m=0] + (1-p_M)\mathbb{1}[m\neq 0]\right]\Gamma_m}.\end{aligned} \tag{6.10}$$

(b) Under the action (1): The SU chooses to transmit. An (ACK/NACK) feedback is sent from the SU receiver to the SU transmitter over a dedicated control channel at

the end of the slot. The possible observations in this case are (ACK, Busy), (ACK, Idle), (NACK, Busy) and (NACK, Idle). The (ACK/NACK) feedback is observed at the end of slot t, while the "Busy" or "Idle" outcome is observed after sensing in slot $t+1$. The belief is then updated. The probabilities of these observations under action $(A_t = 1)$ are given by

$$\begin{aligned}
&\Pr\left[(\text{ACK}, \text{Busy})|\mathbf{\Lambda}_t\right] = \mathbf{\Lambda}_t(0)\left[\lambda(1-p_M) + (1-\lambda)p_F\right], \\
&\Pr\left[(\text{NACK}, \text{Busy})|\mathbf{\Lambda}_t\right] = (1-p_M)\left[\sum_{m=1}^{K-1}\mathbf{\Lambda}_t(m)\right] + \mathbf{\Lambda}_t(K)\left[\lambda(1-p_M) + (1-\lambda)p_F\right], \\
&\Pr\left[(\text{ACK}, \text{Idle})|\mathbf{\Lambda}_t\right] = \mathbf{\Lambda}_t(0)\left[\lambda p_M + (1-\lambda)(1-p_F)\right], \\
&\Pr\left[(\text{NACK}, \text{Idle})|\mathbf{\Lambda}_t\right] = p_M\left[\sum_{m=1}^{K-1}\mathbf{\Lambda}_t(m)\right] + \mathbf{\Lambda}_t(K)\left[\lambda p_M + (1-\lambda)(1-p_F)\right]. \quad (6.11)
\end{aligned}$$

The belief is updated as follows.

(i) If $O(A_t) =$ (ACK, Busy):

$$\begin{aligned}
\mathbf{\Lambda}_{t+1}(m) &= \Pr\left[X_{t+1} = m | A_t = 1, \mathbf{\Lambda}_t, O(A_t) = (\text{ACK,Busy})\right] \\
&= \frac{\left[(1-p_M)\mathbb{1}[m=1] + p_F\mathbb{1}[m=0]\right]\left[\lambda\mathbb{1}[m=1] + (1-\lambda)\mathbb{1}[m=0]\right]}{p_F(1-\lambda) + \lambda(1-p_M)}.
\end{aligned} \quad (6.12)$$

(ii) If $O(A_t) =$ (NACK, Busy):

$$\begin{aligned}
\mathbf{\Lambda}_{t+1}(m) &= \Pr\left[X_{t+1} = m | A_t = 1, \mathbf{\Lambda}_t, O(A_t) = (\text{NACK,Busy})\right] \\
&= \frac{\left[(1-p_M)\mathbb{1}[m\neq 0] + p_F\mathbb{1}[m=0]\right]\Psi_m}{\sum_{m=0}^{K}\left[p_F\mathbb{1}[m=0] + (1-p_M)\mathbb{1}[m\neq 0]\right]\Psi_m},
\end{aligned} \quad (6.13)$$

where $\Psi_m = \sum_{i=1}^{K-1} \mathbf{\Lambda}_t(i)\mathbb{1}[m = i+1] + \mathbb{1}[m = 0]\mathbf{\Lambda}_t(K)(1-\lambda) + \mathbb{1}[m = 1]\mathbf{\Lambda}_t(K)\lambda$.

(*iii*) If $O(A_t) =$ (ACK, Idle):

$$\begin{aligned}\mathbf{\Lambda}_{t+1}(m) &= \Pr\left[X_{t+1} = m | A_t = 1, \mathbf{\Lambda}_t, O(A_t) = (\text{ACK,Idle})\right] \\ &= \frac{\left[p_M \mathbb{1}[m=1] + (1-p_F)\mathbb{1}[m=0]\right]\left[\lambda\mathbb{1}[m=1] + (1-\lambda)\mathbb{1}[m=0]\right]}{(1-p_F)(1-\lambda) + \lambda p_M}.\end{aligned} \tag{6.14}$$

(*iv*) If $O(A_t) =$ (NACK, Idle):

$$\begin{aligned}\mathbf{\Lambda}_{t+1}(m) &= \Pr\left[X_{t+1} = m | A_t = 1, \mathbf{\Lambda}_t, O(A_t) = (\text{NACK,Idle})\right] \\ &= \frac{\left[p_M \mathbb{1}[m \neq 0] + (1-p_F)\mathbb{1}[m=0]\right]\Psi_m}{\sum_{m=0}^{K}\left[(1-p_F)\mathbb{1}[m=0] + p_M\mathbb{1}[m\neq 0]\right]\Psi_m}.\end{aligned} \tag{6.15}$$

6.4.2.6 Policy

The SU policy is a mapping from the belief space to the action space $\{0, 1\}$. The optimal policy maximizes the expected discounted reward and is given by

$$\pi^* = \arg\max_{\pi} \mathbb{E}_\pi \left[\sum_{t=1}^{\infty} \xi^t R\left(X_t, A_t\right) | \mathbf{\Lambda}_1\right], \tag{6.16}$$

where ξ is the discount factor ($\xi < 1$), which describes the importance of the future reward relative to the immediate reward, and $\mathbf{\Lambda}_1$ is the initial belief vector, which is set to the stationary distribution of the chain.

