ABSTRACT

Title of Dissertation: COGNITIVE MULTIPLE ACCESS FOR COOPERATIVE COMMUNICATIONS AND NETWORKING

Amr A. El-Sherif, Doctor of Philosophy, 2009

Dissertation directed by: Professor K. J. Ray Liu
Department of Electrical and Computer Engineering

In cooperative communications different network nodes share their antennas and resources to form a virtual antenna array and improve their performance through spatial diversity. This thesis contributes to the advancement of cooperative communications by developing and analyzing new multiple access cooperation protocols that leverage the benefits of cooperation to upper network layers.

For speech communications networks, we propose a cooperative multiple access protocol that exploits inherent characteristics of speech signals, namely, long periods of silence, to enable cooperation without incurring bandwidth efficiency losses. Using analytical and simulation results we show that the proposed protocol achieves significant increase in network throughput, reduction in delay, and improved perceptual speech quality.

In TDMA networks, we investigate the problem of sharing idle time slots between a group of cooperative cognitive relays helping primary users, and a group of cognitive secondary users. Analytical results reveal that, despite the apparent competition between relays and secondary users, and even in case of mutual interference between the two groups, both primary and secondary users will significantly benefit in terms of maximum stable throughput from the presence of relays.

For random access networks, we find a solution to the problem of achieving cooperation gains without suffering from increased collision probability due to relay transmissions. A novel cooperation protocol is developed and analyzed for that purpose. Analytical and simulation results reveal significant improvements in terms of throughput and delay performance of the network. Moreover, collision probability is decreased.

Finally, in the framework of a cognitive radio network, we study the negative effects of spectrum sensing errors on the performance of both primary and secondary networks. To alleviate those negative effects, we propose a novel joint design of the spectrum sensing and channel access mechanisms. Results show significant performance improvement in the maximum stable throughput region of both networks.

COGNITIVE MULTIPLE ACCESS FOR COOPERATIVE COMMUNICATIONS AND NETWORKING

by

Amr A. El-Sherif

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
2009

Advisory Committee:

Professor K. J. Ray Liu, Chairman
Professor Richard La
Professor Mark Shayman
Professor Dianne O'Leary
Professor Amr Baz

PREFACE

All praise is due to Allah (God)

“And of knowledge, you (mankind) have been given only a little”, translation of the meaning of Verse 85, Chapter 17, the Holy Quran.

DEDICATION

To my Family

ACKNOWLEDGEMENTS

This thesis would have never been completed without the aid and support of countless people over the past four years. First of all, I would like to express my deep gratitude to my advisor, Prof. Ray Liu, for his continuous guidance and support during the years of my Ph.D study and research, for his leadership, patience, motivation, understanding in various aspects, and immense knowledge. I am very honored for conducting my Ph.D. research work under the supervision of Professor Liu.

My gratitude to Professor Richard La, Professor Mark Shayman, Professor Dianne O'Leary, and Professor Amr Baz for agreeing to be in my dissertation committee, and for devoting the time to revising the thesis.

I would like to thank every member in our Signals and Information group for the wonderful time we spent together and for the very nice discussions we had during our weekly group meetings. I have learned so much from them and will always keep the good memories, which we have shared together. I want to specially thank Dr. Ahmed Sadek, Prof. Karim Seddik, Dr. Ahmed Ibrahim, Prof. Andres Kwasinski, and Mohamed Baidas for the research discussions that we had that were all inspiring for me.

I am deeply indebted to my parents and my family for all their prayers and support in the past years. Finally, all praise is due to Allah for making me be able to achieve this important step.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

The proliferation of wireless applications with high demands in terms of signal quality and high data rates, e.g., multimedia service through cellular networks, has increased the attention toward the study of wireless channels. One major challenge is the scarcity of the two fundamental resources for communications, namely, energy and bandwidth. Another major challenges for communicating over wireless channels is the fading nature of those channels. Fading results in random fluctuations in the amplitude of the received signals that can result in the received signal amplitude being very low to the extent that the receiver may not be able to distinguish the signal from thermal noise [1], [2]. Fading is the result of the random scattering from reflectors with different attenuation coefficients that results in multiple copies of the signal arriving at the receiver with different gains, phase shifts and delays. These multiple signal replicas can add together in constructive or destructive ways resulting in the fading phenomenon [1], [2]. Therefore, there is an urgent need for wireless communication protocols that can mitigate the fading effect and improve the system performance.

1.1 Diversity

One solution to the fading nature of the wireless channels is the use of diversity achieving schemes. Diversity can be defined as any technique by which multiple copies of the signal are delivered to the receiver via independently fading channels [2]. The probability of having all the channels in deep fade is much lower than that of any individual channel. Independent channels could be generated in any of the three physical domains: time, frequency, and space. In time, the same signal can be transmitted at different well separated time slots to ensure uncorrelated channel realizations. Diversity in the frequency domain could be achieved by the transmission of the same signal over different frequency bands. Despite the gains in signal quality, time diversity is achieved at the expense of increased system delay, and frequency diversity is always associated with high bandwidth losses. Spatial diversity on the other hand can be achieved through the use of multiple transmit and/or receive antennas. The use of spatial diversity has gained a lot of interest in the recent years since it does not incur a penalty on the system in terms of delay or bandwidth. The diversity order of any scheme is defined as the rate of decay of the probability of error (P_e) with the Signal-to-noise ration (SNR) when using log-log scale [2], i.e.,

$$D = \lim_{SNR \to \infty} -\frac{\log P_e}{\log SNR}. \tag{1.1}$$

Spatial diversity could be achieved through the use of multiple-input-multiple-output (MIMO) systems. A MIMO system is simply one where both the transmitter and the receiver have multiple antennas. This implies that the transmitter has the capability of transmitting a different signal from each antenna and the receiver has as input different signals from each antenna. Therefore, MIMO channels have the ability of adding more degrees of freedom to the conventional single antenna channels, which result in higher channel capacity as was shown in [3] and [4]. MIMO systems can provide performance improvement through diversity gain. For example, if the number of antennas at the transmitter and receiver are M and N, respectively, and assuming independent fading between all antenna element pairs, the probability of error at the receiver side can be shown to decay with the SNR as SNR^{-MN}. Code design to achieve diversity in flat fading MIMO systems, also known as space-time codes, has been the focus of many researchers in the last decade [5–7].

While it is feasible to equip base stations with multiple antennas, small mobile units cannot have more than one antenna due to space constraints. This gave rise to a revolutionary concept, namely, cooperative diversity [8–13]. Cooperative communications benefit from the broadcast nature of the wireless channel to form a distributed MIMO system via relaying.

1.2 Cooperative Diversity

The classical relay channel model based on additive white Gaussian noise (AWGN) channels was presented in [14]. An upper bound on the channel capacity and an achievable lower bound for the additive white Gaussian noise (AWGN) relay channels were provided. Generally, the lower and upper bounds do not coincide except for special cases as in the degraded relay channels. In [15], different coding strategies haven been proposed, which achieve the ergodic capacity with phase fading if the phase information is known locally and if the relays are near the source.

1.2.1 PHY Layer Cooperation

User-cooperation has been first introduced and studied in [9, 10]. A two-user code division multiple access (CDMA) cooperative system, where both users are active and use orthogonal codes, was implemented in this two-part series. Assuming the knowledge of channel phases at the transmitter sides, increased data rates for the cooperating users have been demonstrated.

In [12], the term cooperative diversity was introduced. Several cooperation protocols were presented and their outage capacity was analyzed. Outage capacity can be defined as the probability that the mutual information of a channel falls below a certain required rate. Among the presented protocols are the decode-and-forward, amplify-and-forward, selection relaying, and incremental relaying.

In amplify-and-forward, the relay simply scales the received version and transmits an amplified version of it to the destination. Note that the amplified version is noisy because of the noise added at the relay. Despite of the noise propagation, it was shown in [12] that amplify-and-forward can achieve full diversity gain equal to two, the number of cooperating nodes in this case. In the decode-and-forward, the relay decodes the source symbol before re-transmitting to the destination. In order to achieve a diversity of order two for the single-relay protocol, the relay should be able to decide whether or not it has decoded correctly. This can be achieved through the use of error detecting codes or the use of appropriate SNR threshold at the relay node [16]. If the relay always forwards the source signal the system will achieve a diversity of order limited by errors at the relay node(s) and this is known as error propagation [17]. Symbol error rate performance analysis for the single-node and multi-node decode-and-forward cooperation protocols were provided in [16, 18].

In selective relaying, the relay and the source are assumed to know the fade of the channel between them, and if the signal-to-noise ratio of the signal received at the relay exceeds a certain threshold, the relay performs decode-and-forward on the message. On the other hand, if the channel between the source and the relay falls below the threshold, the relay idles. Furthermore, assuming reciprocity in the channel, the source also knows that the relay idles, and the source transmits a copy of its signal to the destination instead. Selective relaying improves upon the performance of decode-and-forward, as the signal-to-noise ratio threshold at the relay can be designed to overcome the inherent problem in decode-and-forward that the relay is required to decode correctly. Selection relaying was shown in [12] to achieve diversity gain two.

In incremental relaying, it is assumed that there is a feedback channel from the destination to the relay. The destination feeds back an acknowledgement to the relay if it was able to receive the sources message correctly in the first transmission phase, and the relay does not need to transmit then. It was shown in [12], that this protocol has the best spectral efficiency among the proposed protocols because the relay does not need to transmit always, and hence, the second transmission phase becomes opportunistic depending on the channel fade of the direct channel between the source and the destination. Nevertheless, incremental relaying achieves diversity order two [12].

1.2.2 MAC Layer Cooperation

Few works have studied the impact of cooperation on the multiple-access layers. In [19], the authors developed a cognitive multiple access protocol that overcomes the problem of bandwidth efficiency loss. The protocol in [19] exploits source burstiness to enable cooperation during silence periods of different nodes in a TDMA network. In other words, a cooperative relay will detect and utilize empty time slots in the TDMA frame to retransmit failed packets. Therefore, no extra channel resources are allocated for cooperation and the system encounters no bandwidth losses. The authors analyzed the protocols performance from a maximum stable throughput point of view, and their results revealed significant performance gains over conventional cooperation strategies.

Cooperation in random access networks has been considered in [20–22]. In [21], the authors proposed a distributed version of network diversity multiple-access (NDMA) [23] protocol and they provided pairwise error probability analysis to demonstrate the diversity gain. In [20] and [22], the authors presented the notion of utilizing the spatial separation between users in the network to assign cooperating pairs (also groups) to each other. In [22], spread spectrum random access protocols were considered in which nearby inactive users are utilized to gain diversity advantage via cooperation assuming a symmetrical setup where all terminals are statistically identical. However, the previously cited works still focus on physical layer parameters as the diversity gains achieved and the outage probability. User cooperation in slotted ALOHA random access network was investigated in [24], where the gains of cooperation on the stability region of a network consisting of multiple cooperating pairs is characterized.

1.3 Dissertation Outline

From the discussion above, the majority of the work on cooperative communications has focused on the different physical layer aspects of cooperation; protocol design, analysis, possible tradeoffs, etc.

In this thesis, our main goal is to shed more light on the impact of the cooperative communications paradigm on upper network layers. The main questions we try to answer are; how can cooperation be enabled in upper network layers? How network resources be assigned for cooperation? Finally, what are the possible benefits and tradeoffs associated with cooperation in upper network layers?

1.3.1 Content-Aware Cooperative Multiple Access (Chapter 2)

In Chapter 2, we propose a cooperative multiple access protocol that exploits source traffic properties to enable cooperation without incurring any bandwidth efficiency loss. Specifically, this protocol exploits the silence periods typical of speech communications to enable relays to forward speech packets for active calls using part of the free time slots left available by users that are silent. We provide a complete characterization of the protocol's performance using Markov-chains based analysis. It is proved that through cooperation, a speech network can achieve higher throughput, lower delays, and better perceptual speech quality. Because the resources allocated to the relay were previously available for other users channel access, no new exclusive channel resources are needed for cooperation and the system encounters no bandwidth losses. On the other hand, the use of cooperation imposes a tradeoff between the amount of help offered to active calls and the ability of the network to admit new users. This tradeoff is identified and analyzed. It is shown that by judicious control of this tradeoff it is possible to achieve through cooperation significant performance improvements [25, 26].

1.3.2 Opportunistic Multiple Access for Cooperative and Cognitive Networks (Chapter 3)

In Chapter 3, we try to answer the questions of; how to share the under-utilized channel resources between cooperative relays and cognitive secondary nodes? How does the coexistence of primary relays and secondary nodes affect the performance of both primary and secondary networks? And, what is the fundamental tradeoff between them?

To answer these questions, we consider the uplink of a TDMA network as the primary network, and start by studying how multiple cognitive relays can exploit the empty time slots to offer help to the primary nodes. We address the problem of how multiple relays share the resources among themselves, as well as, how relays are assigned to primary nodes by proposing two different relay assignment schemes. Then the presence of secondary nodes and their interaction with the primary network is considered.

The described system is studied form a queuing theory point of view. And the stability regions of both primary and secondary networks are characterized. It is shown that because of possible collisions between relays and secondary users, the set of queues in the network are interacting. In other words, the service process of a given queue is dependent on the state of all other queues. To decouple the interaction between queues and analyze the network, a dominant system approach for the analysis of queues stability is necessary.

Results reveal that although relays occupy part of the empty time slots that would have been available to secondary nodes, it is always beneficial to both primary and secondary nodes that the maximum possible number of relays be employed. On one hand, relays help the primary network achieve higher stable throughput by offering different reliable paths for the packets to reach the destination. On the other hand, relays will help primary nodes empty their queues at a much faster rates, thus providing secondary nodes with more opportunities to transmit their own information. It is interesting to note that even when secondary nodes interfere with relays transmissions, there is a significant improvement in both primary and secondary throughput due to this fast rate of emptying the queues [27, 28].

1.3.3 Joint Design of Spectrum Sensing and Channel Access (Chapter 4)

In Chapter 4, we focus on the effects of spectrum sensing errors on the performance of cognitive radio networks. We mainly try to answer the question of how the spectrum sensing errors affects the performance of the cognitive radio network from a multiple access protocol design point of view, and, how the joint design of spectrum sensing and access mechanisms can mitigate the negative effects of sensing errors.

We start by studying the effects of channel sensing errors on the performance of the multiple access layer, and reach the conclusion that because of spectrum sensing errors, the system suffers from severe degradation in the stability region of both primary and secondary networks. To mitigate the negative effects of sensing errors, we then propose a novel joint design of the spectrum sensing and access mechanisms. The design is based on the observation that, in a binary hypothesis testing problem, the value of the test statistic could be used as a measure of how confident we are in the test outcome. The further the value of the test statistic is from the decision threshold, the more confident we are that the decision is correct.

Therefore, instead of using the hard decisions of the spectrum sensor to decide whether to access the channel or not, a secondary user can have different access probabilities for different values of the test statistic. Design of the joint spectrum sensing and channel access scheme is formulated as an optimization problem. This optimization problem is shown to be non-convex, and several approximations are applied to the objective function to convert it to a convex problem which is easily solved. Results reveal that there is a huge improvement in performance by virtue of the proposed technique, and that the approximations applied do not result in any significant performance penalties [29].

1.3.4 Random Access Cooperative Networks (Chapter 5)

Chapter 5 tries to answer the questions of how to enable cooperation in a random access network without the possible increase in the number of packet collisions? And, since cooperation introduces extra transmissions in the channel, what are the benefits and possible tradeoffs associated with cooperation in this case?

To answer these questions, we start by proposing a cooperative protocol in which a relay node is deployed to help different network nodes to forward their packets to the access point (AP). Presence of the relay will help improve the communication channel through the spatial diversity it creates. The main challenge is, how to minimize the collision probability in order for the relay presence not to degrade the network performance instead of improving it. Through a careful design of the cooperation protocol, this goal is achieved by letting the relay deviate from the channel access mechanism dictated by the 802.11 protocol and to access the channel immediately after each transmission attempt on the channel. The protocol design guarantees an uncontested access to the wireless medium to the relay. Analytical results reveal significant gains in terms of network throughput, delay, the number of supported nodes, due cooperation and the proposed protocol. Furthermore, it is shown that, by virtue of the protocol design, collision probability shows a decrease rather than the expected increase due to extra transmissions on the channel [30].

Chapter 2

Content-Aware Cooperative Multiple Access

As discussed in chapter 1, most of the work on cooperative communications has focused on physical layer aspects of cooperation. Few are the works that considered the questions of; what are the effects of cooperative communications on the performance of upper network layers? And how is it enabled and implemented at these layers? In this chapter, we try to answer these questions by proposing a multiple access protocol that makes use of the source traffic properties to enable cooperation. Because the design of the protocol is highly dependent on the characteristics of the source, we will consider a packet speech communication network. Nevertheless, the main underlying ideas can be extended to other types of sources.

Speech communication has a distinct characteristic that differentiates it from data communication. Speech sources are characterized by periods of silence in between talk spurts. The speech talk-silence patterns could be exploited in statistical multiplexing-like schemes where silent users, release their channel resources, which can then be utilized to admit more users to the network. This comes at the cost of requiring a more sophisticated multiple access protocol. One well-known protocol to address this problem is the Packet Reservation Multiple Access (PRMA) protocol [31], which can be viewed as a combination of TDMA and slotted ALOHA protocols. In PRMA, the channel is slotted, and users in talk spurts contend for the channel in empty time slots. If a user is successful, then the slot is reserved for that user. Users with reservations transmit their speech packets in their reserved slots. When the talk spurt ends, the user stops transmitting packets, and its time slot is free for other users to access.

If a user fails to transmit its packet due to channel errors, that user loses the reservation, and reserved slot becomes free for contention again. Although the effects of this operation is practically unnoticeable in channels where errors are very infrequent, in channels with frequent errors (such as the wireless channel) the loss of reservation due to an error has the potential to adversely affect the efficiency of the protocol. This is because channel errors not only force the discarding of the damaged packets, but also increase the network traffic and access delay, as users with lost packets have to repeat the contention process again.

In this chapter we propose a cooperative multiple access protocol that uses the properties of speech to increase the network capacity and the cooperation efficiency. As is the case with established multiple access protocols, the network capacity is increased by reserving network resources only to the users in a talk spurt. Those users finishing a silence period need to contend for channel access over a shared resource. At the same time, the use of cooperation increases the system performance by helping users in a talk spurt reduce the probability of dropping packets and having to contend again. Cooperation is achieved through the deployment of a relay node. This relay node exploits the silence periods typical

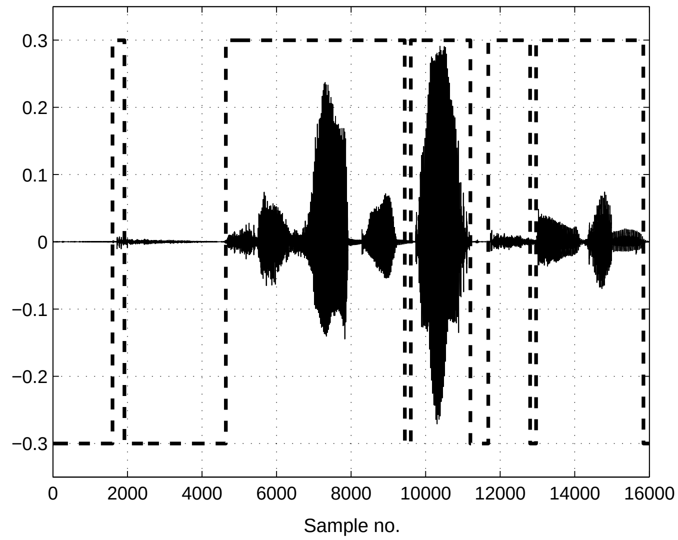

Figure 2.1: A typical speech segment illustrating the on/off characteristic of speech. The dashed lines take a value of 0.3 (chosen arbitrarily so the figure is sufficiently clear) when speech is detected "on" and a value of -0.3 when speech is detected "off".

of speech communications in a new way, it forwards speech packets for active calls using part of the free time slots left available by users that are silent. Most importantly, because the resources allocated to the relay were previously available for other users contending for channel access, no new exclusive channel resources are needed for cooperation and the system encounters no bandwidth losses. On the other hand, the use of cooperation imposes a tradeoff between the amount of help offered to active calls and the probability of a successful contention for channel access. This kind of tradeoff is, in some sense, similar to the diversity-multiplexing tradeoff in MIMO and other user cooperative systems. By judicious control of this tradeoff it is possible to achieve through cooperation significant performance improvements.

2.1 System Model

2.1.1 Speech Source Model

Speech sources are characterized by periods of silence in between talk spurts that account for roughly 60 % of the conversation time [32]. This key property could be exploited to significantly improve the utilization of channel resources but with the cost of requiring a more sophisticated multiple access protocol. Figure 2.1 shows the voice signal amplitude for a speech sequence and illustrates the alternation of speech between talking and silence periods. In the figure, the dashed line indicates the state of the speech sequence, "on" or "off", over the time. In order to make the figure sufficiently clear we have chosen arbitrarily a value of 0.3 when speech is detected "on" (talking state) and a value of -0.3 when speech is detected "off" (silence state). In practice, the detection of the speech state is performed in the source encoder through an algorithm named VAD (voice activity detector).

To model this alternation, each speech source in a conversation is modeled as a two state Markov chain, namely, talk (TLK) and silence (SIL) states (Fig. 2.2). In the figure, γ represents the transition probability from the talking state to the silence state and σ is the transition probability from the silence state to the talking state. The value of these two probabilities depend, of course, on the speech model but also on the time unit used to model state transitions in the Markov chain. In packet speech communications scenarios, as the one considered in this chapter, it is convenient for the purpose of mathematical analysis, to choose the same basic time unit for the Markov chain as the one used for channel access [33]. As will be discussed in the next section, the channel is divided into TDMA time frames, each of duration T seconds. Hence, it is suitable to choose the basic time unit for the

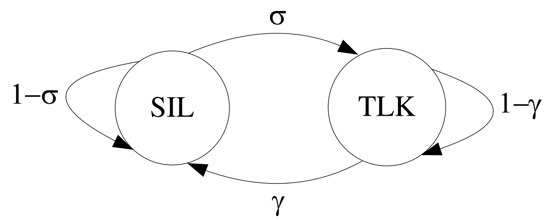

Figure 2.2: Speech source model using a Markov Chain. The two states TLK and SIL, correspond to a speaker being in a talk spurt or a period of silence, respectively.

Markov chain equal to T seconds also, which means that state transitions are only allowed at the frame boundaries.

It is also customary in the Markov chain modeling of speech sources to assume that the waiting time in any state has an exponential distribution [32]. Then, with these assumptions in mind, the state transition probabilities for the Markov chain can be calculated as follows: The transition probability from the talking state to the silence state is the probability that a talk spurt with mean duration t_1 ends in a frame of duration T, which can be calculated as,

$$\gamma = 1 - e^{-T/t_1}. \tag{2.1}$$

Similarly, the transition probability from the silence state to the talking state is the probability that a silence gap of mean duration t_2 ends during a frame of duration T, and is calculated as,

$$\sigma = 1 - e^{-T/t_2}. \tag{2.2}$$

2.1.2 Network Model

We consider a network with three types of nodes as illustrated in Fig. 4.2: source nodes, a relay node, and a base station. We will focus exclusively on the uplink

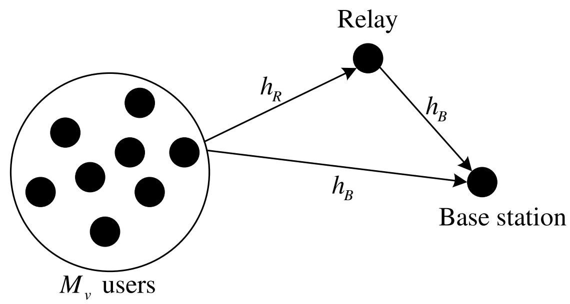

Figure 2.3: Network and Channel model.

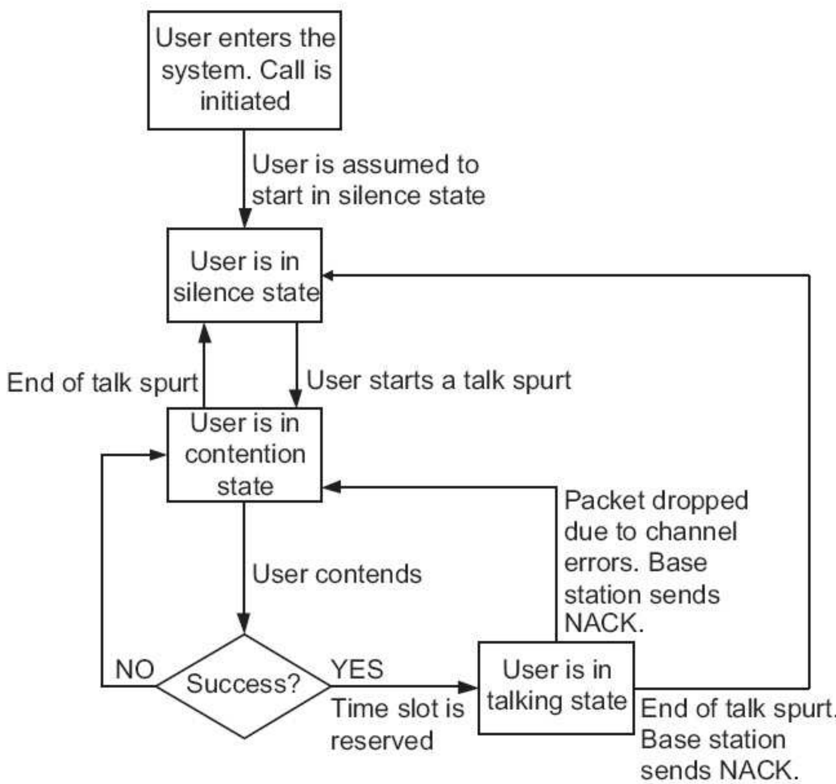

Figure 2.4: Flowchart of the PRMA protocol.

channel and assume a packet network carrying speech traffic. Medium access in the network is based on the Packet Reservation Multiple Access (PRMA) protocol [31]. The PRMA protocol can be viewed as a combination of TDMA and slotted ALOHA protocols where the channel is subdivided into time frames and each frame is in turn subdivided into N time slots. Figure 2.4 shows a simple diagram of how PRMA works. Those users in the process of starting a talk spurt contend for the channel over empty time slots, independently of each other and with a fixed access probability that we will denote as p_v. If a user is successful in the contention process, then a slot is reserved for that user; otherwise, the base station feeds back a NULL message to make the slot available for contention in the next time frame.

Users with reserved slots keep the reservation, in principle for the duration of the talk spurt, and use them to transmit their corresponding speech packets. Upon ending a talk spurt, a user enters a silence state where it is not generating or transmitting any packets. In this case, the base station feeds back a NULL message to declare that the previously reserved time slot is free again for other users to use (we assume immediate feedback of the NULL message). Note that the PRMA protocol exploits the on-off nature of speech to improve the utilization of the channel by reserving slots only to calls in a talk spurt. Nevertheless, note also that because users contend for an empty time slot with certain probability, some slots may be left unused even in situations of access congestion.

The state of every time slot (free or reserved) in the current frame is determined by the base station feedback at the end of each time slot in the previous frame. It is assumed that the feedback channel is error free, thus there is no uncertainty in the state of any time slot. Moreover, it is assumed that the base station will also feed back a NULL message in response to errors due to packet collisions during the contention process, as well as errors due to wireless channel impairments. This means that a user will lose its reservation if it faces a channel error while transmitting its packet.

Because speech communication is very sensitive to delay, speech packets require prompt delivery. In PRMA, the voice packets from calls that fail the contention to access the channel are placed in a waiting queue. If a packet remains undelivered for a pre-specified maximum delay of D_{max} frames, the packet is dropped from the user's queue.

2.1.3 Channel Model

The received signal at the base station can be written as

$$y_B = \sqrt{G r_B^{-\alpha}} h_B x + n_B; \tag{2.3}$$

similarly, the received signal at the relay

$$y_R = \sqrt{G r_R^{-\alpha}} h_R x + n_R; \tag{2.4}$$

where x is the transmitted signal, G the transmit power, assumed to be the same for all users and the relay, r_B and r_R denotes the distance from any user to the base station and to the relay, respectively, α is the path loss exponent, and h_B and h_R are the channel fading coefficients for the user-base station and user-relay links, respectively, which are modeled as zero-mean complex Gaussian random variables with unit variance. The additive noise terms n_B and n_R are modeled as zero-mean complex Gaussian random variables with variance N_0. We assume that the channel coefficients are constant for the transmission duration of one packet. In this work, we only considered the case of a symmetric network, where all the inter-users channels are assumed to be statistically identical (see Fig. 4.2).

The success and failure of packet reception is characterized by outage events and outage probabilities, which is defined as follows. For a target signal-to-noise (SNR) ratio β (called *outage SNR* [34]), if the received SNR as a function of the fading realization h is given by $\mathrm{SNR}(h)$, then the outage event O is the event that $\mathrm{SNR}(h) < \beta$, and $\Pr\{\mathrm{SNR}(h) < \beta\}$ denotes the outage probability. This definition is equivalent to the capture model in [35], [36]. The SNR threshold β is a function of different parameters in the communication system; it is a function of the application, the data rate, the signal-processing applied at encoder/decoder sides, error-correction codes, and other factors. For example, varying the data rate and fixing all other parameters, the required SNR threshold β to achieve certain system performance is a monotonically increasing function of the data rate. Also, increasing the signal-processing and encode/decoder complexity in the physical layer reduces the required SNR threshold β for a required system performance.

For the channel model in (2.3) and (2.4), the received SNR of a signal transmitted between any user and the base station can be specified as follows

$$\mathrm{SNR}_B = \frac{|\, h_B \,|^2 \, r_B^{-\alpha} G}{N_0}. \tag{2.5}$$

Since the SNR in (2.5) is a monotone function of $|\, h_B \,|^2$, the outage event for an outage SNR β is equivalent to

$$\{h_B : \mathrm{SNR}_B < \beta\} = \left\{ h_B : |\, h_B \,|^2 < \frac{\beta N_0 r_B^{\alpha}}{G} \right\}. \tag{2.6}$$

Accordingly, and knowing that $|\, h_B \,|^2$ has an exponential distribution, the outage probability is

$$P_{OB} = \Pr\left\{ |\, h_B \,|^2 < \frac{\beta N_0 r_B^{\alpha}}{G} \right\} = 1 - \exp\left(-\frac{\beta N_0 r_B^{\alpha}}{G} \right). \tag{2.7}$$

Similar relations hold for the outage probability between any user and the relay (P_{OR}).

2.2 Content-Aware Cooperative Multiple Access Protocol

Transmission errors, which are inherent to wireless communication channels, have a significant impact on the PRMA network performance [37]. On one side, if a user experiences an error while contending for access to a time slot, it would fail on the try and would have to contend again in another free slot. Moreover, if a user that already holds a reserved slot experiences an error while sending a speech packet, the user would have to give it up and go through the contention process again because the base station would send a NULL feedback upon receiving a packet with errors, which would also indicate that the reserved slot is free. These effects translate into an increase in the number of contending users and, thus, a significant increase in network traffic and in delay to gain a slot reservation. These effects ultimately severely degrade the speech quality. In fact, the congestion may reach a level where all users experience reduced speech quality due to packets dropped due to excessive delay [31].

By enabling cooperation in the speech network, one can benefit from the spatial diversity offered by cooperation to mitigate the wireless channel impairments. Here we propose the deployment of a single relay node into the network. This node will have the task of helping users holding slot reservations to forward their packets by operating in an incremental decode-and-forward mode [12]. In this mode, the relay first decodes the received packet, and then re-encodes and forwards a regenerated version of the packet to the base station if necessary. In order to decide whether to forward the packet or not, the relay utilizes limited feedback from the base station in the form of automatic repeat request (here we consider the NULL feedback

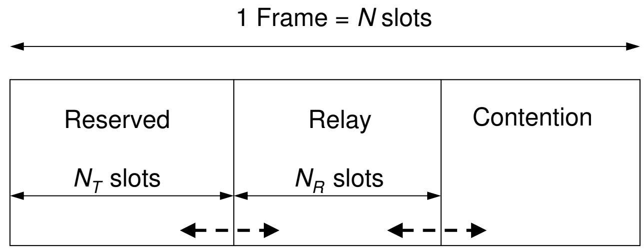

Figure 2.5: The organization of time slots in a frame.

as the repeat request message to the relay). This means that the relay will only forward the packets that were not successfully received by the base station. If the relay successfully forwards the packet to the base station then the user owning that packet will not lose its reservation and will continue sending new packets in the upcoming frames. The rationale in introducing a relay is that it would result in a more reliable end-to-end link and, hence, a reduction in the number of users losing their reserved time slots. This leads to a further reduction in the average number of contending users, and therefore, much lower access delay and packet dropping probability, which ultimately improves speech quality.

To incorporate the relay operation into the network, we propose the frame structure illustrated in Fig. 2.5. The first N_T slots create a variable size (from frame to frame) compartment of slots reserved for the talking users. Of the remaining $(N - N_T)$ free slots, a fraction p_r is assigned to the relay and the remaining free slots are made available for contention. The ordering of slots in a frame is first the N_T slots reserved for the talking users, followed by N_R slots assigned to the relay, and the remaining slots are used in the contention process.