Let $V(\mathbf{\Lambda}_t)$ denote the value function, which is defined as the maximum expected reward that can be incurred starting from time slot t given belief $\mathbf{\Lambda}_t$. The value function must satisfy the Bellman equation in dynamic programming for all t

$$V(\mathbf{\Lambda}_t) = \max_{A_t \in \{0,1\}} \left[R_{A_t}(\mathbf{\Lambda}_t) + \xi \sum_{O(A_t)} \Pr\left[O(A_t)|\mathbf{\Lambda}_t\right] V\left(\Phi(\mathbf{\Lambda}_t | A_t, O(A_t))\right)\right], \tag{6.17}$$

where $R_{A_t}(\mathbf{\Lambda}_t)$ is the expected immediate reward in slot t under action A_t, $O(A_t)$ represents the observation under action A_t and the function $\Phi(\mathbf{\Lambda}_t|A_t, O(A_t))$ represents the belief update under action A_t and observations $O(A_t)$. For action $A_t \in \{0, 1\}$, $R_{A_t}(\mathbf{\Lambda}_t)$ is given by

$$R_0(\mathbf{\Lambda}_t) = wr_{P,1}(1 - \mathbf{\Lambda}_t(0)) \text{ and } R_1(\mathbf{\Lambda}_t) = (1 - w)r_{S,1}\mathbf{\Lambda}_t(0), \tag{6.18}$$

where $\mathbf{\Lambda}_t(0)$ is the first component of the belief vector $\mathbf{\Lambda}_t$, which represents the belief that the PU state is idle.

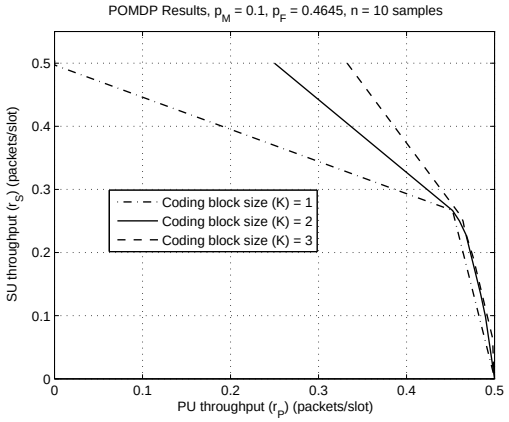

Figure 6.7: Throughput of the POMDP optimal policy (smaller p_M, larger p_F).

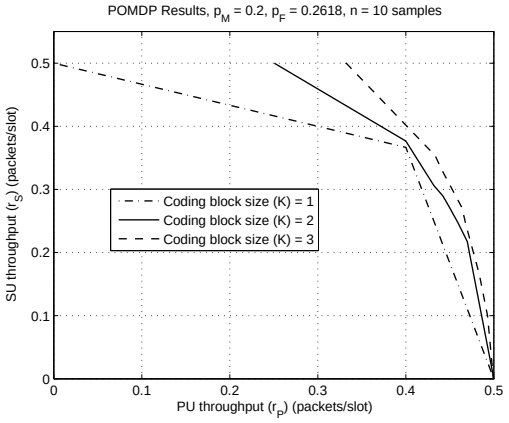

Figure 6.8: Throughput of the POMDP optimal policy (larger p_M, smaller p_F).

POMDPs with discretized belief space are PSPACE-hard problems [101]. Here, we limit ourselves to small values of K and compute the optimal policy using the value iteration algorithm applied to a discrete finite uniform grid in the belief space[4].

[4]Although many POMDP solvers are available, we developed our own solver tailored to the specific needs of our problem such as (i) the belief is updated based on observations depending on two different states of the PU (the ACK/NACK in a slot and the sensing outcome in the following slot), and (ii) we compute the average PU and SU throughputs rather than the expected reward.

Once the optimal SU transmission policy is computed, we run a system simulation to compute the PU and SU throughputs. We consider the suboptimal greedy policy which requires less complexity in Section 6.4.3.

In Figs. 6.7 and 6.8, we vary the PU protection weight parameter $w \in [0, 1]$ and show the relationship between the SU throughput, r_S, and PU throughput, r_P, jointly achieved under the POMDP formulation for fixed channel utilization $u = 0.5$ and discount factor $\xi = 0.9$. For the same value of r_P, the SU throughput r_S increases with coding block size K. For r_S to reach its maximum possible value 0.5, the PU throughput r_P must drop to zero when $K = 1$, while for $K = 2$ or $K = 3$ the PU can still sustain a non-zero throughput r_P with $r_S = 0.5$. This can be explained as follows. For $u = 0.5$ and $K = 1$ ($\lambda = 0.5$), the busy/idle states of the PU spectrum form an i.i.d. sequence and hence the spectrum cannot be predicted at the SU. For this case, three possible transmission strategies at the SU are as follows. The first strategy is to transmit in all slots irrespective of the PU and this corresponds to $r_S = 1 - u = 0.5$ and $r_P = 0$ due to continuous collisions. The second strategy is to trust its sensing outcome in each slot and this corresponds to $r_S = (1 - u)(1 - p_F)$ and $r_P = u(1 - p_M)$ and is given by the breaking point on the curve of $K = 1$. The third strategy is to remain silent at all slots and corresponds $r_P = u = 0.5$ and $r_S = 0$. By time sharing between the first and second strategies or between the second and third strategies, the points on the linear parts on the curve of $K = 1$ in Fig. 6.7 and 6.8 can be achieved. On the other hand, increasing K to 2 or 3 introduces more correlation (hence more memory) to the PU states and hence the Markov chain becomes more amenable to be tracked at

the SU facing the sensing errors. For the SU to achieve the maximum throughput $r_S = 1 - u$, it can keep transmitting until receiving a collision. Then, it remains silent for the following $K-1$ slots since the PU will be busy in these slots and hence SU packets, if transmitted, would be lost by collisions. This corresponds to PU throughput $r_P = \left(\frac{K-1}{K}\right) u$. This strategy converges to the optimal throughput pair $(r_P = u, r_S = 1 - u)$ as $K \to \infty$. For $r_S < 1 - u$, this same strategy is used at the SU while also remaining silent over more slots based on the tracking outcome. This provides the PU with more collision-free slots leading to a higher value of r_P and justifies our earlier claim that the optimal policy guarantees at least $K - 1$ collision free PU transmissions during each PU busy period.

Figure 6.9 shows how to choose the weight w defined in the reward function of the POMDP formulation to provide some level of protection to the PU. For a given target PU throughput, the corresponding weight w can be found from Fig. 6.9 and this weight is used in computing the optimal access policy at the SU.

6.4.3 Tracking Phase - Single PU - Greedy Policy

Although it is possible to find the optimal policy for the POMDP as in Section 6.4.2, this requires high complexity and is not generally feasible for real time processing or for energy limited nodes. In this part, we discuss the suboptimal greedy policy and show that its performance is close to that achieved by the POMDP with much less complexity.

The greedy policy aims at finding the action that maximizes the expected immediate

reward in a slot without taking into account the effect of that action on the future reward. Hence, from Eq. (6.18), the greedy policy can be found to be:

$$A_t^{\text{Greedy}} = \begin{cases} 0, & \text{if } \mathbf{\Lambda}_t(0) \leq \frac{wr_{P,1}}{wr_{P,1}+(1-w)r_{S,1}}, \\ 1, & \text{otherwise.} \end{cases} \tag{6.19}$$

Clearly, the greedy policy is a threshold policy on the probability that the PU is idle $\mathbf{\Lambda}_t(0)$, which is easy to implement. Figure 6.10 shows the performance of the greedy policy for various values of K.

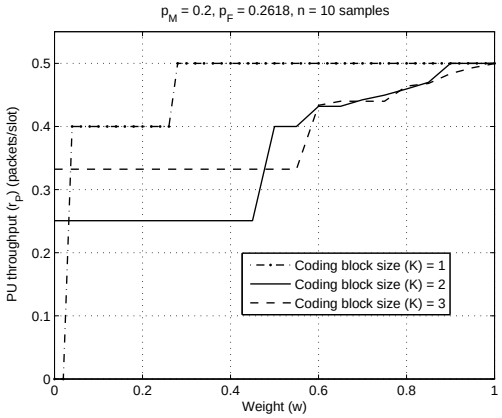

Figure 6.9: PU throughput as a function of the weight w for the POMDP optimal policy.

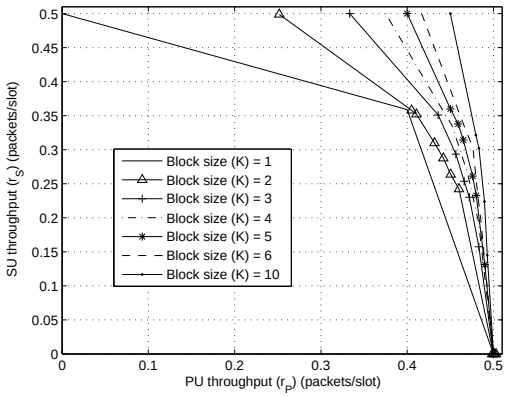

Figure 6.10: Throughput performance of the greedy policy.

6.4.4 Tracking Phase - Multiple PUs

In this section, we generalize the model to the case of multiple primary channels that evolve independently. At the beginning of every slot, the SU chooses a channel to sense and possibly access depending on the sensing outcome. Unlike the case of a single PU where the PU was protected by choosing the weight w to guarantee some PU throughput, in this case, the PUs are protected by guaranteeing a maximum

collision probability $p_C^{(i)}$ to the ith PU, $i \in \mathcal{M} = \{1, 2, ..., M\}$. As in [34], this can be achieved by choosing a spectrum sensor operating point at the SU with misdetection probability $p_M = \min_i\{p_C^{(i)}\}$. The SU then transmits with rate $r_{S,1}$ only if the channel is sensed to be idle. In general, this choice of sensor operating point is a conservative choice since the ith PU will experience an effective collision probability less than $p_C^{(i)}$, and hence achieves a throughput that is larger than $u_i(1 - p_C^{(i)})$, where u_i is the channel utilization of the ith PU. At the end of each slot, the belief is updated given the sensing outcome and possible SU receiver feedback (ACK/NACK). At the beginning of the following slot, the SU chooses a channel to sense, then the Markov chains of the PUs make a transition, and then the sensing outcome of the chosen channel becomes available to the SU. If the channel is sensed as idle, the SU transmits while it remains silent if it is sensed to be busy. If the SU transmits, it receives a feedback (ACK/NACK) from its receiver. The belief is then updated based on the observations and the cycle repeats. Clearly, three possible observations are possible at the end of a slot based on which the belief is updated: (Busy), (Idle, ACK) and (Idle, NACK). We assume coordination between the SU-TX and SU-RX where in each slot the SU-TX informs the SU-RX of the channel it chooses to sense through a dedicated control channel. For ease of exposition, we subsequently assume symmetric system where all the PU Markov chains are identical (evolving independently as in Fig. 6.2 with same λ) and the collision constraint is the same for all i (equal to p_C). The asymmetric case can be similarly handled.

Let X_t^i be the true state of the ith PU channel in slot t, and $\mathcal{M} = \{1, 2, ..., M\}$ be the set of PUs in the system. The action taken by the SU at the beginning of slot

t is $A_t \in \mathcal{M}$, where $A_t = i$ means that the SU chooses to sense the ith channel for possible opportunistic access. We proceed to define the parameters of the POMDP formulation.

6.4.4.1 Belief Vector

Let $\mathbf{\Lambda}_t^{(i)}$, with $i \in \mathcal{M}$, be the belief vector of the state of the ith PU's channel at the end of slot t, where $\mathbf{\Lambda}_t^{(i)}(m)$ is the probability that the ith PU's Markov chain is in state m at the end of slot t given all the previous actions and observations, where $m \in \{0, 1, ..., K\}$. Also define $\mathbf{\Lambda}_t = \left[\mathbf{\Lambda}_t^{(1)}; \mathbf{\Lambda}_t^{(2)}; ...; \mathbf{\Lambda}_t^{(M)}\right]$.

6.4.4.2 Observations

Under action $A_t = j$, the three possible observations on the jth channel are (Busy), (Idle, ACK) and (Idle, NACK).

$$\Pr\left[\text{Busy}|\mathbf{\Lambda}_t\right] = \left[\mathbf{\Lambda}_t^{(j)}(0) + \mathbf{\Lambda}_t^{(j)}(K)\right]\left[\lambda(1-p_M) + (1-\lambda)p_F\right] + (1-p_M)\left[\sum_{m=1}^{K-1} \mathbf{\Lambda}_t^{(j)}(m)\right], \tag{6.20}$$

$$\Pr\left[(\text{Idle}, \text{ACK})|\mathbf{\Lambda}_t\right] = \left[\mathbf{\Lambda}_t^{(j)}(0) + \mathbf{\Lambda}_t^{(j)}(K)\right](1-\lambda)(1-p_F), \tag{6.21}$$

$$\Pr\left[(\text{Idle}, \text{NACK})|\mathbf{\Lambda}_t\right] = p_M\left[\sum_{m=1}^{K-1} \mathbf{\Lambda}_t^{(j)}(m) + \lambda\mathbf{\Lambda}_t^{(j)}(0) + \lambda\mathbf{\Lambda}_t^{(j)}(K)\right]. \tag{6.22}$$