Through the base station ACK to a successful contention, all network members will know that a new user has gained a reservation. This user's reserved slot will be appended to the end of the reserved slots compartment. When a user gives up its reservation, the base station feedback will inform all users that this time slot will be free in the next frame. To keep the reserved slots compartment contiguous, any user whose reserved slot is after the freed up slot will shift its transmission one slot earlier in the next frame. It should be noted that rearrangement of reserved slots is achieved through the information provided by the base station feedback, and no extra scheduling is required.

In any specific frame, a fraction p_r of the non-reserved slots is assigned to the relay. In some frames, when the number of failed packets is smaller than the number of relay slots, the relay might not use all its assigned slots to correct all the failed packets. In such a case, when the base station detects the correction of all failed packets, it sends a special feedback message declaring the end of the relay slots and the start of the contention slots. Moreover, since the relay is helping talking users only, no slots are assigned to the relay when there are no users with reserved slots. The maximum number of time slots assigned to the relay is then,

$$N_R = \begin{cases} 0, & N_T = 0 \\ \text{round}(p_r(N - N_T)), & N_T > 0 \end{cases} \tag{2.8}$$

It is clear that the value of p_r determines how much help the relay will offer to talking users; also it determines the reduction in the number of free time slots available for contention. Therefore, the introduction of cooperation poses a tradeoff between the amount of help the relay offers to existing users and the ability of the network to admit new users because of the reduction in the number of contention slots. Since such tradeoff is governed by p_r, the choice of the value of this parameter is crucial for the optimal performance of the system. This issue will be addressed later in this chapter.

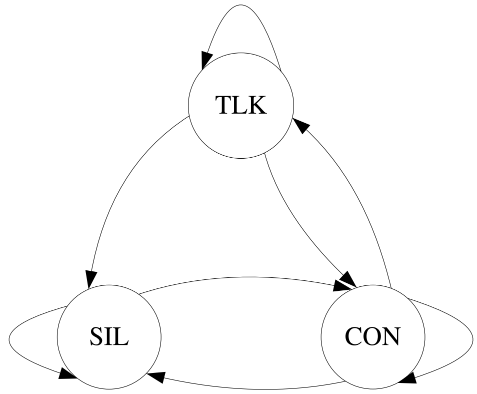

Figure 2.6: User's terminal model

2.3 Dynamic State Model

In this section, we develop an analytical model to study and measure the network performance. Based on the models discussed above, a user can be in one of three states: "SIL" when in a silence period, "CON" when contending for channel access, and "TLK" when holding a slot reservation. The dynamics of user transitions between these three states can be described by the Markov chain of Fig. 2.6 [33]. A user in SIL state moves to CON state when a new talk spurt begins. When there is an available slot, with probability p_v, a user in CON state will send the packet at the head of its queue. If contention succeeds, a user in CON state transits to TLK state, where it will have the slot reserved in subsequent frames. A user moves from CON state to SIL state if its talk spurt ends before gaining access to the channel. A user in TLK state transits to SIL state when its talk spurt ends, and transits to CON state if its packet is not received correctly by the base station. This later transition could be avoided if the relay is able to help that user.

Again, we will consider one complete frame as the time step for the Markov chain. Although the actions of different users are independent, the transition probabilities between different states for a given user are in general dependent on the number of users in CON and TLK states. These numbers will affect the probability with which a user succeeds in contention. Moreover, the number of users in TLK state will determine the number of slots assigned to the relay, and hence the relay's ability to help users.

In order to take these dependencies into consideration, the whole network will be modeled as the two-dimensional Markov chain (M_C, M_T), where M_C and M_T are random variables denoting the number of users in CON and TLK states, respectively. Assuming there are M_v users in the network, then the number of users in the SIL state is $M_S = M_v - M_C - M_T$. In what follows, we will analyze this Markov chain and calculate its stationary distribution which will allow for the derivation of different performance measures.

Let $S_1 = (M_{C_1}, M_{T_1})$, and $S_2 = (M_{C_2}, M_{T_2})$ be the system states at two consecutive frames. Then,

$$M_{C_2} = M_{C_1} + m_{SC} + m_{TC} - m_{CS} - m_{CT}, \tag{2.9}$$

$$M_{T_2} = M_{T_1} + m_{CT} - m_{TS} - m_{TC}, \tag{2.10}$$

where m_{ij} denotes the number of users departing from state $i \in \{S, C, T\}$ to state $j \in \{S, C, T\}$, for example, m_{SC} is the number of users departing from SIL state to CON state. This implies that the transition probability between any two states can be determined in terms of the distributions of m_{SC}, m_{CS}, m_{CT}, m_{TS}, and m_{TC}. Next we will calculate these distributions.

2.3.1 Distribution of m_{SC}

From section 2.1.1, and since all users are independent, the number of users making a transition from the SIL state to the CON state, m_{SC}, follows a binomial distribution with parameter σ, where σ is defined in (2.2). Then,

$$\Pr(m_{SC} = i) = \binom{M_S}{i} \sigma^i (1-\sigma)^{M_S - i}, \quad i = 0, ..., M_S. \tag{2.11}$$

2.3.2 Distribution of m_{TS}

From section 2.1.1, and from users independence, the number of users making a transition from the TLK state to the SIL state, m_{TS}, is binomially distributed with parameter γ, where γ is defined in (2.1). Then,

$$\Pr(m_{TS} = i) = \binom{M_T}{i} \gamma^i (1-\gamma)^{M_T - i}, \quad i = 0, ..., M_T. \tag{2.12}$$

2.3.3 Distribution of m_{TC}

A user leaves the TLK state to the CON state if its transmitted packet fails to reach the base station successfully, and if the relay did not help that user. Also, a user in TLK state will leave to SIL state if its talk spurt ends in the current frame irrespective of the reception state of its last transmitted packet. This means that this user will not attempt to retransmit its last packet in the talk spurt and the relay will not try to help this user.

Given the number of users making transitions from TLK state to SIL state (their talk spurt ended and have no packets to transmit), m_{TS}, the number of erroneous packets from the remaining users in the TLK state, ε, follows a binomial distribution with parameter P_{OB}, the outage probability of the link between any user and the base station as defined in (5.2). Therefore,

$$\Pr(\varepsilon = i | m_{TS}) = \binom{M'_T}{i} P_{OB}^i (1-P_{OB})^{M'_T - i}, \quad i = 0, ..., M'_T, \tag{2.13}$$

where $M'_T = M_T - m_{TS}$, the number of remaining users in the TLK state. Assume that the relay can successfully receive ε_R packets out of the ε erroneous packets.

Then, conditioned on ε, the number of successfully received packets by the relay, ε_R, is also binomially distributed but with parameter P_{OR}, the outage probability of the user-relay link, then,

$$\Pr(\varepsilon_R = i|\varepsilon) = \binom{\varepsilon}{i}(1 - P_{OR})^i P_{OR}^{\varepsilon-i}, \quad i = 0, ..., \varepsilon. \tag{2.14}$$

For each of the slots assigned to the relay, a packet among the ε_R packets in the relay's queue is selected at random and forwarded. It follows that for $\varepsilon_R \geq N_R$, the number of successfully forwarded packets ε_F is binomially distributed with parameter P_{ORB}, the outage probability of the link from relay to base station,

$$\Pr(\varepsilon_F = i|\varepsilon_R) = \binom{N_R}{i}(1 - P_{ORB})^i P_{ORB}^{N_R-i}, \quad i = 0, ..., N_R, \tag{2.15}$$

where N_R is the maximum number of time slots assigned to the relay. For $\varepsilon_R < N_R$, the distribution of the number of successfully forwarded packets follows (2.15) for $i = 0, ..., \varepsilon_R - 1$. The probability that $\varepsilon_F = \varepsilon_R$ is

$$\Pr(\varepsilon_F = \varepsilon_R|\varepsilon_R) = \sum_{i=0}^{N_R-\varepsilon_R} \binom{\varepsilon_R + i - 1}{i} P_{ORB}^i (1 - P_{ORB})^{\varepsilon_R}, \tag{2.16}$$

which accounts for all the possible combinations of successful and failed packet transmissions before the ${\varepsilon_R}^{th}$ successful packet.

Finally, the probability that i users make the transition from TLK state to CON state is the probability that from the ε erroneous packets, the relay successfully forwards $(\varepsilon_F = \varepsilon_R - i)$ packets. Then the distribution of m_{TC} is given by

$$\Pr(m_{TC} = i|m_{TS}) = \sum_{k=i}^{M_T-m_{TS}} \sum_{l=k-i}^{k} \Pr(\varepsilon_F = \varepsilon_R - i|\varepsilon_R = l)\Pr(\varepsilon_R = l|\varepsilon = k) \times \Pr(\varepsilon = k|m_{TS}). \tag{2.17}$$

2.3.4 Distribution of m_{CT}

Upon a successful contention, a user transits from the CON to the TLK state. This transition occurs at the end of each free slot where contention can take place. Thus, the number of contending users will vary from slot to slot. Suppose there are M_T reserved slots and the relay uses m_R slots in a given frame. Then there are $(N - M_T - m_R)$ free slots for contention. We want to calculate the distribution of the number of users that moved from CON state to TLK state at the end of the last free slot. This distribution could be calculated using the following recurrence model. Let $q(M'_C)$ be the probability that a user succeeds in contention when there are M'_C contending users, and p_v the users' channel access probability, then,

$$q(M'_C) = M'_C p_v (1 - p_v)^{M'_C - 1}(1 - P_{OB}), \tag{2.18}$$

which is the probability that only one user has permission to transmit and the channel was not in outage during packet transmission.

Define $R_k(M'_C)$ as the probability that M'_C users remain in the CON state at the end of the k^{th} available slot, $(k = 0, 1, 2, ..., N - M_T - m_R)$. Conditioning on the outcome of the $(k-1)^{st}$ time slot, it follows that,

$$R_k(M'_C) = R_{k-1}(M'_C)[1 - q(M'_C)] + R_{k-1}(M'_C + 1)q(M'_C + 1),$$
$$M'_C = 0, 1, ..., M_C, \tag{2.19}$$

where M_C is the number of users in the CON state at the beginning of the frame. The initial condition for this recursion is

$$R_0(M'_C) = \begin{cases} 1, & M'_C = M_C \\ 0, & M'_C \neq M_C \end{cases} \tag{2.20}$$

and the boundary condition $q(M_C+1)=0$, which follows from the fact that the total number of contending users is M_C. Finally, the distribution of m_{CT} is

$$\Pr(m_{CT}=i|m_R) = R_{N-M_T-m_R}(M_C-i), \quad i=0,...,M_C, \tag{2.21}$$

i.e., the probability that i users succeed in contention in the current frame is equal to the probability that (M_C-i) users remain in the contention state at the end of that frame.

To calculate the distribution of m_R, the number of time slots the relay actually uses out of its N_R assigned slots, we condition on ε_R and ε_F, the number of packets in the relay queue and the number of packets successfully forwarded by the relay, respectively. The relay will use all of the N_R assigned slots to successfully forward ε_F packets in the following two cases; (i) $\varepsilon_R \geq N_R$, i.e., the relay has more packets to forward than it has assigned time slots, (ii) $\varepsilon_R < N_R$ but $0 \leq \varepsilon_F < \varepsilon_R$, which means that the relay has enough time slots, but is unable to successfully forward all the packets in its queue (due to transmission errors). Therefore, $\Pr(m_R = N_R|\varepsilon_R, \varepsilon_F) = 1$.

For the remaining case where $\varepsilon_R < N_R$ and $\varepsilon_F = \varepsilon_R$, it follows from (2.16) by a simple change of variables that

$$\Pr(m_R = i|\varepsilon_R, \varepsilon_F) = \binom{i-1}{i-\varepsilon_R} {P_{ORB}}^{i-\varepsilon_R}, \quad i=\varepsilon_R,...,N_R \tag{2.22}$$

which is the probability that there are i failed transmissions before the ${\varepsilon_R}^{th}$ successful transmission.

2.3.5 Distribution of m_{CS}

A user makes a transition from the CON state to the SIL state if its talk spurt ends before gaining access to the channel. Conditioning on the number of users that successfully accessed the channel, m_{CT}, and through the same argument as in 2.3.2, we have

$$\Pr(m_{CS} = i | m_{CT}) = \binom{M_C - m_{CT}}{i} \gamma^i (1-\gamma)^{M_C - m_{CT} - i}, \quad i = 0, ..., M_C - m_{CT}. \tag{2.23}$$

In what follows, we will not seek to remove the conditioning on m_{CT} from this distribution because this is the form we will be interested in when calculating the state transition matrix later in this chapter.

At this point, it is important to remark that all the distributions calculated above are state dependent because they generally depend on M_C and M_T. This means that we have to calculate a different set of distributions for each possible state of the system.

2.3.6 State transition probabilities

Here we consider the state transition matrix $\mathbf{P}$. An element $P(S_1, S_2)$ of this matrix is the transition probability from state $S_1 = (M_{C_1}, M_{T_1})$ to state $S_2 = (M_{C_2}, M_{T_2})$. It easily follows from (2.9) and (2.10), and the distributions developed above that the transition probability $P(S_1, S_2)$ is given by

$$\begin{aligned} P(S_1, S_2) = \sum_{x=0}^{M_{C_1}} \sum_{y=0}^{M'} \sum_{z=0}^{M_{T_1}} & \Pr(m_{CS} = x | m_{CT} = y, S_1) \\ & \times \Pr(m_{TC} = M_{T_1} - M_{T_2} + y - z | m_{TS} = z, S_1) \\ & \times \Pr(m_{SC} = M_{C_2} - M_{C_1} + x + y - z | S_1) \\ & \times \Pr(m_{CT} = y | S_1) Pr(m_{TS} = z | S_1), \end{aligned} \tag{2.24}$$

where $M' = \min(M_{C_1} - x, N - M_{T_1} - N_{R_1})$, since the number of CON to TLK transitions, m_{CT}, cannot exceed the number of users remaining in the CON state after the CON to SIL transitions, or the total number of available time slots for contention. It should be noted that,

$$P(S_1, S_2) = 0, \quad \text{if } M_{T_2} > \min(M_{T_1} + M_{C_1}, N), \tag{2.25}$$

since the number of users in TLK state in the next frame cannot exceed the total number of time slots in a frame or the number of users in TLK and CON states in the current frame. Also,

$$\Pr(m_{TC} = M_{T_1} - M_{T_2} + y - z|S_1) = 0, \quad \text{if } M_{T_1} - M_{T_2} + y - z > M_{T_1} - z, \tag{2.26}$$

because the number of users leaving the SIL state cannot be larger than the number of users initially in this state. And,

$$\Pr(m_{SC} = M_{C_2} - M_{C_1} + x + y - z|S_1) = 0, \quad \text{if } M_{C_2} - M_{C_1} + x + y - z > M_{S_1}, \tag{2.27}$$

because the number of users leaving the TLK state to the CON state cannot exceed the difference between the number of users in the TLK state and the number of users moving from TLK state to SIL state.

cannot be larger than the number of users initially in this state.

Finally, the stationary distribution vector $\boldsymbol{\pi}$ can be calculated as the left eigenvector of the maximum eigenvalue of the matrix $\boldsymbol{P}$, or in other words, can be obtained by solving $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$.

2.4 Performance Analysis

To assess the performance of the speech network under our proposed cooperative protocol, four measures will be considered: network throughput, multiple access delay, packet dropping probability, and speech quality.

2.4.1 Network Throughput

The throughput can be defined as the aggregate average amount of data transported through the channel in a unit time. In our case, the number of packets successfully transmitted in a given frame can be decomposed into two components linked by the tradeoff between the use of cooperation and reduction in the number of contention slots. The first component corresponds to the contending users who succeed in gaining access to the channel. The second corresponds to the talking users who succeed in transmitting their packets to the base station, either directly or through the help of the relay. Thus, the throughput can be expressed as

$$Th = \frac{E\left\{E\left\{m_{CT}|S_1\right\} + M_T - E\left\{m_{TC}|S_1\right\}\right\}}{N}, \tag{2.28}$$

where $E\{\cdot\}$ is the expectation operator. The term $E\left\{m_{CT}|S_1\right\}$ corresponds to the successful contention, whereas the number of successfully transmitted packets is expressed as $(M_T - E\left\{m_{TC}|S_1\right\})$, the number of users in TLK state minus the expected number of users leaving the TLK state to the CON state, which are the users with failures in their transmissions. Finally, the outermost expectation is with respect to the stationary distribution of the system's Markov chain.