6.4.4.3 Belief Update

Under action $A_t = j$, the belief is updated as follows:

For $r \neq j$:

$$\mathbf{\Lambda}_{t+1}^{(r)}(0) = (1-\lambda)\left[\mathbf{\Lambda}_t^{(r)}(0) + \mathbf{\Lambda}_t^{(r)}(K)\right], \tag{6.23}$$

$$\mathbf{\Lambda}_{t+1}^{(r)}(1) = \lambda\left[\mathbf{\Lambda}_t^{(r)}(0) + \mathbf{\Lambda}_t^{(r)}(K)\right], \tag{6.24}$$

$$\mathbf{\Lambda}_{t+1}^{(r)}(m) = \mathbf{\Lambda}_t^{(r)}(m-1), m = 2, 3, ..., K. \tag{6.25}$$

For $r = j$:

• Under observation $O(A_t) = \text{Busy}$, the belief is updated according to

$$\mathbf{\Lambda}_{t+1}^{(r)}(m) = \frac{[p_F \mathbb{1}[m=0] + (1-p_M)\mathbb{1}[m \neq 0]]\,\Gamma_m}{\sum_{m=0}^{K}[p_F \mathbb{1}[m=0] + (1-p_M)\mathbb{1}[m \neq 0]]\,\Gamma_m}, \tag{6.26}$$

where Γ_m is as given by Eq. (6.9).

• Under observation $O(A_t) = (\text{Idle}, \text{ACK})$, the belief is updated according to

$$\mathbf{\Lambda}_{t+1}^{(r)}(m) = \mathbb{1}[m=0]. \tag{6.27}$$

• Under observation $O(A_t) = (\text{Idle}, \text{NACK})$, the belief is updated according to

$$\mathbf{\Lambda}_{t+1}^{(r)}(m) = \frac{\mathbb{1}[m \neq 0]\Gamma_m}{\sum_{m=1}^{K}\mathbb{1}[m \neq 0]\Gamma_m}. \tag{6.28}$$

In this case, the belief vector is of dimension $M(K+1) \times 1$. Obtaining the optimal POMDP solution is computationally prohibitive. We proceed by comparing the optimal solution for $K = 1$ with the greedy solutions for $K > 1$ and show that using the suboptimal greedy policy ensures gains with larger K compared with the optimal solution for $K = 1$.

6.4.4.4 Reward

Assuming a collision channel, the SU collects a reward of $r_{S,1}$ only if the chosen channel is idle and sensed to be idle (no false alarm). The reward under action $A_t = j \in \mathcal{M}$ is given by

$$R\left(X_t^{(1)}, X_t^{(2)}, ..., X_t^{(M)}, A_t = j\right) = \begin{cases} r_{S,1}, & \text{if } X_t^{(j)} = 0, \text{ no false alarm} \\ 0, & \text{otherwise.} \end{cases}, \quad (6.29)$$

The expected immediate reward under action $A_t = j$ is given by

$$R_j\left(\mathbf{\Lambda}_t\right) = r_{S,1}(1-\lambda)(1-p_F)\left(\mathbf{\Lambda}_t^{(j)}(0) + \mathbf{\Lambda}_t^{(j)}(K)\right) \quad (6.30)$$

6.4.4.5 Greedy Policy

Choose channel s to sense, where

$$s = \arg\max_{q \in \mathcal{M}} \left[\mathbf{\Lambda}_t^{(q)}(0) + \mathbf{\Lambda}_t^{(q)}(K)\right](1-\lambda)(1-p_F)r_{S,1} \quad (6.31)$$

In the numerical results, we consider the case with $M = 3$. All the other parameter values are as before, and the channel utilization is fixed to $u = 0.5$ for each of the three PUs. By the symmetry of the system, it is clear that the PUs achieve equal throughputs. From Fig. 6.11 and Fig. 6.12, we can see that the greedy policy achieves throughput gain for $K > 1$ for both the SU and the PU compared with the optimal solution for $K = 1$. This is due to the more predictability of the PU channels for higher K.

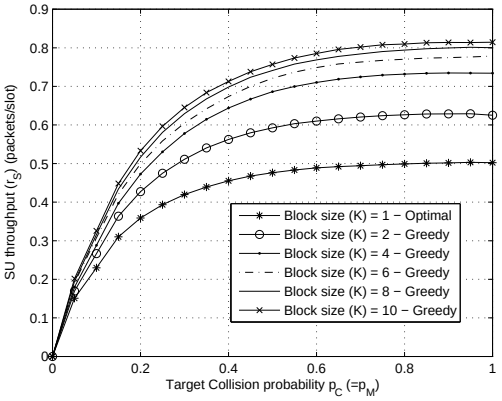

Figure 6.11: SU throughput vs. target collision probability, $M = 3$ PUs.

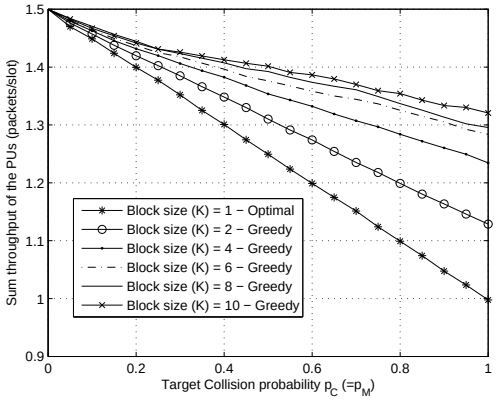

Figure 6.12: PU throughput vs. target collision probability, $M = 3$ PUs.