2.4.2 Multiple Access Delay

The delay is the number of frames a user spends in the CON state before gaining access to the channel. This delay is a function of the probability with which a user succeeds in contention during a given frame. This success probability depends on the network state at the instant the user enters the CON state, and will differ from frame to frame according to the path the network follows in the state space. Therefore, for exact evaluation of the multiple access delay, one should condition on the state at which the user of interest enters the CON state for the first time. Starting from this state, the delay is obtained from the calculation of the statistics of all possible paths the network follows in the state space till the user succeeds in the contention process. It is possible to show that for a network with N time slots per frame and M_v users, the total number of states is given by $(M_v - N/2 + 1)(N+1)$ for $M_v \geq N$. For a network with $M_v = N = 10$, the number of states is 66. With such large number of states, finding an exact expression for the multiple access delay becomes prohibitively complex.

To get an approximate expression for the delay, we will assume that when the user enters the contention state the system state will not change until that user succeeds in contention. Thus, the success probability will be constant throughout the whole contention process, and the delay at any given state will follow a geometric distribution with parameter $p_s(i)$, the success probability at any state i. The approximate average delay is given by

$$D_{avg} = \sum_{i \in \Omega} \frac{\pi(i)}{p_s(i)}, \tag{2.29}$$

where Ω is the set of states where $M_C \neq 0$ and $\pi(i)$ is the i^{th} element of the stationary distribution vector $\boldsymbol{\pi}$.

The last step is to calculate the success probability $p_s(i)$. Given the assumption that all users are statistically identical, the probability that a user succeeds during contention in a given frame is equal to the probability that at least one user succeeds during contention in that frame. Given that the frame starts with M_C contending users, and from the recursion of (2.19), the probability that no user succeeds in contention is

$$R_{N-M_T-N_R}(M_C) = \left(1 - M_C p_v (1-p_v)^{M_C - 1}(1 - P_{OB})\right)^{(N-M_T-N_R)}. \tag{2.30}$$

Therefore, the probability that at least one user succeeds in contention is

$$\begin{aligned} p_s(i) =& 1 - R_{N-M_T-N_R}(M_C) \\ =& 1 - \left(1 - M_C p_v (1-p_v)^{M_C-1}(1-P_{OB})\right)^{(N-M_T-N_R)}. \end{aligned} \tag{2.31}$$

2.4.3 Packet Dropping Probability

Speech communication is delay sensitive and requires prompt delivery of speech packets. In the PRMA protocol, packets start to be dropped if they are delayed in the network for more than a maximum allowable delay of D_{max} frames. Based on the assumption that the speech coder generates exactly one speech packet per frame, every user will maintain a queue of length D_{max}. Whenever the queue is full at the start of a frame, the oldest packet is dropped until the user succeeds in reserving a time slot. After gaining a reservation, in each frame the oldest packet in the queue is transmitted and the new incoming packet is added at the end of the queue. If the talk spurt ends before getting a slot reservation, all the packets in the queue are dropped. Because of channel errors, a user with a reserved time slot may lose its reservation and return to the group of contending users, thus risking further packet dropping.

To analyze the packet dropping probability, we adopt the method developed in [38] and [39] for the analysis of the PRMA protocol. First, we focus on the case when a user is trying to access the channel for the first time. Given that the system is in state i with M_C contending users and M_T users holding slot reservations, consider a contending user whose talk spurt started at the current frame. The talk spurt consists of L packets, where L is a random variable. The user will start to contend for a time slot in the current frame and continue in subsequent frames until it succeeds or the talk spurt ends. The user waits in the

CON state for D frames to obtain a reservation. Using the assumption developed in section 2.4.2 that the delay D is geometrically distributed, the probability that a user waits for d frames given the system is in state i is

$$P_D(d|i) = (1 - P_s(i))(P_s(i))^d, \quad d = 0, 1, ... \tag{2.32}$$

We need to distinguish between two different cases relating the length of the talk spurt L and the maximum allowable delay D_{max}. Moreover, we should note the assumption that when a user transits to the silence state all remaining packets in the buffer are dropped.

1. $L \leq D_{max}$: In this case, the buffer is long enough to store the whole talk spurt. If reservation is obtained before the talk spurt ends, j packets are lost if the transition from TLK to SIL occurred after the $(L - j)^{th}$ transmission which has a probability of $\gamma(1-\gamma)^{L-j}$. Otherwise, all the talk spurt packets are discarded. As a function of the waiting time d, the number of dropped packets is

$$n_d(d) = \begin{cases} j, & 0 \leq d < L \\ L, & d \geq L \end{cases} \tag{2.33}$$

and the distribution of the number of dropped packets is given by

$$\Pr\{n_d|L \leq D_{max}, i\} = \begin{cases} \gamma(1-\gamma)^{L-j} \sum_{d=0}^{L} P_D(d,i), & n_d = 0 \\ \sum_{d=L+1}^{\infty} P_D(d,i), & n_d = L \end{cases} \tag{2.34}$$

2. $L > D_{max}$: In this case, after waiting D_{max} frames, one packet is dropped per frame until slot reservation is achieved. The dropped packet is the oldest in the queue with an associated delay of D_{max}. The number of dropped packets

as a function of the delay is given by

$$n_d(d) = \begin{cases} j, & 0 \leq d \leq D_{max} - 1 \\ k, & d = D_{max} + k - 1, \\ & k = 1, 2, ..., (L - D_{max}) \\ L, & d \geq L \end{cases} \tag{2.35}$$

and its distribution

$$\Pr\{n_d | L > D_{max}, i\} = \begin{cases} \gamma(1-\gamma)^{L-j} \sum_{d=0}^{D_{max}-1} P_D(d,i), & n_d = 0 \\ (1 - P_s(i))(P_s(i))^{n_d}, & n_d = 1, 2, ..., (L - D_{max}) \\ \sum_{d=L}^{\infty} P_D(d,i), & n_d = L \end{cases} \tag{2.36}$$

We note here that although all the summations mentioned above have closed form expressions, they tend to become complex and lengthy. Therefore, we avoid writing them here, so as to keep the presentation compact. This will apply to the next section.

The expected number of dropped packets for the above two cases, namely $E\{n_d | L \leq D_{max}, i\}$ and $E\{n_d | L > D_{max}, i\}$, can be easily calculated using the corresponding distributions and then combined to get the total expected number of dropped packets as

$$E\{n_d | i\} = \sum_{l=1}^{D_{max}} E\{n_d | L \leq D_{max}, i\} P_L(l) + \sum_{l=D_{max}+1}^{\infty} E\{n_d | L > D_{max}, i\} P_L(l), \tag{2.37}$$

where $P_L(l)$ is the probability mass function of the length of the talk spurt. From the speech source model discussed in section 2.1.1, the talk spurt duration, L, is geometrically distributed with parameter γ, i.e.,

$$P_L(l) = \gamma(1-\gamma)^{l-1}, \quad l = 1, 2, ... \tag{2.38}$$

Finally, the packet dropping probability is the ratio of the average number of dropped packets per talk spurt to the average number of packets generated per talk spurt, i.e,

$$P_{d_0} = \frac{1}{\gamma} \sum_{i \in \Omega} E\{n_d | i\} \pi(i), \tag{2.39}$$

where the sum is over Ω, the set of states with $M_C \neq 0$ (since packets are dropped only when the user is in the CON state).

Next we consider the packet dropping probability due to the first transition from the TLK state to the CON state, which is caused by channel errors. First, we need to make the following assumptions:

- Any user in TLK state has obtained its reservation with the first packet in the talk spurt. This means no packets were dropped in the first contention process. Furthermore, this packet is delayed by $D_0 = D_{avg}$ frames, i.e., this packet is delayed by the average multiple access delay calculated in section 2.4.2.

- The first channel error occurs while transmitting the j^{th} packet of the talk spurt. Since the first packet was delayed by D_0 frames, the remaining maximum delay for the subsequent packets in the talk spurt is $D_1 = D_{max} - D_0$ frames.

- There are L packets in the talk spurt, and L_1 packets following and including the j^{th} packet which encountered a channel error.

Based on the time instant when the user left TLK state to CON state, we need to analyze three cases:

Case 1:: Transmission instant of the j^{th} packet is after the end of the talk spurt. This means, $D_0 + (j-1) \geq L$, or $L_1 \leq D_0$. In this case all the remaining

L_1 packets are discarded without any contentions and

$$E\{n_d|L_1 \leq D_0, i\} = L_1.$$

Case 2: Transmission instant of the j^{th} packet is before the end of the talk spurt and the remaining time till the end of the talk spurt is less than the maximum remaining delay D_1. That is, $0 < L - D_0 - (j-1) \leq D_1$, or $D_0 < L_1 \leq D_{max}$. In this case, no packets are dropped if the user gets a reserved slot before the end of the talk spurt. Otherwise, all L_1 are discarded.The number of dropped packets as a function of the waiting time is

$$n_d(d) = \begin{cases} 0, & 0 \leq d \leq L_1 - D_0 \\ L_1, & d > L_1 - D_0 \end{cases} \tag{2.40}$$

and its distribution

$$Pr\{n_d|D_0 < L_1 \leq D_{max}, i\} = \begin{cases} \sum_{d=0}^{L_1-D_0} P_D(d), & n_d = 0 \\ \sum_{d=L_1-D_0+1}^{\infty} P_D(d), & n_d = L_1 \end{cases} \tag{2.41}$$

Case 3: $L - D_0 - (j-1) > D_1$, or $L_1 > D_{max}$. In this case, the j^{th} packet is dropped after waiting for D_1 frames and a packet will be dropped every frame till the user gets access to the channel. If the talk spurt ends before accessing the channel, all the packets in the buffer are discarded. Therefore,

$$n_d(d) = \begin{cases} 0, & 0 \leq d \leq D_1 - 1 \\ k, & d = D_1 + k - 1, \quad k = 1, 2, ..., (L_1 - D_{max}) \\ L_1, & d > L - D_0 \end{cases} \tag{2.42}$$

and

$$Pr\{n_d|L_1 > D_{max}, i\} = \begin{cases} \sum_{d=0}^{D_1-1} P_D(d,i), & n_d = 0 \\ (1-P_s(i))(P_s(i))^{n_d}, & n_d = 1,2,...,(L_1 - D_{max}) \\ \sum_{d=L_1-D_0}^{\infty} P_D(d,i), & n_d = L_1. \end{cases} \tag{2.43}$$

Having the distributions of the number of dropped packets for each case, one can calculate the corresponding expected number of dropped packets, $E\{n_d|L1 \leq D_0, i\}$, $E\{n_d|D_0 < L_1 \leq D_{max}, i\}$, and $E\{n_d|L_1 > D_{max}, i\}$.

The next step is to average with respect to L_1, the number of remaining packets in the talk spurt after the first error. From the earlier assumptions, we have

$$Pr\{L_1 = l_1|L = l\} = u(1-u)^{(l-l_1-1)}, \quad l_1 = 1, 2, ..., l-1, \tag{2.44}$$

where u is the user's transition probability from the TLK state to the CON state. A user leaves the TLK state to the CON state if: (i) packet transmission failed, (ii) relay did not help that user, and (iii) talk spurt did not end during current frame (had the talk spurt ended, the transition would have been to the SIL state). Now, we need to consider the talk spurt length L. If $(L-1) \leq D_{max}$, cases 1 and 2 above would occur, otherwise all three cases would occur. Therefore,

$$E\{n_d|L, i\} = E\{n_d|L-1 \leq D_{max}, i\} + E\{n_d|L-1 > D_{max}, i\}, \tag{2.45}$$

where

$$E\{n_d|L-1 \leq D_{max}, i\} = \sum_{L_1=1}^{D_0} E\{n_d|L_1 \leq D_0, i\} P_r\{L_1|L\} + \sum_{L_1=D_0+1}^{D_{max}} E\{n_d|D_0 < L_1 \leq D_{max}, i\} P_r\{L_1|L\}, \tag{2.46}$$

and

$$E\{n_d|L-1>D_{max},i\} = \sum_{L_1=1}^{D_0} E\{n_d|L_1 \le D_0, i\}P_r\{L_1|L\} + \sum_{L_1=D_0+1}^{D_{max}} E\{n_d|D_0 < L_1 \le D_{max}, i\}P_r\{L_1|L\} + \sum_{L_1=D_{max}+1}^{L-1} E\{n_d|L_1 > D_{max}, i\}P_r\{L_1|L\}. \quad (2.47)$$

Finally,

$$E\{n_d|i\} = \sum_{L=2}^{D_{max}+1} E\{n_d|L-1 \le D_{max}, i\}P_r\{L\} + \sum_{L=D_{max}+2}^{\infty} E\{n_d|L-1 > D_{max}, i\}P_r\{L\}, \quad (2.48)$$

where $P_r\{L\}$ is defined in (2.38). As in (2.39), the packet dropping probability due to the first error in the talk spurt is

$$P_{d_1} = \frac{1}{\gamma}\sum_{i\in\Omega} E\{n_d|i\}\pi(i). \quad (2.49)$$

After regaining access to the channel, a second transmission error may occur and the user has to go through contention again and may lose some packets. This process may be repeated several times till the end of the talk spurt. The number of such cycles is a random variable because of the random nature of channel errors. Furthermore, calculation of the packet dropping probability due to the second and later errors becomes intractable due to the more complex scenarios to be considered. To deal with this issue, the following approximation will be used [39]. Let k be the average number of transitions from TLK to CON states during a talk spurt, $k = u/\gamma$, where u is the user's transition probability from the TLK state to the CON state. If u is small, $k \le 1$, and P_d can be approximated by

$$P_d = P_{d_0} + P_{d_1},$$

and if u is large, then $k > 1$ and P_d can be approximated by

$$P_d = P_{d_0} + kP_{d_1}.$$

2.4.4 Speech quality measure

We will base the voice quality assessment of our protocol on the predictive model developed in [40]. This model uses source codec parameters, end-to-end delay and packet dropping probability to predict the value of the conversational Mean Opinion Score (MOS_c) [41], a perceptual voice quality measure based on the ITU-T PESQ quality measure standard [42, p. 862] that takes values in the range from 1 (bad quality) to 5 (excellent quality). For the GSM AMR 12.2 kbps voice codec [43], the (MOS_c) can be estimated using [40] as,

$$\begin{aligned} MOS_c = 3.91 &- 0.17P_d + 1.57 \cdot 10^{-3} D + 6.51 \cdot 10^{-3} P_d^2 \\ &- 2.40 \cdot 10^{-5} D^2 - 7.53 \cdot 10^{-6} P_d D - 10^{-4} P_d^3 \\ &+ 2.62 \cdot 10^{-8} D^3 + 1.38 \cdot 10^{-7} P_d D^2 \\ &- 5.51 \cdot 10^{-8} P_d^2 D, \end{aligned} \tag{2.50}$$

where P_d is the packet dropping probability and D is the average delay we calculated earlier.

2.5 Numerical Results and Discussions

In this section, we compare the performance of the proposed cooperative multiple access protocol and the PRMA protocol without cooperation. The parameters settings are as follows. The speech source model has a mean talk spurt and a mean silence period duration of $t_1 = 1$ and $t_2 = 1.35$ seconds, respectively. The speech encoder has a 12.2 kbps data rate, and we assume each packet carries 114 bits of speech data as in the GSM system. Therefore, if each user sends a single packet per frame, the frame duration is 9.35 ms. The maximum allowable delay is 20 ms, i.e., $D_{max} = 2$ frames; this value of 20 ms is chosen based on the acceptable delay for conversational interactive speech. For this setup it is possible to accept delays of up to 100-150 ms [44]. However, there are other sources of delay such as coding delay (typically 20 ms), network delay, delay at other transcoders in the network, echo cancelers, etc.; so a value of 20 ms is a safe choice to ensure good end-to-end delay behavior. Each frame is divided into $N = 10$ time slots, contention permission probability $p_v = 0.3$, SNR threshold $\beta = 15$ dB and path loss exponent $\alpha = 3.7$. The distance between any user and the base station is 100 m, between any user and the relay is 50 m, and between the relay and base station is 100 m.

Figures 2.7-2.9 show the different performance measures vs. transmit power for a fixed number of users $M_v = 15$. It is noted from the figures that there is a good match between the analytical and simulation results, which validates our derived analytical expressions. Fig. 2.7 depicts the gain in throughput due to cooperations. For example, at a low power level of 50 mW, the non-cooperative throughput is around 0.35 while the cooperative throughput with $p_r = 0.3$ is around 0.53, which amounts to an 80% increase.