6.5 Sensing network-coded spectrum with real radio measurements

Next, we validate our spectrum sensing results with actual radio measurements. In the hardware experiments, wireless tests are executed in the 2.462 GHz channel with 10dBm transmission power and 256 Kbps rate. We use two configurable RF front-ends (RouterStation Pro from Ubiquiti) as the PU transmitter and receiver. The PU transmitter sends packets over the air to the PU receiver according to the Markov chain in Fig. 6.2. We use USRP N210 as the software-defined-radio platform that acts as the SU that detects the idle slots of the PU in a synchronous slotted system. Although the Gaussian power profiles used through the chapter is commonly used since it occurs in narrowband channels with Rayleigh fading, we aim in this part at testing our proposed algorithms on real spectrum power profile obtained via measurements. In fact, it has been observed that in the indoor experiment setup, and for the devices used, the signal profile is not Gaussian. However,

it has been shown that even for the real measurements, a larger block size K is beneficial for both the PU and the SU.

For the throughput maximization problem, the SU receives physical RF signals and runs an energy detector implemented on the GNU Radio, and passes the channel sensing results to the tracking algorithm. In these testbed experiments, the distribution of the measured power at the SU transmitter does not match Gaussian characteristics and hence the energy threshold values τ (Eq. (6.2)) previously used in simulations, based on Gaussian assumptions, cannot be directly used. In Fig. 6.13, we plot the empirical misdetection probability p_M obtained by applying energy detector with $n = 10$ sensing samples to the measured data as a function of the corresponding threshold. For a given target misdetection probability p_M, the SU can choose the energy threshold accordingly.

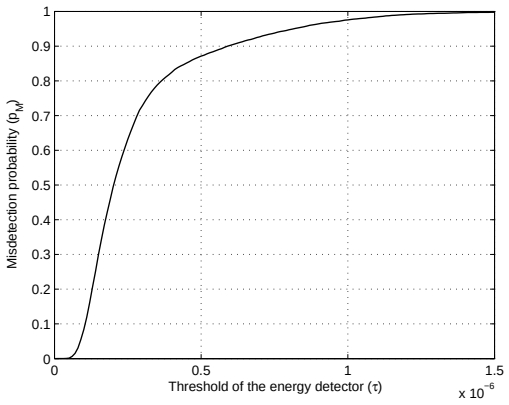

Figure 6.13: Misdetection probability p_M vs. the energy threshold for the experiment data.

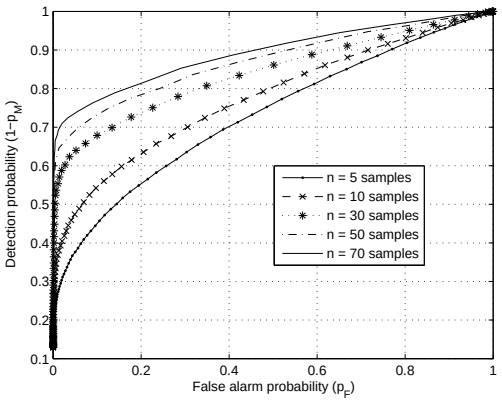

Figure 6.14: ROC curves for the real radio experiment measurements.

Figure 6.14 shows the relationship between the detection probability and the false alarm probability for different numbers of sensing samples n. In testbed exper-

iments, the noise variance was measured to be $\sigma^2 = -68.7887$ dBm and the average received power is -68.2286 dBm leading to an SNR= 0.561 dB. From Fig. 6.13, we can see that for $n = 10$ sensing samples per slot and target misdetection probability $p_M = 0.2$, the corresponding energy threshold τ is 1.285×10^{-7}. The corresponding false alarm probability is $p_F = 0.4905$, which is obtained from Fig. 6.14. We only observe the spectrum predictability gain since we fix the channel utilization to $u = 0.5$.

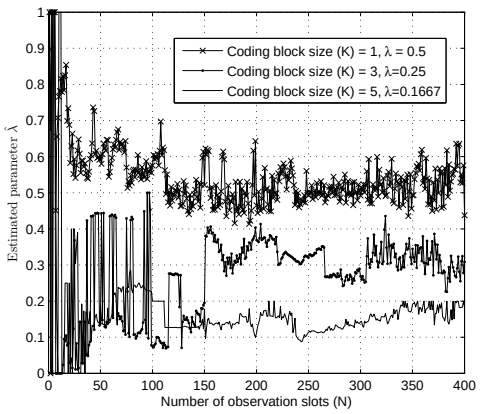

Figure 6.15: Estimated parameter $\hat{\lambda}$ vs. N under real measurements.

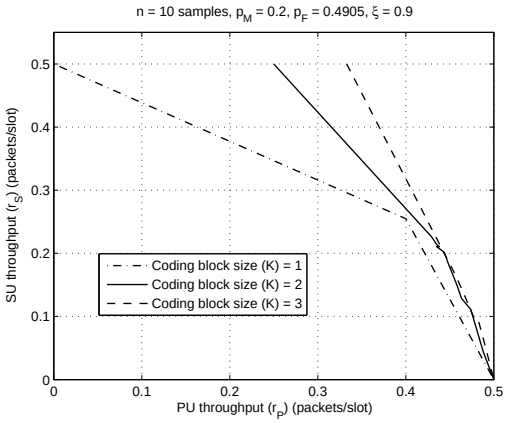

Figure 6.16: Throughput of the POMDP optimal policy under real measurements.

Figure 6.15 shows the performance of the Baum-Welch algorithm when applied to the measured data. We see that although a higher K leads to a better estimate for the same number of observed slots, none of the estimated values exactly matches the true value of λ after 400 slots. However, for practical purposes, 150 observation slots lead to a fairly accurate estimate. Figure 6.16 shows the performance of the POMDP optimal policy with measured data. It is clear that as K increases, for the same r_P, the SU can achieve higher throughput r_S due to the spectrum predictability

gain and this supports the use of network coding for better throughput performance of both PU and SU.

6.6 Summary and Conclusions

We showed how the SU can leverage the structure of the idle/busy periods on the PU spectrum with network-coded transmissions. We considered perfect PU channels without any spectrum availability gain due to network coding, but even in this case the SU can largely benefit from improved spectrum predictability due to the structure induced by network coding on the PUs' channels. We applied the CUSUM and Viterbi algorithms for the SU to minimize the time to detect an idle slot. Although a larger coding block size leads to better predictability of the PU channel state, the long busy periods involved in block transmission may limit the performance gain depending on the desired level of PU protection. When the SU's objective is to maximize the throughput, the problem is formulated as a POMDP for tracking the PUs' spectrum state evolution. We first considered the optimal POMDP policy in the case of a single PU channel and then considered the suboptimal low-complexity greedy policy. It was shown that increasing the coding block size is always beneficial for the SU throughput. Similar conclusions were shown for the case of multiple PUs where we showed that the greedy policy with block sizes $K \geq 2$ outperforms the optimal policy with $K = 1$, i.e., with no network-coding. Finally, we verified these results with real radio measurements obtained by actual radio transmissions over the air.