The gains in delay and packet dropping probability are depicted in Figures 2.8 and 2.9, respectively, where again we can see significant decrease in delay and packet dropping probability in the low power region. It is noted that increasing the power decreases cooperation gains, which is due to the fact that at low power levels the performance is limited by outage events, which is where the relay plays a

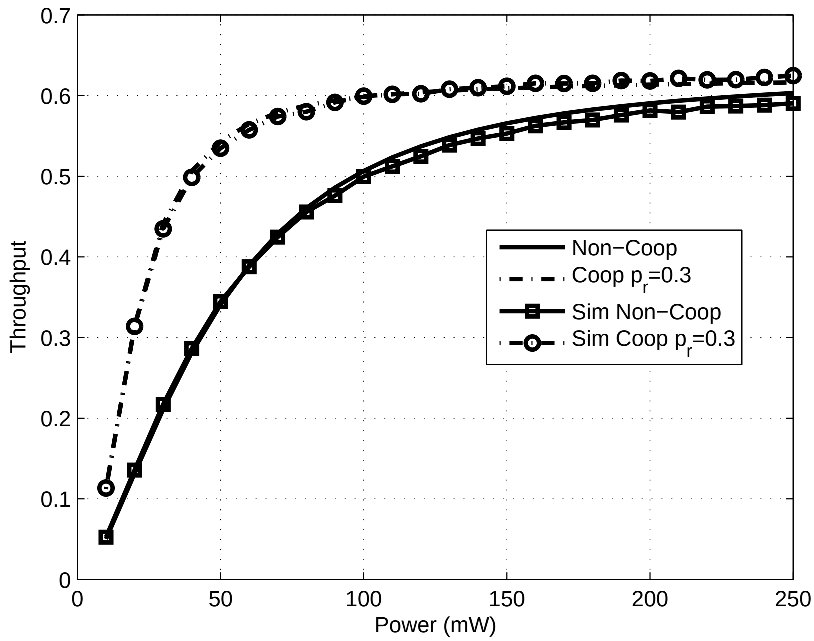

Figure 2.7: Network performance measures for 15 users and transmission power varying from 10mW to 250mW: Throughput.

role in reducing the probability of such events. On the other hand, at high power levels outage probability is low and the performance is limited by packet collisions.

For 75 mW transmission power, Fig. 2.10 shows the throughput against number of users. A significant gain is achieved in throughput and in the number of users maximizing this throughput. We see a 45 % increase in throughput and the number of users maximizing the throughput increases from 14 to 20 users. But in a speech network the maximum number of supported users should be defined by the speech quality and not only the network throughput. Fig. 2.11 depicts the mean opinion score (MOS) speech quality measure against the number of users. At MOS of 3.5 for example, which is an acceptable quality, we see that our protocol increases the number of supported users from 23 to 26, or a 13 % increase in the number of users.

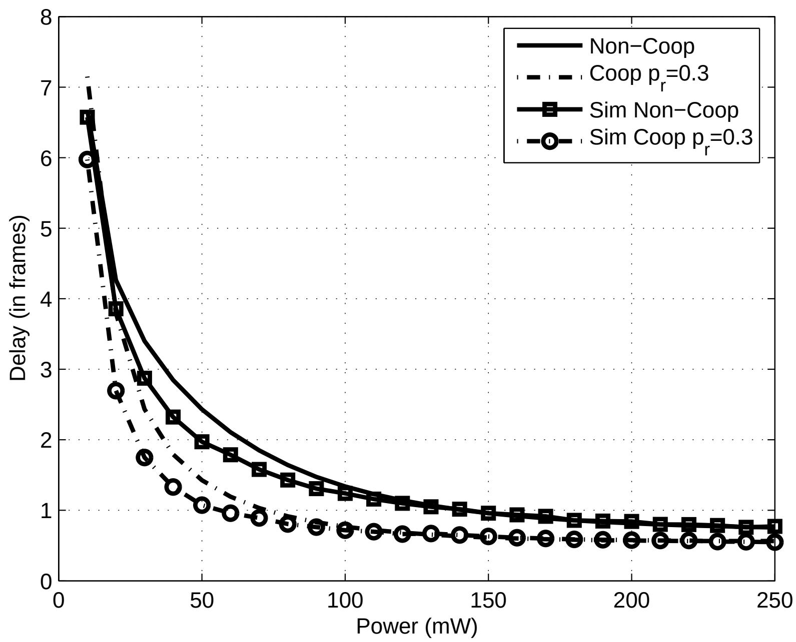

Figure 2.8: Network performance measures for 15 users and transmission power varying from 10mW to 250mW: Approximate delay of (2.29).

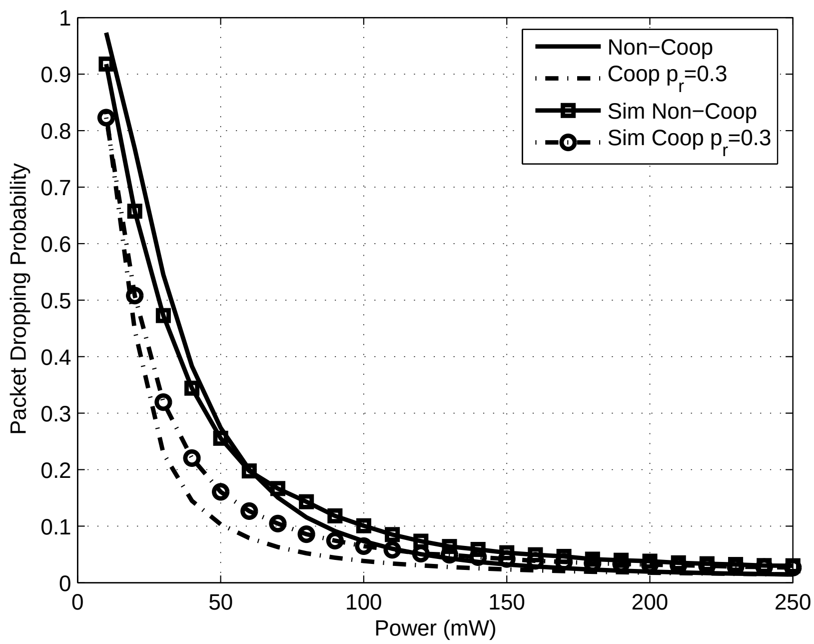

Figure 2.9: Network performance measures for 15 users and transmission power varying from 10mW to 250mW: Packet dropping probability.

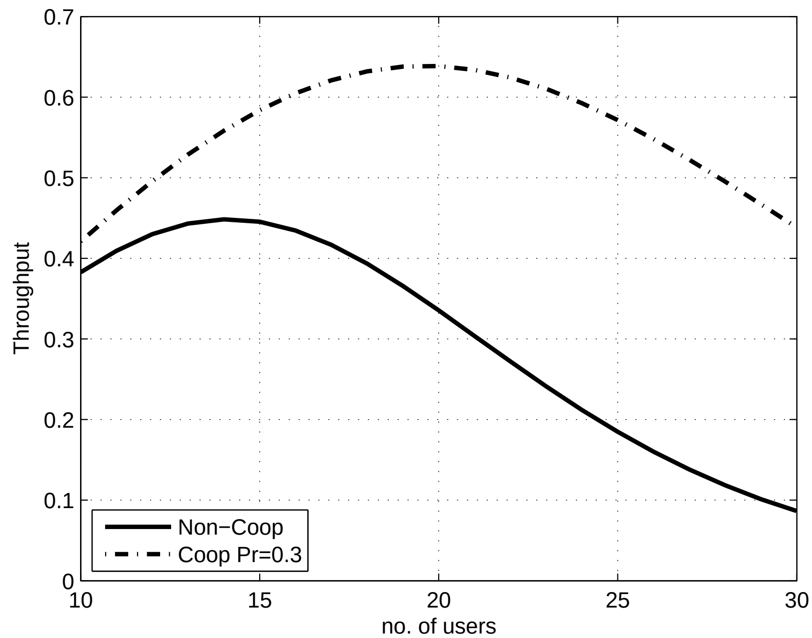

Figure 2.10: Throughput vs. number of users.

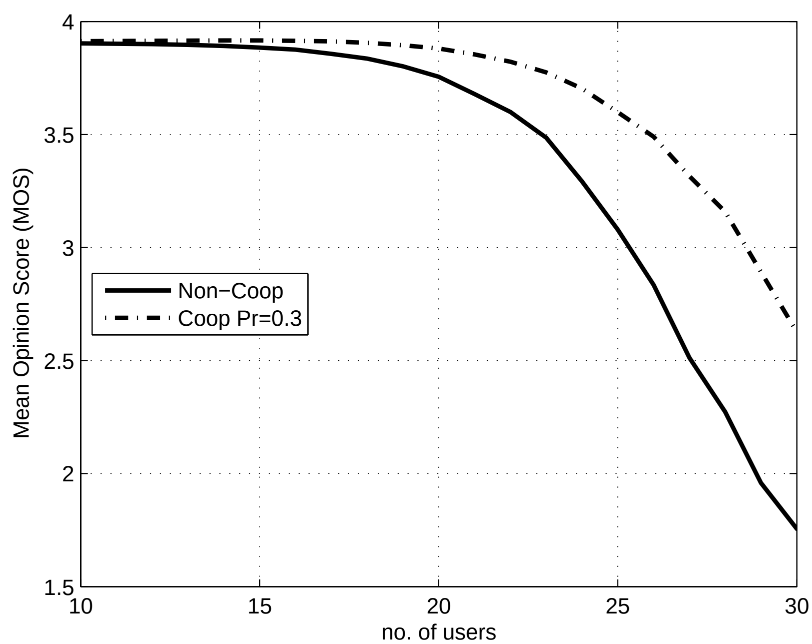

Figure 2.11: Speech quality vs. number of users.

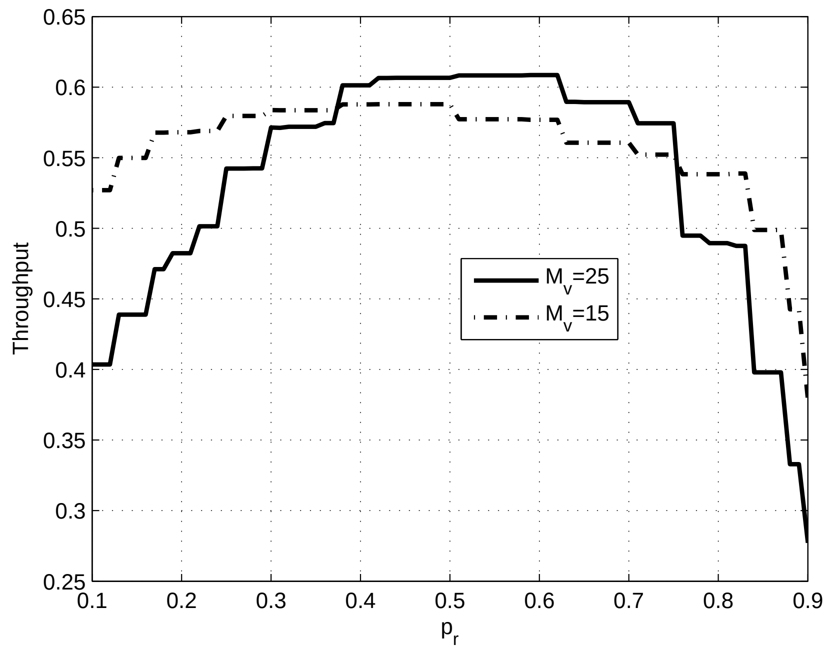

Figure 2.12: Throughput as a function of p_r for 75mW transmission power.

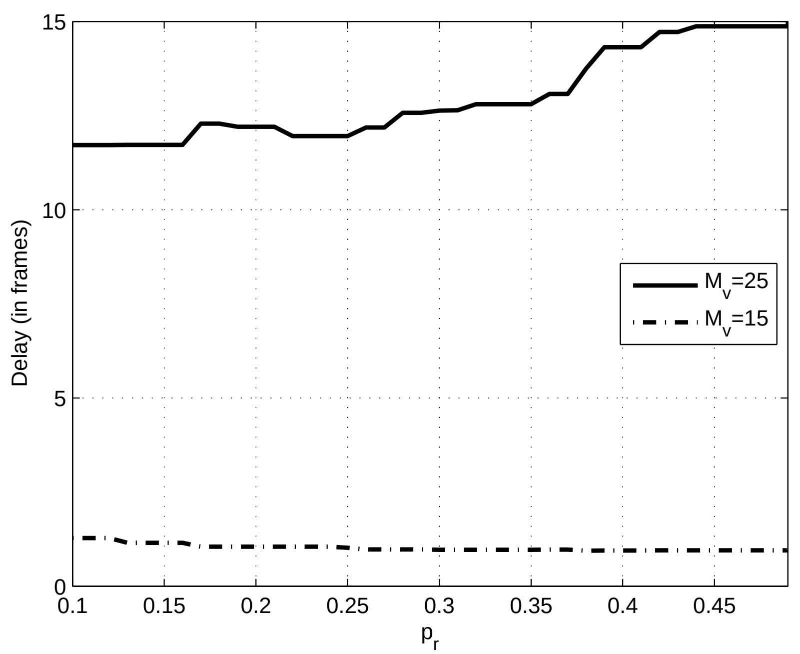

Figure 2.13: Delay as a function of p_r for 75mW transmission power .

To study the effect of the amount of resources assigned to the relay, Fig. 2.12 depicts the throughput as a function of p_r for the case of a congested network ($M_v = 25$) and a network with a moderate number of users ($M_v = 15$). It is noted that throughput is initially increasing with p_r, since increasing the p_r increases the relay's ability to help more users combat channel fading, hence decreasing outage probability and increasing the average number of successful packets per frame. Then throughput starts to decrease as p_r increases because of the network's inability to accept new users since the relay is occupying a larger portion of the contention slots, thus leading to a reduction in the average number of TLK users and a reduction in throughput. Delay performance as a function of p_r is shown in Fig. 2.13. While for a moderately loaded network the delay decreases with increasing p_r (up to the value of $p_r = 0.5$ after which delay increases dramatically), a congested network suffers from an increase in delay, which is associated with increased packet dropping probability and decreased speech quality. This is mainly due to the reduction in the number of contention slots in favor of the relay. This effect appears in the congested network only because of the larger average number of contending users compared to the moderately loaded case. Therefore, the introduction of cooperation introduces a tradeoff between the amount of help provided by the relay, and the network's ability to serve users starting a talk spurt. From Figures 2.12 and 2.13 we can see that by assigning about 30% to 50% of the free resources to the relay good throughput performance is achieved while the delay is kept at an acceptable level.

It should be noted that in this chapter we assumed a perfect feedback channel. However, situations may arise where the feedback message could take longer than expected or be received in error by the relay or network users. Since it is through feedback messages that users and relay determine the state of each slot in the upcoming frame and whether a packet needs retransmission or not, delays or errors in these messages lead to ambiguity in the state of different time slots and packet transmissions. One possible solution to deal with such an imperfect feedback channel is to make different nodes take different actions in response to lost or delayed feedback. For instance, a delayed or lost feedback after a packet transmission by a user with a reserved slot can be considered by that user and the relay as a NACK message; thus this packet will be considered for help by the relay or retransmission by the user in the next frame, while for other users this should be considered as an ACK message so the time slot involved will still be considered as reserved in the next frame and no collisions occur with the original user's transmissions. For a contention slot, the situation should be different. All users shall assume the delayed or lost feedback as a NACK message, so no slot reservation is made and any user involved in the contention process will retry in the next time slot.

Chapter 3

Opportunistic Multiple Access for Cooperative and Cognitive Networks

In chapter 2, a multiple access protocol that makes use source traffic characteristics to enable cooperation was considered. Specifically, silence periods in a speech signal were exploited to enable relay transmissions helping network users, or could be used to admit new users to the network. Furthermore, the tradeoff associated with sharing the idle time slots between the relay and new network users was investigated.

In a general data network, sources are also bursty in nature. Therefore, the periods of silence in which the source is not transmitting any data could be exploited to enable cooperation without incurring any bandwidth efficiency losses [19]. As the idle resources in the speech network of chapter 2 were used to admit new users, idle resources in the data network might also be used to introduce a secondary network to share these idle resources. This concept is well known as dynamic spectrum sharing, or cognitive radios [45].

From the analogy with the speech network of chapter 2, it is clear that cooperative communications and cognitive radios are closely related problems in the sense that the available unused or under-used channel resources can be utilized to improve the primary system performance via cooperation, or it can be shared by a secondary system to transmit new information. Despite this fact, these two problems have been studied independently.

In this chapter, we try to answer the questions: How can the under-utilized channel resources be shared between cooperative relays and cognitive secondary nodes? How does the coexistence of primary relays and secondary nodes affect the performance of both primary and secondary networks? And, what is the fundamental tradeoff between them? At a first glance one might jump to the conclusion that since relays are part of the primary network thus having higher priority over secondary nodes, then the primary network will benefit from cooperation while secondary nodes will suffer from reduced channel access opportunities. We will prove that this argument is not correct, and that even in the situation of interfering relays and secondary transmissions, both networks will benefit from the presence of relays in terms of maximum stable throughput.

To answer the questions posed above, we consider the uplink of a TDMA network as the primary network, and start by studying how cognitive relays can exploit the empty time slots to offer help to the primary nodes. We address the problem of how multiple relays share the resources among themselves, as well as how relays are assigned to primary nodes, by proposing two different relay assignment schemes. Then the presence of secondary nodes and their interaction with the primary network is considered. This chapter mainly focuses on the opportunistic multiple access aspect of the cognitive radio problem, as opposed to its dynamic spectrum sharing aspect.

To access the channel, secondary nodes depend on sensing the channel for primary activity, either primary nodes or relay nodes transmissions. Since sensing is not perfect, collisions might occur between primary and secondary transmissions. In order to have an upper and lower bound on the system's performance, two extreme cases are considered in this chapter. The first is when secondary nodes have the ability to perfectly sense relays' transmissions, and thus access the channel when all primary nodes and relay nodes queues are empty. In the second case secondary nodes cannot sense relays' transmissions at all. Since the cognitive principle is based on the idea that the presence of the secondary system should be transparent to the primary system (in this case both primary and relay nodes), appropriate countermeasures should be adopted at the secondary nodes to minimize interference with relay transmissions.

Because of the possible collisions between secondary and relay transmissions, the nodes queues are interacting. To analyze the stability of the system's queues we resort to a stochastic dominance approach. Analyzing the stability of interacting queues is a difficult problem that has been addressed for ALOHA systems initially in [46]. Later in [47], the dominant system approach was explicitly introduced and employed to find bounds on the stable throughput region of ALOHA with a collision channel model. Many other works followed that to study the stability of ALOHA. In [48], necessary and sufficient conditions for the stability of a finite number of queues were provided; however, the stable throughput region was only explicitly characterized for a 3-terminals system. In [49], the authors provided tighter bounds on the stable throughput region for the ALOHA system using the concept of stability ranks, which was also introduced in the same paper. The stability of ALOHA systems under a multi-packet reception model (MPR) was considered in [50] and [51]. Characterizing the stable throughput region for interacting queues with M > 3 terminals is still an open problem.

The stability region is characterized for the two above mentioned scenarios and compared to the case where the primary network doesn't employ relays. Analytical and numerical results reveal that although relays occupy part of the empty time slots that would have been available to secondary nodes, it is always beneficial to both primary and secondary nodes that the maximum possible number of relays be employed. On one hand, relays help the primary network achieve higher stable throughput by offering different reliable paths for the packets to reach the destination. On the other hand, relays will help primary nodes empty their queues at a much faster rates, thus providing secondary nodes with more opportunities to transmit their own information. It is interesting to note that even when secondary nodes interfere with relays transmissions, there is a significant improvement in both primary and secondary throughput due to this fast rate of emptying the queues.

3.1 System Models

We consider the uplink of a TDMA cellular network as the primary network. The primary network consists of M_p source nodes numbered $1, 2, ..., M_p$ communicating with a base station (BS) d_p located at the center of the cell as illustrated in Fig. 4.2. As part of the primary network, M_r cognitive relay nodes numbered $1, 2, ..., M_r$ are deployed to help primary nodes forward their packets to the base station. The relay nodes will exploit the under-utilized channel resources (time slots in this case) to forward primary packets without incurring any loss in the bandwidth

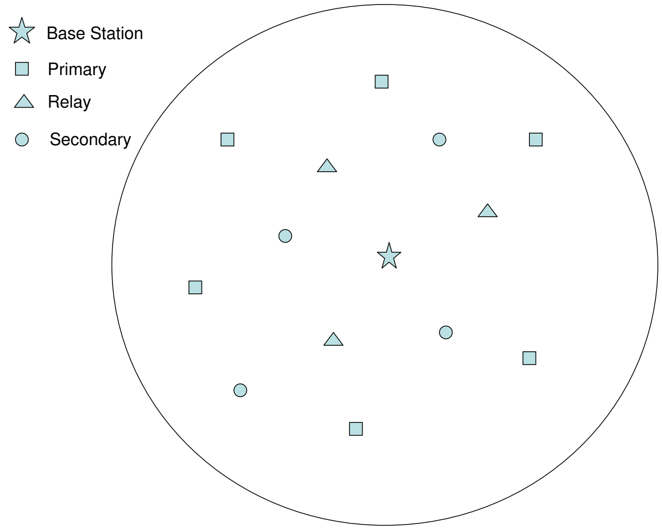

Figure 3.1: Network model

efficiency. A secondary network, consisting of M_s nodes numbered $1, 2, ..., M_s$, tries to exploit the unutilized channel resources to communicate their own data packets using slotted ALOHA as their multiple access protocol. We consider a circular cell of radius R. The BS is located at the center of the cell, and the different nodes are uniformly distributed within the cell area. Let $\mathcal{M}_p = 1, 2, ..., M_p$ denote the set of primary nodes, $\mathcal{M}_r = 1, 2, ..., M_r$ denote the set of relay nodes, and $\mathcal{M}_s = 1, 2, ..., M_s$ denote the set of secondary nodes. .

3.1.1 Channel Model

The wireless channel between a node and its destination is modeled as a Rayleigh flat fading channel with additive white Gaussian noise. The signal received at a receiving node j from a transmitting node i at time t can be modeled as

$$y_{ij}^t = \sqrt{G_i \rho_{ij}^{-\gamma}} h_{ij}^t x_i^t + n_j^t, \tag{3.1}$$

where G_i is the transmitting power, assumed to be the same for all nodes, ρ_{ij} denotes the distance between the two nodes, γ the path loss exponent, h_{ij}^t is the channel fading coefficient between nodes i and j at time t and is modeled as i.i.d zero mean, circularly symmetric complex gaussian random process with unit variance. The term x_i^t denotes the transmitted packet with average unit power, and n_j^t denote i.i.d additive white Gaussian noise processes with zero mean and variance N_0. Since the arrivals, the channel gains, and the additive noise processes are all assumed stationary, we can drop the index t without loss of generality.

As in the previous chapter, success and failure of packet reception are characterized by outage events and outage probabilities,

$$O_{ij} \triangleq \left\{ h_{ij} : \frac{|h_{ij}|^2 \rho_{ij}^{-\gamma} G_i}{N_0} \leq \beta \right\}, \tag{3.2}$$

and

$$\Pr\{O_{ij}\} = P_{ij}^o = 1 - \exp\left(-\frac{\beta N_0 \rho_{ij}^{\gamma}}{G_i} \right). \tag{3.3}$$

3.1.2 Queuing Model

Each primary, relay, or secondary node has an infinite buffer for storing fixed length packets. The channel is slotted in time and a slot duration equals the packet transmission time. The arrivals at the i^{th} primary node's queue ($i \in \mathcal{M}_p$), and the j^{th} secondary node's queue ($i \in \mathcal{M}_s$) are Bernoulli random variables, i.i.d from slot to slot with mean λ_i^p and λ_j^s, respectively. Hence, the vector $\Lambda = [\lambda_1^p, ..., \lambda_{M_p}^p, \lambda_1^s, ..., \lambda_{M_s}^s]$ denotes the average arrival rates. Arrival processes are assumed to be independent from one node to another.

Primary users access the channel by dividing the channel resources, time in this case, among them, hence, each node is allocated a fraction of the time. Let

$\mathbf{\Omega_p} = [\omega_1^p, \omega_2^p, ..., \omega_{M_p}^p]$ denote a resource-sharing vector, where $\omega_i^p \geq 0$ is the fraction of time allocated to node $i \in \mathcal{M}_p$, or it can represent the probability that node i is allocated the whole time slot [52]. The set of all feasible resource-sharing vectors is specified as follows

$$\digamma_p = \left\{ \mathbf{\Omega_p} = (\omega_1^p, \omega_2^p, ..., \omega_{M_p}^p) \in \Re^{+M_p} : \sum_{i \in \mathcal{M}_p} \omega_i^p \leq 1 \right\}. \tag{3.4}$$

In a communication network, the stability of the network's queues is a fundamental performance measure. Stability can be loosely defined as having a certain quantity of interest kept bounded. In our case, we are interested in the queue size being bounded. More rigourously, stability can be defined as follows (for the primary network alone). Denote the queue sizes of the transmitting nodes at any time t by the vector $\mathbf{Q}^t = [Q_i^t, i \in \mathcal{M}_p]$. We adopt the following definition of stability used in [48]. Queue $i \in \mathcal{M}_p$ of the system is stable if,

$$\lim_{t \to \infty} \Pr \left\{ Q_i^t < x \right\} = F(x) \text{ and } \lim_{x \to \infty} F(x) = 1. \tag{3.5}$$

If $\lim_{x \to \infty} \lim_{t \to \infty} \inf \Pr \left\{ Q_i^t < x \right\} = 1$, the queue is called substable. From the definition, if a queue is stable then it is also substable. If a queue is not substable, then it is unstable. An arrival rate vector $\Lambda = [\lambda_1^p, ..., \lambda_{M_p}^p]$ is said to be stable if there exists a resource sharing vector $\Omega_p \in \digamma_p$ such that all the queues are stable. The multidimensional stochastic process $\mathbf{Q}^t$ can be easily shown to be an irreducible and aperiodic discrete-time Markov chain process with a countable number of states and state space $\in \mathrm{Z}_+^{M_p}$. For such a Markov chain, the process is stable if and only if there exists a positive probability for every queue being empty [49], i.e.,

$$\lim_{t \to \infty} \Pr \left\{ Q_i^t = 0 \right\} > 0, \quad i \in \mathcal{M}_p. \tag{3.6}$$

If the arrival and service processes of a queueing system are strictly stationary, then one can apply Loynes's theorem to check for stability conditions [53]. This theorem states that if the arrival process and the service process of a queueing system are strictly stationary, and the average arrival rate is less than the average service rate, then the queue is stable; if the average arrival rate is greater than the average service rate then the queue is unstable.

3.2 Cognitive Cooperative Protocol with Multiple Relays

In a TDMA system without relays, if a node does not have a packet to transmit, its time slot remains idle, i.e., wasted channel resources. The possibility to utilize these wasted channel resources to provide some sort of spatial diversity and increased reliability to the TDMA system by employing a single cooperative relay node was investigated in [19]. Here we consider the case of a network with multiple relay nodes. We assume that relays can sense the communication channel to detect empty time slots. This assumption is reasonable for the orthogonal multiple-access scheme used, as there is no interference, and the relay can employ coherent or feature detectors that have high detection probability [54]. In the presence of interference, knowledge of the interference structure can help in the detection. The second assumption we make is that the errors and delay in packet acknowledgement feedback is negligible, which is reasonable for short length ACK/NACK packets as low rate codes can be employed in the feedback channel.

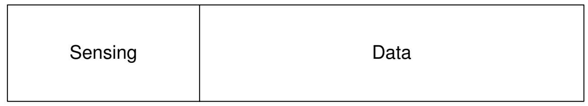

Figure 3.2: Time slot structure, showing the sensing period used by the relays to detect primary presence.

3.2.1 Cooperation Protocol

In this section, the cooperation protocol is presented. For the purpose of protocol description and analysis we will assume that the relay selection phase has already taken place, and that every primary node has assigned to it the best relay from the group of available relays. Note that every primary node gets help from only one relay, but a relay might help more than one primary node. Later in this section we will propose two different relay selection criteria and compare between them. The cooperative protocol operates as follows.

- At the beginning of each time slot, a node transmits the packet at the head of its queue (if any) to the destination. Due to the broadcast nature of wireless channel, relays can listen to the transmitted packets.
- If the packet is not received correctly by the destination, a NACK message is fed back from the destination declaring the packet's failure. If the relay assigned to the packet owner was able to decode the packet correctly, it stores the packet in its queue and sends back an ACK message to declare successful reception of the packet at the relay. We assume that the ACK/NACK feedback is immediate and error free.
- The node drops the packet from its queue if it is correctly received by either the destination or the relay (an ACK is received from either the destination or the relay).

- At the beginning of each time slot, relays sense the channel to check whether the time slot is empty (not utilized for packet transmission) or not (see Fig. 4.1).

- Relays distribute the available time slots in a TDMA fashion. Therefore, if a time slot is detected as empty, this free time slot will be assigned to relay $i \in \mathcal{M}_r$ with probability ω_i^r. As it is the case with TDMA networks, $\mathbf{\Omega_r} = (\omega_1^r, \omega_2^r, ..., \omega_{M_r}^r)$ denote a resource-sharing vector, and the set of all feasible resource-sharing vectors is specified as follows

$$\mathcal{F}_r = \left\{ \mathbf{\Omega_r} = (\omega_1^r, \omega_2^r, ..., \omega_{M_r}^r) \in \Re^{+M_r} : \sum_{i \in \mathcal{M}_r} \omega_i^r \leq 1 \right\}. \tag{3.7}$$

- Relay i then transmits the packet at the head of its queue.

The proposed protocol is cognitive in the sense that it introduces a relay in the network that tries detecting unutilized channel resources and use them to help other nodes by forwarding packets lost in previous transmissions.

In this chapter, we make the assumption assume that there is enough guard time at the beginning of each time slot that enables sensing, and that channel sensing is error free. In the next chapter we will consider the case with imperfect channel sensing, and the effect of sensing errors on the network's performance.

3.2.2 Stability Analysis

In this section we characterize the maximum stable throughput region of the cooperative protocol and compare it against the maximum stable throughput of TDMA without cooperation.

For the whole system to be stable, all queues therein should be stable. Hence, the stability region of the network is the intersection of the stability regions of the source nodes' queues, and the relay nodes' queues.

Source Nodes Stability

First, we consider the stability region of the system defined by the source nodes' queues. A source node succeeds in transmitting a packet if either the destination or its assigned relay receive the packet successfully. Therefore, the success probability of node i can be calculated as

$$P_i = \Pr\left\{\overline{O_{id}}\bigcap\overline{O_{ir_i}}\right\} = (1-P_{id}^o) + (1-P_{ir_i}^o) - (1-P_{id}^o)(1-P_{ir_i}^o), \tag{3.8}$$

where $\overline{O_{ij}}$ denotes complement of the event that the channel between node i and receiver $j \in (r_i, d)$ (r_i denotes the relay node assigned to node i, and d the destination) is in outage (i.e., the event that the packet was received successfully). If source node i has no relay assigned to it, its success probability is then given by

$$P_i = \Pr\left\{\overline{O_{id}}\right\} = 1 - P_{id}^o. \tag{3.9}$$

Therefore, for each queue $i \in \mathcal{M}_p$, the queue behaves exactly as in a TDMA system with success probability determined by (3.8) or (3.9). From Loynes's theorem, the primary nodes' stability region $\mathcal{R}_p$ is defined as

$$\mathcal{R}_p = \{(\lambda_1^p, ..., \lambda_{M_p}^p) \in R^{+M_p} : \lambda_i^p < \omega_i^p P_i, \forall i \in \mathcal{M}_p, (\omega_1^p, ..., \omega_{M_p}^p) \in \digamma_p\}, \tag{3.10}$$

which can be easily shown to be equivalent to

$$\mathcal{R}_p = \{(\lambda_1^p, ..., \lambda_{M_p}^p) \in R^{+M_p} : \sum_{i\in\mathcal{M}_p} \frac{\lambda_i^p}{P_i} \leq 1\}. \tag{3.11}$$

Relay Nodes Stability

Next we consider the stability of the relays' queues. In order to apply Loynes' theorem, it is required that the arrival and service processes of the relays' queues are stationary. Let Q_j^t denote the j^{th} ($j \in \mathcal{M}_r$) relay queue size at time t. Then its evolution can be modeled as

$$Q_j^{t+1} = (Q_j^t - Y_j^t)^+ + X_j^t, \tag{3.12}$$

where X_i^t represents the number of arrivals in time slot t and Y_i^t denotes the possibility of serving a packet at this time slot from the i^{th} relay queue (Y_i^t takes values $\{0, 1\}$). Function $(\cdot)^+$ is defined as $(x)^+ = \max(x, 0)$. Now we establish the stationarity of the arrival and service processes. If source nodes' queues are stable, then by definition the departure processes from these nodes are stationary. A packet departing from a node's queue is stored in the relay's queue (i.e., counted as an arrival) if simultaneously the following two events happen: the node-destination channel is in outage and the node-relay channel is not in outage. Hence, the arrival process to the queue can be modeled as follows

$$X_j^t = \sum_{i \in S_j} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t \neq 0\} \bigcap O_{id} \bigcap \overline{O_{ij}} \right], \tag{3.13}$$

where $\mathbf{1}[\cdot]$ is the indicator function and A_i^t denotes the event that slot t is assigned to source node i. $\{Q_i^t \neq 0\}$ denotes the event that node i queue is not empty, i.e., the node has a packet to transmit, and according to Little's theorem [55] it has probability $\lambda_i^p / (\omega_i^p P_i)$, where P_i is node i success probability and is defined in (3.8) and (3.9). Finally, S_j denotes the set of source nodes to which relay j assigned to help. The random processes involved in the above expression are all stationary, hence, the arrival process to the relay is stationary. The average arrival rate to the relay's queue can then be computed as

$$\lambda_j^r = E[X_j^t] = \sum_{i \in S_j} \lambda_i^p \frac{P_{id}^o(1 - P_{ij}^o)}{P_i}. \quad (3.14)$$