Chapter 7: Conclusions

In this dissertation, we studied cognitive wireless networks from a cross-layer (PHY/MAC) perspective. We showed that by accounting for the interactions between the layers, significant performance enhancements can be achieved. We studied both cases of unicasting and multicasting networks. Modern transmission techniques such as cooperative communications and network coding have been investigated in cognitive environments.

We first considered a random access system where the nodes adjust their MAC layer access probabilities according to their decentralized channel state information (CSI) obtained at the PHY layer (transmission control). From a stable throughput point of view, for a collision channel, it was shown that if the channels tend to be in the bad states, then random access with transmission control outperforms orthogonal access since the decentralized CSI allows the users to avoid transmission when their channels are bad and hence reducing contention between the users. By further enhancement of the PHY layer by using sophisticated receivers with MPR capability, this restriction is alleviated and random access with transmission control always outperforms orthogonal access. In the latter case, the optimal transmission policy at the users is to transmit whenever backlogged. We then studied the effect

of such cross-layer approach on the average delay performance for the case of statistically symmetric users and channels over a collision channel. It was shown that for the same average arrival rates to the users, transmission control leads to a smaller average delay compared to the case with no transmission control. Hence, although acquiring channel state information for time varying channels requires more complexity, it renders the distributed random access protocol better than orthogonal access.

Next, we studied a cognitive wireless network where a set of secondary users (SUs) opportunistically accesses the spectrum licensed to a primary user (PU). We showed that unavoidable sensing errors at the PHY layer can lead to detrimental effects on the stable throughput of the PU due to the interference involved. In order to mitigate such negative effects, we proposed a cross-layer cooperative scheme between the SUs and the PU, where all the SUs that were able to decode a PU's unsuccessful packet collaboratively forward that packet in the next idle slots using distributed orthogonal space-time block code. It was shown that this protocol has the property that the gain in PU throughput by relaying increases as the number of SUs with whom the PU shares the spectrum increases. Furthermore, it was shown that the SUs can also benefit from relaying. The PU benefits by having spatial diversity to his transmissions, while the SUs benefit by helping the PU to empty his queue and hence having access to a higher fraction of idle slots. Hence, relaying can be an incentive for the PUs to share their resources with more SUs leading to an overall increase in the spectral efficiency, as targeted in cognitive networks.

Motivated by the observation that schemes based on sensing are conservative

since they waste the transmission opportunities when the primary transmitter is busy while the channel between the secondary transmitter and primary receiver (cross-channel) is in deep fade, we proposed and analyzed several access schemes that exploit the knowledge of the cross channel statistics for maximizing the SU stable throughput subject to some protection metric to the PU. In particular, we compared between a scheme with no sensing where the sensing duration is exploited for data transmission and a scheme where the secondary user senses the channel and randomizes his transmissions at all slots. We showed that if the PU and SU receivers cannot successfully decode transmissions in presence of interference, then the scheme with sensing is preferred; while if the receivers can handle the interference, then no sensing is better as it provides the SU with more data transmission duration and hence more throughput. This means that using sophisticated receivers that can handle transmissions in presence of interference alleviates the need for complex SU transmitters with high sensing performance. This is preferred for instance if the SU destination is an eNodeB capable of handling high complexity.

We then turned our attention to the case of multicasting networks where network coding comes naturally into the picture. We first explored the problem of relay assisted multicasting. The relay is a cognitive node that assists the source in multicasting its traffic only whenever the source is idle and hence, avoiding allocating any explicit resources to the relay. The relay is also capable of using network coding on the packets he has in queue. It was shown that if the channel from the relay to the destinations is better on average than the channel between the source and the destinations, then cognitive relaying is beneficial. Furthermore, network

coding at the relay can lead to further enhancements as the number of destinations gets larger. Hence, at the expense of complexity due to network coding and adding a relay node, significant throughput gains can be achieved. This can be applied, for instance, to downlink broadcast channels such as the Multicast-broadcast single-frequency network (MBSFN) in the 4G-LTE systems.

Finally, we studied how an SU can leverage the shaping effects of network coding when applied to a PU spectrum. We first studied these effects in two different objectives: (i) quickest detection of an idle slot and (ii) throughput maximization. For quickest detection, it was shown that network coding is not necessary beneficial from a quickest detection perspective since it might take the PU's spectrum a longtime to switch to the idle state. We studied both cases of unknown spectrum dynamics at the SU using the CUSUM algorithm and the known spectrum dynamics using Viterbi algorithm. For throughput maximization, it was shown that the structure induced on the PU spectrum facilitates learning the spectrum dynamics at the SU. Furthermore, the shaping effects lead to better tracking of the PU spectrum dynamics by exploiting the side information about the PU spectrum dynamics for mitigating the sensing errors at the PHY layer. Our analytical and simulation results are supported by real-time signal measurements in an SDR testbed. The conclusion of that part is that using network coding at the PU not only possibly leads to a higher PU throughput but also, if properly exploited at an SU, leads to higher SU throughput. This highlights the importance of using network coding in practical multicasting systems.

Bibliography

[1] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2416–2434, October 1998.

[2] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 112–119, December 2005.

[3] V. Kawadia and P. R. Kumar, "A cautionary perspective on cross layer design," *IEEE Wireless Communications Magazine*, vol. 12, no. 2, pp. 3–11, December 2005.

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

[5] FCC, "Et docket no. 03-108," [Online] Available: http://www.fcc.gov/oet/cognitiveradio/.

[6] D. Chen, S. Yin, Q. Zhang, and S. Li, "Mining spectrum usage data: a large-scale spectrum measurement study," in *Proc. of international conference on Mobile computing and networking. ACM Mobicom 2009, Beijing, China*, Sept. 2009.

[7] M. A. McHenry, "Spectrum white space measurements," in *New America Foundation Broadband Forum*, June 2003.

[8] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, Royal Institute of Technology, Sweden, May 2000.