Similarly, we establish the stationarity of the service process of the j^{th} relay queue. The service process of the relay queue depends by definition on the empty slots available from primary nodes and the channel from relay to destination being not in outage. By assuming that source nodes' queues are stable, they offer stationary empty slots to the relay. Also the channel statistics is stationary, hence, the relay's service process is stationary. The service process of the j^{th} relay's queue can be modeled as

$$Y_j^t = \sum_{i \in \mathcal{M}_p} \mathbf{1}\left[A_i^t \bigcap \{Q_i^t = 0\} \bigcap \overline{O_{jd}^t} \bigcap U_j^t\right], \quad (3.15)$$

where U_j^t is the event that the current idle time slot is assigned to relay j to service its queue, which has probability ω_j^r according to the TDMA resource sharing policy employed by the relays. The average service rate of the relay can then be determined from the following equation

$$\mu_j^r = E[Y_j^t] = \left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right)(1 - P_{jd}^o)\omega_j^r. \quad (3.16)$$

Using Loynes' theorem, the stability condition for the j^{th} relay queue is $\lambda_j^r < \mu_j^r$. The stability region $\mathcal{R}_r$ of the system comprised of the relays' queues is then defined as

$$\mathcal{R}_r = \{(\lambda_1^p, ..., \lambda_{M_p}^p) \in R^{+M_p} : \lambda_j^r < \mu_j^r, \forall j \in \mathcal{M}_r, (\omega_1^r, ..., \omega_{M_r}^r) \in \digamma_r\}, \quad (3.17)$$

which can be easily shown to be equivalent to

$$\mathcal{R}_r = \{(\lambda_1^p, ..., \lambda_{M_p}^p) \in R^{+M_p} : \sum_{j \in \mathcal{M}_r} \frac{\sum_{i \in S_j} \lambda_i^p \frac{P_{id}^o(1-P_{ij}^o)}{P_i}}{\left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right)(1 - P_{jd}^o)} \leq 1\}. \quad (3.18)$$

Finally, the maximum stable throughput region of the complete system defined by the source nodes and relays queues is given by the intersection of the maximum stable throughput regions of source nodes queues and relays queues, which can be shown to be equal to

$$\mathcal{R} = \mathcal{R}_p \bigcap \mathcal{R}_r = \mathcal{R}_r. \tag{3.19}$$

From 3.18 it is noted that the stability region for the cooperative protocol is bounded by a hyperplane. Since the stability of TDMA is also determined by a hyperplane, when comparing both stability regions it is enough to compare the intersection of these hyperplanes with the coordinate axes . Considering the i^{th} source node, this intersection for the cooperative protocol is equal to

$$\lambda_i^{p*}(\text{Coop}) = \frac{P_i(1 - P_{jd}^o)}{P_{id}^o(1 - P_{ij}^o) + (1 - P_{jd}^o)}, \tag{3.20}$$

where it was assumed that relay node j is assigned to source node i. The corresponding value for TDMA is given by

$$\lambda_i^{p*}(\text{TDMA}) = 1 - P_{id}^o. \tag{3.21}$$

It is clear that the stability region for TDMA is completely contained inside the stability region of the cooperative protocol if $\lambda_i^{p*}(\text{Coop}) > \lambda_i^{p*}(\text{TDMA})$ for all $i \in \mathcal{M}_p$. Using (3.20) and (3.21), this condition is equivalent to

$$P_{jd}^o < P_{id}^o. \tag{3.22}$$

These conditions have the following intuitive explanation. If the channel between the relay and destination has higher success probability that the channel between the terminal and destination, then it is better to have the relay help the terminal transmit its packets. Note that (3.22) implies that TDMA can offer better performance for the terminal whose outage probability does not satisfy (3.22).

3.2.3 Relay Selection

In this section we propose two different relay assignment criteria and compare their performance in terms of maximum stable throughput.

Nearest Neighbor

It is noted from (3.8) that the probability of a successful source node transmission (correctly received by either the destination or the relay) is an increasing function of the success probability of the source-relay link. This probability is in turn a decreasing function of the distance between source node and relay node as seen from (3.3). Therefore, in order to maximize source nodes service rates, one can assign relays to source nodes based on the nearest neighbor criterion. That is, each source nodes gets help from its closest relay.

Although this criterion for relay selection maximizes source nodes service rates, it suffers from some performance degradation if the success probability of the source-destination link is better than that of the relay-destination link, as discussed above.

Based on this observation, the relay selection criterion is modified such that a source node selects its nearest neighbor relay from the group of relays that are closer to the destination than the source node itself. For the implementation of such a selection criterion, we assume that each user can know its distance to the destination through, for example, calculating the average received power. Then through a simple distributed protocol each node sends out a Hello message searching for its nearest neighbors. This can be done using time of arrival (TOA) estimation for example; see [56] and [57]. Each source node then selects the nearest neighbor relay node with a distance closer to the destination than the source node itself.

Maximum Success Probability

The nearest neighbor criterion considers only the maximization of the source node service rate, and makes sure that there will be no performance degradation due to the selection of an ill positioned relay.

In order to maximize the network's stability region, the relay selection process should be able to take the service rates of the relays into consideration. Intuitively, it is beneficial (from a stability point of view) to favor the relays with higher service rates over the ones with lower service rates. To take the relay-destination link into consideration, we propose the following criterion where source node i selects a relay according to

$$\begin{aligned} &\arg\max_{j\in\mathcal{M}_r}(1-P^o_{ij})(1-P^o_{jd}) \\ &\text{s.t. } P^o_{jd} < P^o_{ij}, \end{aligned} \tag{3.23}$$

i.e., node i selects the relay that maximizes the overall packet success probability over both source-relay and relay-destination links, under the constraint that the relay-destination link has a higher success probability that the source-destination link. Using the definition of the outage probability (4.3), it can be shown that the relay selection criterion of (3.23) is equivalent to

$$\begin{aligned} &\arg\min_{j\in\mathcal{M}_r}\rho_{ij}+\rho_{jd} \\ &\text{s.t. } \rho_{jd} < \rho_{ij}, \end{aligned} \tag{3.24}$$

where ρ_{ij} is the distance between source node i and relay node j, and ρ_{id} the distance between relay node j and the destination. Therefore, the maximum success probability criterion reduces to a minimization of the sum of source-relay relay-destination distances. It is noted that for a given source node, the optimal relay location is at the midpoint of the line between the source node and the destination. This selection criterion can be implemented using the same distributed protocol described above.

3.3 Opportunistic Multiple Access for Secondary Nodes

In the previous section, the problem of utilizing the idle channel resources to enable cognitive relays to help source nodes forward their packets was considered. Aside from being used by relays, these idle channel resources could be used by a group of secondary (unlicensed) nodes to transmit their own data packets. Therefore, the use of these idle channel resources (time slots, in our network) offers either diversity to the primary nodes through the group of relays, or multiplexing through the group of secondary nodes that send new information over the channel.

In this section, we study the effect of sharing the idle time slots between relays and secondary nodes on the performance of both primary and secondary networks. Mainly, we focus on how the secondary network's throughput is affected when part of the idle channel resources are used by the relays, and how the primary network throughput is affected when secondary transmissions interfere with relay transmissions. Furthermore, we study the possibility that secondary nodes work as relays for the primary network. By working as relays, the secondary nodes aim at creating more transmission opportunities for themselves by helping primary nodes empty their queues at a faster rate.

The secondary network consists of M_s nodes forming an ad-hoc network, in which nodes are grouped into source-destination pairs where each source node communicates with its associated destination node. To access the channel, secondary nodes will sense the channel at the beginning of each time slot, as shown in Fig. 4.1, to detect primary activity. As with the relays, we assume that the primary detection process is error free. To share the idle time slots among the secondary network, secondary nodes employ slotted ALOHA as a multiple access protocol. Therefore, whenever an idle slot is detected, secondary nodes with nonempty queues will attempt to transmit their packets with channel access probability α_s.

Since both relay and secondary nodes sense the channel at the beginning of each time slot, it is not necessary that secondary nodes will be able to detect relay transmissions. In such situations, secondary packets will collide with relay packets. To take these collision events into consideration, we will study two extreme cases. The first is when the secondary nodes are always unable to detect relays transmissions, thus always, colliding with relays if they decide to transmit at the same time slot. The second case occurs when secondary nodes are all the time able to detect relays presence successfully, thus no interference at all. The study of these two cases enables us to find inner and outer bounds on the maximum stable throughput region of the network.

Furthermore, we consider the tradeoff between the amount of help offered to the primary network through relays, and the achievable throughput of the secondary network. To study this tradeoff, we consider the case where relays limit their access to the channel, therefore, providing secondary nodes with uncontested access to the idle time slots. This is made possible by letting relays make their transmission attempts in an empty time slot with probability an access α_r. In other words, when a relay has a packet to transit, and it encounters an idle time slot, it will transmit its packet with probability α_r, and defer transmitting, in order to offer allow secondary nodes to use that slot, with probability $1 - \alpha_r$.

3.3.1 Case I: No Interference

Here we consider the case when secondary nodes are always able to successfully detect relays transmissions. Therefore, no interference is exhibited by relay nodes from secondary transmissions.

In order to share resources with secondary nodes, and enable secondary nodes to access the idle time slots, relays will limit their access to the channel by utilizing a transmission probability α_r. In other words, when a relay detects an empty time slot, it transmits the packet at the head of its queue with probability α_r, and remains silent with probability $1-\alpha_r$. In this case, TDMA is still used to organize relays access to idle time slots, and we assume that all relays will use the same probability α_r. Therefore, relays will collectively use a fraction α_r of the idle time slots to offer help to primary nodes, and secondary nodes will have a guaranteed access to at least a fraction $1 - \alpha_r$ of the idle time slots. The actual figure will be higher since relays will not have packets to transmit all the time. It should also be noted that, since all relays have the same access probability, this scheme will not affect the relay selection process. In other words, if relay j is the optimal relay for node i, then either it will remain its optimal relay, or under some conditions, it will be better that node i does not use any relay.

Primary Network Stability Analysis

Since service processes of the primary nodes are not affected by how relays access the channel, the stability region of the system comprised of primary queues is defined as in (3.11).

To characterize the stability region of the system composed of relays queues, we first note that the arrival process to a relay queue is not affected by the relay's channel access mechanism. Therefore, the arrival process for relay queue $j \in \mathcal{M}_r$ is defined as in (3.13), and its average arrival rate given by (3.13). For the j^{th} relay service process, stationarity of the service process could easily be established using the same arguments used in the previous section. Therefore, the service process of the j^{th} relay's queue can be modeled as

$$Y_j^t = \sum_{i \in \mathcal{M}_p} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t = 0\} \bigcap \overline{O_{jd}^t} \bigcap U_j^t \bigcap P_r \right], \tag{3.25}$$

where U_j^t is the event that the current idle time slot is assigned to relay j to service its queue, which has probability ω_j^r according to the TDMA resource sharing policy employed by the relays, and P_r is the event that relay j has permission to access the channel in the current time slot, which has a probability α_r. The average service rate of the relay can then be determined from the following equation

$$\mu_j^r = E[Y_j^t] = \left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i} \right) (1 - P_{jd}^o) \omega_j^r \alpha_r. \tag{3.26}$$

Using Loynes' theorem, the stability condition for the j^{th} relay queue is $\lambda_j^r < \mu_j^r$, and the stability region $\mathcal{R}_r$ of the system comprised of the relays' queues can be shown to be defined as follows,

$$\mathcal{R}_r = \{(\lambda_1^p, ..., \lambda_{M_p}^p) \in R^{+M_p} : \sum_{j \in \mathcal{M}_r} \frac{\sum_{i \in S_j} \lambda_i^p \frac{P_{id}^o (1 - P_{ij}^o)}{P_i}}{\left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right) (1 - P_{jd}^o)} \alpha_r \leq 1 \}, \tag{3.27}$$

which can also be shown to be equal to the stability region of the whole primary network (primary nodes and relays).

From (3.27) it is noted that the stability region in this case is also bounded by a hyperplane. The intersection with the i^{th} coordinate axis gives the maximum allowable arrival rate for the i^{th} source node, which is equal to

$$\lambda_i^{p*}(\text{Coop}) = \frac{P_i(1-P_{jd}^o)\alpha_r}{P_{id}^o(1-P_{ij}^o)+(1-P_{jd}^o)\alpha_r} \tag{3.28}$$

which is a monotonically increasing function in α_r. Therefore, from the point of view of primary network stability, it is always beneficial to assign most of the idle resources to relays.

The condition in (3.22), defining when the cooperative protocol outperforms TDMA, translates into

$$\left(1-P_{jd}^o\right)\alpha_r > \left(1-P_{id}^o\right), \tag{3.29}$$

which tells us that, by limiting their access to the channel, relays appear to primary nodes as having higher outage probabilities. Therefore, according to the value of α_r and different relays' outage probabilities, some of the relays might be rendered unusable, and situations might arise in which no relay is used at all.

Secondary Network Stability Analysis

Switching to the analysis of the secondary nodes stability, we recall that the secondary network consists of M_s nodes numbered $1, 2, ..., M_s$, and having average arrival rates $[\lambda_1^s, ..., \lambda_{M_s}^s]$. Upon the detection of an idle time slot, a node with non-empty queues will try to transmit the packet at the head of its queue with access probability α_s. A node's sensing and channel access decisions are independent from other nodes. We further assume for mathematical tractability that all nodes have the same access probability α_s.

To study the stability region of secondary nodes, we note first that, since secondary nodes are employing slotted ALOHA for multiple access, the secondary nodes queues are interacting. In other words, the service rate of a given queue is dependent on the state of all other queues, i.e., whether they are empty or not. Studying the stability conditions for interacting queues is a difficult problem that has been addressed for ALOHA systems [47], [49] [50]. The concept of dominant systems was introduced and employed in [47] to help findi bounds on the stability region of ALOHA with collision channel. The dominant system in [47] was defined by allowing a set of terminals with no packets to transmit to continue transmitting dummy packets. In this manner, the queues in the dominant system stochastically dominate the queues in the original system. Or in other words, with the same initial conditions for queue sizes in both the original and dominant systems, the queue sizes in the dominant system are not smaller than those in the original system.

To study the stability of the interacting system of queues consisting of secondary nodes queues, we make use of the dominant system approach to decouple the interaction between queues. We define the dominant system as follows

- Arrivals at each queue in the dominant system are the same as in the original system.
- Time slots assigned to primary node $i \in \mathcal{M}_p$ are identical in both systems.
- The outcomes of the "coin tossing" (that determines transmission attempts of relay and secondary nodes) in every slot are the same.
- Channel realizations for both systems are identical.
- The noise generated at the receiving ends of both systems is identical.
- In the dominant system, secondary nodes attempt to transmit dummy packets when their queues are empty.

Given identical initial queue sizes for both the original and dominant systems, secondary nodes queues in the dominant system are never shorter than those in the original one. This is true because in the dominant system, secondary nodes suffer from an increased collision probability, thus longer queues, compared to the original one since secondary nodes always have a packet to transmit (possibly a dummy packet). This implies that relay nodes' queues empty faster in the original system and therefore relays see a lower probability of collision as compared to the dominant system, and as a result will have shorter queues. Consequently, stability conditions for the dominant system are sufficient for the stability of the original system.

To prove the necessary conditions, we follow an argument similar to that used by [47] and [50] for ALOHA systems to prove the "indistinguishability" of the dominant and original systems at saturation. Consider the dominant system in which secondary nodes transmit dummy packets. If along some realizations of secondary queues of nonzero probability, secondary queues never empty, then the original system and the dominant system are "indistinguishable". Thus, with a particular initial condition, if secondary queues in the dominant system never empty with nonzero probability (i.e., it is unstable), then secondary queues in the original system must be unstable as well. This means that the boundary of the stability region of the dominant system is also a boundary for the stability region of the original system. Thus, conditions for stability of the dominant system are sufficient and necessary for the stability of the original system.