[9] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks (Elsevier) Journal*, vol. 50, no. 4, pp. 2127–2159, Sept. 2006.

[10] Y. C. Liang, K. C. Chen, G. Y. Li, and P. Mähönsen, "Cognitive radio networking and communications: An overview," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3386–3407, Sept. 2011.

[11] N. Devroye, M. Vu, and V. Tarokh, "Cognitive radio networks," *IEEE Signal Processing Magazine*, pp. 12–23, Nov. 2008.

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

[13] M. Vu, S. S. Ghassemzadeh, and V. Tarokh, "Interference in a cognitive network with beacon," in *Proc. IEEE Wireless Communications and Networking Conference*, Mar. 2008, pp. 876–881.

[14] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 116–130, First Quarter 2009.

[15] N. Abramson, "The ALOHA system: Another alternative for computer communications," in *Proc. Fall Joint Computer Conf., AFIPS Conf. Proc., Montvale, NJ.*, 1970, p. 281285.

[16] F. Digham, M. S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Trans. Commun*, vol. 55, no. 1, pp. 21–24, Jan. 2007.

[17] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[18] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[19] G. Scutari and S. Barbarossa, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2387–2399, Sept. 2005.

[20] A. K. Sadek, K. J. R. Liu, and A. Ephremides, "Cognitive multiple access via cooperation: Protocol design and performance analysis," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3677–3696, Oct. 2007.

[21] B. Rong and A. Ephremides, "Cooperation above the physical layer: the case of a simple network," in *IEEE International Symposium on Information Theory (ISIT), Seoul, Korea*, June/July 2009.

[22] B. Rong and A. Ephremides, "Protocol-level cooperation in wireless networks: Stable throughput and delay analysis," in *the 7th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Seoul, Korea*, June 2009.

[23] A. Fanous and A. Sultan, "Optimal sensing duration based on primary feedback in energy limited cognitive networks," in *International Conference on Communications (ICC), Budapest, Hungary*, June 2013.

[24] F. E. Lapiccirella, S. Huang, X. Liu, and Z. Ding, "Feedback-based access and power control for distributed multiuser cognitive networks," in *Proc. Information Theory and Applications Workshop (ITA)*, 2009.

[25] S. Huang, X. Liu, and Z. Ding, "Distributed power control for cognitive user access based on primary link control feedback," in *Proc. International Conference on Computer Communications (INFOCOM)*, 2010.

[26] S. Huang, Z. Ding, and X. Liu, "Cognitive spectrum access control based on intrinsic primary ARQ information," in *Proc. International Conference on Communications (ICC)*, 2010.

[27] A. Sultan, "Leveraging primary feedback and spectrum sensing for cognitive access," in *Proc. IEEE Wireless Communications and Networking Conference*, Apr. 2012.

[28] T. V. Nguyen, H. Shin, T. Quek, and M. Win, "Sensing and probing cardinalities for active cognitive radios," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1833–1848, Apr. 2012.

[29] R. Ahlswede, N. Cai, S. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[30] T. Ho, R. Koetter, M. Medard, M. Effros, J. Shi, and D. Karger, "A random linear network coding approach to multicast," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[31] A. Eryilmaz, A. Ozdaglar, and M. Médard, "On the delay and throughput gains of coding in unreliable networks," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5511–5524, Dec. 2008.

[32] S. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, p. 371381, Feb. 2003.

[33] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.

[34] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive Mac for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[35] Z. Sun, G. J. Bradford, and J. N. Laneman, "Sequence detection algorithms for PHY-layer sensing in dynamic spectrum access networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 97–109, Feb. 2011.

[36] S. Wang and J. Zhang, "Opportunistic spectrum scheduling by jointly exploiting channel correlation and PU traffic memory," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 394–405, Mar. 2013.

[37] D. MacKay, "Fountain codes," *IEE Proc.-Commun*, vol. 152, no. 6, pp. 1062–1068, 2005.

[38] A. Fanous, Y. Sagduyu, and A. Ephremides, "Reliable spectrum sensing and opportunistic access in network-coded communications," *IEEE J. Sel. Areas Commun.*, Mar. 2014, To appear.

[39] B. Tsybakov and W. Mikhailov, "Ergodicity of slotted ALOHA systems," *Probl. Inform. Transm.*, vol. 15, no. 4, p. 7387, 1979.

[40] R. Rao and A. Ephremides, "On the stability of interacting queues in a multiaccess system," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 918–930, Sept. 1988.

[41] W. Luo and A. Ephremides, "Stability of N interacting queues in random-access systems," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1579–1587, July 1999.

[42] S. Ghez, S. Verdú, and S. Schwartz, "Stability properties of slotted ALOHA with multipacket reception capability," *"IEEE Trans. Autom. Control"*, vol. 33, no. 7, pp. 640–649, July 1988.

[43] V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite user slotted ALOHA with multipacket reception," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, July 2005.

[44] S. Adireddy and L. Tong, "Exploiting decentralized channel state information for random access," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 537–561, Feb. 2005.

[45] S. Wang, C. Lin, and Y. Hong, "On the stability and delay of channel aware slotted ALOHA with imperfect CSI," in *International Communications Conference (ICC) Proceedings, Beijing, China*, May 2008.

[46] Y. Hong, C. Lin, and S. Wang, "On the stability region of two-user slotted ALOHA with cooperative relays," in *International Symposium on Information Theory (ISIT), Nice, France*, June 2007.

[47] A. Fanous and A. Ephremides, "Transmission control of two-user slotted ALOHA over Gilbert-Elliott channel: Stability and delay analysis," in *International Symposium on Information Theory (ISIT), St. Petersburg, Russia*, July/August 2011.

[48] W. Szpankowski, "Stability conditions for some multiqueue distributed systems: Buffered random access systems," *"Adv. Appl. Probab."*, vol. 26, pp. 498–515, 1994.

[49] R. Loynes, "The stability of a queue with non-interdependent inter-arrival and service times," *Proc. Camb. Philos. Soc.*, vol. 58, pp. 497–520, 1962.

[50] B. Rong and A. Ephremides, "Stable throughput, rate control, and delay in multi-access channels," in *the 8th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Avignon, France*, May/June 2010.

[51] M. Sidi and A. Segall, "Two interfering queues in packet-radio networks," *IEEE Trans. Commun*, vol. 31, no. 1, pp. 123–129, Jan. 1983.

[52] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[53] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 3945–3958, Sept. 2009.

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

[55] A. Fanous and A. Ephremides, "Effect of secondary nodes on the primary's stable throughput in a cognitive wireless network," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2012.

[56] S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Stable throughput tradeoffs in cognitive shared channels with cooperative relaying," in *International Conference on Computer Communications (Infocom)*, Apr. 2011.

[57] I. Krikidis, N. Devroye, and J. S. Thompson, "Stability analysis for cognitive radio with multi-access primary transmission," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 72–77, Jan. 2010.

[58] A. A. El-Sherif, A. K. Sadek, and K. J. R. Liu, "Opportunistic Multiple Access for Cognitive Radio Networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 704–715, Apr. 2011.

[59] I. Krikidis, J. N. Laneman, J. S. Thompson, and S. McLaughlin, "Protocol design and throughput analysis for multi-user cognitive cooperative systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4740–4751, Sept. 2009.

[60] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Spacetime codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[61] A. Fanous and A. Ephremides, "Stable throughput in a cognitive wireless network," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 523–533, Mar. 2013.

[62] A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?" *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2662–2671, May 2009.

[63] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Englewood Clifffs, NJ: Prentice-Hall, 1987.

[64] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer Texts in Electrical Engineering, 2010.

[65] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[66] H. Wang and X.-G. Xia, "Upper bounds of rates of complex orthogonal space-time block codes," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2788–2796, Oct. 2003.

[67] J. Unnikrishnan and V. V. Veeravalli, "Cooperative Sensing for Primary Detection in Cognitive Radio," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 18–27, Feb. 2008.

[68] S. M. Ross, *Introduction to probability Models*, 6th ed. Academic Press, 1997.

[69] A. Fanous and A. Ephremides, "Access schemes for mitigating the effects of sensing errors in cognitive wireless networks," *IEEE Trans. Wireless Commun.*, To appear.

[70] A. A. El-Sherif and K. J. R. Liu, "Joint design of spectrum sensing and channel access in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1743 – 1753, June 2011.

[71] A. M. Arafa, K. G. Seddik, A. K. Sultan, T. ElBatt, and A. A. El-Sherif, "A soft sensing-based cognitive access scheme exploiting primary feedback," in *Proc. of the 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, May 2010.

[72] A. E. Shafie, "To sense or not to sense ?" June 2012, available online: http://arxiv.org/abs/1206.6153.

[73] P. Nain, "Analysis of a two-node ALOHA-network with infinite capacity buffers," in *Proc. Int. Seminar on Computer Networking and Performance Evaluation, Tokyo, Japan*, September 1985, pp. 49–63.

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

[75] S. Boyd and L.Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[76] H. Takagi, *Queueing Analysis*, 6th ed. North-Holland, 1993, vol. 3.

[77] R. Tandra and A. Sahai, "Is interference like noise when you know its codebook?" in *International Symposium on Information Theory (ISIT), Seattle, USA*, July 2006.

[78] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part i: System description," *IEEE Trans. Commun*, vol. 51, pp. 1927–1938, Nov. 2003.

[79] P. Fan, C. Zhi, C. Wei, and K. B. Letaief, "Reliable relay assisted wireless multicast using network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, pp. 749–762, June 2009.

[80] A. Fanous and A. Ephremides, "Network-level cooperative protocols for wireless multicasting: Stable throughput analysis and use of network coding," in *Information Theory Workshop (ITW), Dublin, Ireland*, September 2010.

[81] Y. E. Sagduyu and A. Ephremides, "On network coding for stable multicast communication," in *IEEE Military Communications Conference MILCOM*, Oct. 2007.

[82] N. Baldo, A. Asterjadhi, and M. Zorzi, "Dynamic spectrum access using a network coded cognitive control channel," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2575–2587, Aug. 2010.

[83] S. Wang, Y. E. Sagduyu, J. Zhang, and J. H. Li, "The impact of induced spectrum predictability via wireless network coding," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 758–769, Feb. 2012.

[84] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[85] H. V. Poor and O. Hadjiliadis, *Quickest Detection*, 1st ed. Cambridge University Press, 2009.

[86] G. D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[87] C. Ghosh, C. Cordeiro, D. P. Agrawal, and M. B. Rao, "Markov chain existence and Hidden Markov models in spectrum sensing," in *IEEE International Conference on Pervasive Computing and Communications (PerCom), 2009*, Mar. 2009.

[88] Q. Zhao and J. Ye, "Quickest detection in multiple on-off processes," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 5994–6006, Dec. 2010.

[89] A. Schliep, "Learning hidden Markov model topology," in *German Conference on Bioinformatics*, 1999.

[90] H. Li, C. Li, and H. Dai, "Quickest spectrum sensing in cognitive radio," in *42nd Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2008.

[91] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 2008.

[92] B. Chen and P. Willett, "Detection of hidden Markov model transient signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1253–1268, 2000.

[93] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[94] A. Fanous, Y. E. Sagduyu, and A. Ephremides, "Opportunistic spectrum access in wireless communications with network coding," in *the 11th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Tsukuba Science City, Japan*, May 2013.

[95] T. Nguyen, B. L. Mark, and Y. Ephraim, "Hidden Markov process based dynamic spectrum access for cognitive radio," in *45th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2011.

[96] O. Madani, S. Hanks, and A. Condon, "On the undecidability of probabilistic planning and infinite horizon partially observable decision problems," in *Proc. National Conference on Artificial Intelligence*, 1999.

[97] M. L. Putterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability & Statistics, 2005.

[98] O. Habachi, S. H, M. van der Shaar, and Y. Hayel, "Online learning based congestion control for adaptive multimedia transmission," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1460–1469, Mar. 2013.

[99] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscribers lines," *IEEE Trans. Commun*, vol. 54, no. 5, pp. 922–933, May 2006.

[100] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations Research*, vol. 21, no. 5, pp. 1071–1088, 1973.

[101] C. H. Papadimitrou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Math. Oper. Res.*, vol. 12, no. 3, pp. 441–450, 1987.