The service process of a secondary node depends on the idle time slots unused by the primary and relay nodes. Therefore, the service process of the k^{th} secondary node can be modeled as

$$Y_k^t = \sum_{i\in\mathcal{M}_p}\sum_{j\in\mathcal{M}_r} \mathbf{1}\Bigg[A_i^t \bigcap \{Q_i^t = 0\} \bigcap U_j^t \bigcap \left\{\overline{\{Q_j^t \neq 0\} \bigcap P_r}\right\} \bigcap \overline{O_{kd}^t} \bigcap P_s \bigcap_{l\in\mathcal{M}_s\backslash k} \{\overline{P_s}\}\Bigg], \tag{3.30}$$

which is the event that the primary node for which the current time slot is assigned has an empty queue, and the relay for which the current time slot has either an empty queue or does not have permission to transmit (the event $U_j^t \bigcap \left\{\overline{\{Q_j^t \neq 0\} \bigcap P_r}\right\}$). Event P_s is the event that a secondary node has a permission to transmit, which has a probability α_s. Therefore, the event $\bigcap P_s \bigcap_{l\in\mathcal{M}_s\backslash k} \{\overline{P_s}\}$ is that only one secondary node is transmitting in the current time slot; otherwise a collision will occur and all packets involved will be lost. Finally, $\overline{O_{kd}^t}$ denotes the event that the k^{th} secondary node link to its destination is not in outage.

Assuming that primary and relay nodes' queues are stable, then they offer stationary empty slots. Also the channel statistics are stationary; hence, the secondary service process is stationary. The average secondary service rate is then given by

$$\mu_k^s = E[Y_k^t] = \left(1 - \sum_{i\in\mathcal{M}_p}\frac{\lambda_i^p}{P_i} - \sum_{j\in\mathcal{M}_r}\frac{\sum_{i\in S_j}\lambda_i^p \frac{P_{id}^o(1-P_{ij}^o)}{P_i}}{(1-P_{jd}^o)}\right)\alpha_s(1-\alpha_s)^{M_s-1}(1-P_{kd}^o). \tag{3.31}$$

From (4.15), it can be easily shown that the optimum value for the secondary access probability is

$$\alpha_s = \frac{1}{M_s}.$$

Using Loyne's theorem along with (4.15), and from (3.27), the stability region of the system defined by the primary nodes, relay nodes, and secondary nodes can be written as

$$\mathcal{R} = \mathcal{R}_r \bigcap \mathcal{R}_r = \Bigg\{ (\lambda_1^p, ..., \lambda_{M_p}^p, \lambda_1^s, ..., \lambda_{M_s}^s) \in R^{+(M_p+M_s)} :$$

$$\sum_{j \in \mathcal{M}_r} \frac{\sum_{i \in S_j} \lambda_i^p \frac{P_{id}^o(1-P_{ij}^o)}{P_i}}{\left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right)(1 - P_{jd}^o)} \alpha_r \leq 1,$$

$$\lambda_k^s \leq \left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i} - \sum_{j \in \mathcal{M}_r} \frac{\sum_{i \in S_j} \lambda_i^p \frac{P_{id}^o(1-P_{ij}^o)}{P_i}}{(1 - P_{jd}^o)}\right)$$

$$\times \frac{1}{M_s}(1 - \frac{1}{M_s})^{M_s - 1}(1 - P_{kd}^o), k \in \mathcal{M}_s \Bigg\}. \tag{3.32}$$

Dependence of secondary nodes service rates in (4.15) on the parameter α_r appears only through primary nodes success probabilities P_i defined in (3.8) and (3.9), and relay nodes arrival rates, which are dependent on α_r through the relay assignment process. Clearly, a higher α_r will result in primary nodes getting better service from relays; thus, primary queues will have higher services rates. Therefore, there will be a higher probability that primary queues are empty. Since relay service rates explicitly depend on α_r as shown in (3.26), higher α_r will also mean a higher probability of empty relay queues. This results in more idle time slots for secondary nodes to exploit.

Case II: Maximum Interference

In the last section, the ideal case in which secondary nodes can sense relays' presence was considered. Here we consider the worst case scenario where secondary nodes cannot sense relay transmissions at all. In this case, collisions between relays and secondary transmissions are inevitable. In case of a collision all packets involved are lost, and a retransmission is necessary.

Again, to study the tradeoff between assigning idle resources to relays or secondary nodes, the case where relays limit their access to the idle time slot and have access probability α_r is considered. Since the cognitive principle is based on the idea that the presence of the secondary system should be transparent to the primary system, the secondary nodes access probability α_s will now play a crucial role of limiting secondary interference to the primary network.

Because of the possible collisions between secondary and relay transmissions, relay and secondary nodes queues form a system of interacting queues.

To study the stability of the interacting system of queues consisting of the relay and secondary nodes queues, we make use of the dominant system approach to decouple the interaction between the queues. The dominant system here will be different from the one we used in the previous section because now relays are involved in collisions. The dominant system in this case is defined as follows. For $j \in \{1, 2\}$, define D_j as

- Arrivals at each queue in D_j are the same as in the original system.
- Time slots assigned to primary node $i \in \mathcal{M}_p$ are identical in both D_j and the original systems.
- The outcomes of the "coin tossing" (that determines transmission attempts of the relay and secondary nodes) in every slot are the same.
- Channel realizations for both systems are identical.
- The noise generated at the receiving ends of both systems is identical.
- In D_1 relays will attempt to transmit dummy packets if their queues are empty. Since secondary queues are interacting among themselves (an interaction that needs to be decoupled as well), secondary nodes will attempt to

transmit dummy packets only if they are informed (with the aid of a "genie") that relays are not transmitting in the current time slot.

- In D_2, secondary nodes attempt to transmit dummy packets when their queues are empty, and relays operate normally.

Stability conditions for the above defined dominant system could be shown to be necessary and sufficient for the stability of the original system through similar arguments to the ones used in the previous section.

Dominant System D_1

Under this dominant system, relays will be transmitting dummy packets if their queues are empty. Since service processes of the primary nodes are not affected by how relays access the channel, the stability region of the system comprised of primary queues is defined as in (3.11).

We start by characterizing the stability region of the system defined by the relays' queues. As in the previous section, the average arrival rate to the relay is unchanged and is given by (3.14). Relays' service processes now depend on the state of secondary queues in addition to the empty slots available from primary nodes, and the channel from a relay to the destination not being in outage. The service process of the j^{th} relay queue can then be modeled as

$$Y_j^t = \sum_{i \in \mathcal{M}_p} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t = 0\} \bigcap U_j^t \bigcap P_r \bigcap \overline{O_{jd}^t} \bigcap_{k \in \mathcal{M}_s} \left\{ \overline{\{Q_k^t \neq 0\} \bigcap P_s} \right\} \right], \quad (3.33)$$

which accounts for the events that, the primary node owning the current time slot has an empty queue, the current time slot is assigned to relay j, the relay has permission to transmit, the relay-destination link is not in outage, and finally, no secondary node is transmitting, which is either due to empty queues or lack of permission to transmit. The average service rate of the j^{th} relay is then given by

$$\mu_j^r = E[Y_j^t] = \left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i}{P_i}\right) \omega_j^r \alpha_r (1 - P_{jd}^o) \prod_{k \in \mathcal{M}_s} \left(1 - \frac{\lambda_k^s}{\mu_k^s} \alpha_s\right). \tag{3.34}$$

Next, we consider the service processes for the secondary queues. Beside the idle time slots unused by the primary nodes and other secondary nodes queues, the service process of a secondary node now depends on whether or not relays have permission to transmit. Therefore, the service process of the k^{th} secondary node can be modeled as

$$Y_k^t = \sum_{i \in \mathcal{M}_p} \sum_{j \in \mathcal{M}_r} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t = 0\} \bigcap U_j^t \bigcap \overline{P_r} \bigcap \overline{O_{kd}^t} \bigcap P_s \bigcap_{l \in \mathcal{M}_s \backslash k} \{\overline{P_s}\} \right], \tag{3.35}$$

which is the event that the primary node for which the current time slot is assigned has an empty queue, the relay has no permission to transmit, the k^{th} secondary node has permission to transmit, all other secondary nodes do not have permission, and the secondary-destination link is not in outage. The average secondary service rate is then given by

$$\mu_k^s = E[Y_k^t] = \left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right) (1 - \alpha_r) \alpha_s (1 - \alpha_s)^{M_s - 1} (1 - P_{kd}^o). \tag{3.36}$$

Using Loynes' theorem and (3.14), (3.34), and (4.30), the stability region for the dominant system D_1 for a given α_r and α_s can be written as follows,

$$\mathcal{R}(D_1) = \Bigg\{ (\lambda_1^p, ..., \lambda_{M_p}^p, \lambda_1^s, ..., \lambda_{M_s}^s) \in R^{+(M_p + M_s)} :$$

$$\sum_{j \in \mathcal{M}_r} \frac{\sum_{i \in S_j} \lambda_i^p \frac{P_{id}^o (1 - P_{ij}^o)}{P_i}}{\left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right) (1 - P_{jd}^o) \alpha_r \prod_{k \in \mathcal{M}_s} \left(1 - \frac{\lambda_k^s}{\mu_k^s} \alpha_s\right)} \leq 1,$$

$$\lambda_k^s \leq \left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right) (1 - \alpha_r) \alpha_s (1 - \alpha_s)^{M_s - 1} (1 - P_{kd}^o), k \in \mathcal{M}_s \Bigg\} \tag{3.37}$$

Dominant System D_2

Under this dominant system, secondary nodes will be transmitting dummy packets if their queues are empty. As it is the case in previous sections, the stability region of the system comprised of primary queues is defined as in (3.11).

To characterize the stability region of the system defined by the relays queues, we note that as in the previous section, the average arrival rate to the relay is unchanged and is given by (3.14). The service process of the j^{th} relay queue is modeled as

$$Y_j^t = \sum_{i \in \mathcal{M}_p} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t = 0\} \bigcap U_j^t \bigcap P_r \bigcap \overline{O_{jd}^t} \bigcap_{k \in \mathcal{M}_s} \left\{ \overline{\bigcap P_s} \right\} \right], \tag{3.38}$$

which differs from (3.33) in the term accounting for the state of secondary queues. Here we have only the event that no secondary node has permission to transmit, because even if the queues are empty, secondary nodes continue to transmit dummy packets. The average service rate of the j^{th} relay is then given by

$$\mu_j^r = E[Y_j^t] = \left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i}{P_i} \right) \omega_j^r \alpha_r (1 - P_{jd}^o)(1 - \alpha_s)^{M_s}. \tag{3.39}$$

Next, we consider the service processes for the secondary queues. Here the secondary service process is dependent on the states of different relay queues. Therefore, the service process of the k^{th} secondary node can be modeled as

$$Y_k^t = \sum_{i \in \mathcal{M}_p} \sum_{j \in \mathcal{M}_r} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t = 0\} \bigcap U_j^t \bigcap \overline{\{Q_j^t \neq 0\} \bigcap P_r} \bigcap \overline{O_{kd}^t} \bigcap P_s \right.$$
$$\left. \bigcap_{l \in \mathcal{M}_s \backslash k} \{\overline{P_s}\} \right], \tag{3.40}$$

and the average secondary service rate is then given by

$$\mu_k^s = E[Y_k^t] = \left(1 - \sum_{i\in\mathcal{M}_p} \frac{\lambda_i^p}{P_i} - \sum_{j\in\mathcal{M}_r} \frac{\sum_{i\in S_j} \lambda_i^p \frac{P_{id}^o(1-P_{ij}^o)}{P_i}}{(1-P_{jd}^o)(1-\alpha_s)^{M_s}}\right)$$
$$\times \alpha_s(1-\alpha_s)^{M_s-1}(1-P_{kd}^o). \quad (3.41)$$

Using Loynes' theorem and (3.14), (3.39), and (4.32), the stability region for the dominant system D_2 for a given α_r and α_s can be written as follows,

$$\mathcal{R}(D_2) = \Bigg\{ (\lambda_1^p, ..., \lambda_{M_p}^p, \lambda_1^s, ..., \lambda_{M_s}^s) \in R^{+(M_p+M_s)} :$$
$$\sum_{j\in\mathcal{M}_r} \frac{\sum_{i\in S_j} \lambda_i^p \frac{P_{id}^o(1-P_{ij}^o)}{P_i}}{\left(1 - \sum_{i\in\mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right) \alpha_r (1-P_{jd}^o)(1-\alpha_s)^{M_s}} \leq 1,$$
$$\lambda_k^s \leq \left(1 - \sum_{i\in\mathcal{M}_p} \frac{\lambda_i^p}{P_i} - \sum_{j\in\mathcal{M}_r} \frac{\sum_{i\in S_j} \lambda_i^p \frac{P_{id}^o(1-P_{ij}^o)}{P_i}}{(1-P_{jd}^o)(1-\alpha_s)^{M_s}}\right)$$
$$\times \alpha_s(1-\alpha_s)^{M_s-1}(1-P_{kd}^o), k \in \mathcal{M}_s \Bigg\} \quad (3.42)$$

Finally, the whole stability region can be determined by taking the union over all possible values of α_s as follows,

$$\mathcal{R} = \bigcup_{\alpha_s\in[0,1]} \left\{\mathcal{R}(D_1) \bigcup \mathcal{R}(D_2)\right\}. \quad (3.43)$$

The dependence of the stability region of (3.43) on the resource assignment parameter α_r is very complex to characterize. Looking at the stability conditions of the system D_1, it can be noted from (3.36) that, on one hand, the dependence of the secondary queues service rates on the primary queues service rates makes it beneficial to assign more idle slots to relays. On the other hand, because of the possible collisions with relay transmissions, which is modeled with the term $(1-\alpha_r)$ in (3.36), it is better from a secondary network point of view to reduce the amount of resources assigned to relays. Similarly, from (3.34) it is noted that the relay's service rate has two competing components that depend on α_r. The first rises from primary queues service rates which are increasing in α_r. The second is the probability that secondary queues are empty, in which as discussed above, its dependence on α_r is not easily identified. If we then look at the dominant system D_2, it will be immediately clear from (3.39) and (4.32) that both primary and secondary nodes will benefit from assigning more resources to relays.

3.3.2 Results and Discussions

First we present results for the proposed selection schemes by considering the following scenario. $M_p = 20$ source nodes, and $M_r = 1, ..., 20$ relay nodes are deployed uniformly in a circular cell of radius $R = 200m$, with the BS located at the center of the cell. The propagation path loss is taken equal to $\gamma = 3.7$ and the SNR threshold $\beta = 35$dB. The transmitted signal power is $G = 100$mW, and the noise power is $N_0 = 10^{-11}$. For ease of illustration we consider the aggregate network arrival rate $\lambda_p = \sum_i \lambda_p^i$, $i \in \mathcal{M}_p$.

Fig. 3.3 compares the maximum stable throughput of the cooperative versus non-cooperative networks as a function of the number of relays in the network. Furthermore, it compares the performance of the two proposed relay selection schemes. It is clear that the cooperative protocol outperforms its non-cooperative counterpart; even with a single relay (which of course is not helping all the nodes) a 25% increase in throughput is achieved. As the number of relays increases we notice a fast increase in throughput; for example, with 5 relays the throughput is increased by 128%. Increasing the number of relays to 10 leads to a 167% increase in throughput. This is mainly because increasing the number of relays increases the number of source nodes that are getting help from these relays, hence leading

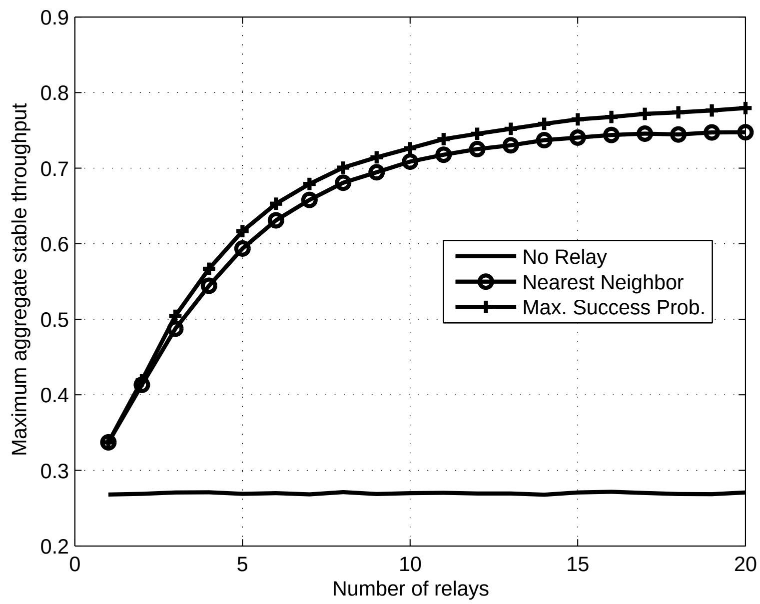

Figure 3.3: Maximum aggregate stable throughput vs. number of relays. $M_p = 20$ primary nodes.

to higher throughput.

From Fig. 3.3, it can be seen that the "maximum success probability" relay selection criterion outperforms the"nearest neighbor" criterion by a margin of 3% to 4% on average. Furthermore, it is noted that the gap between the two criteria increases with increasing number of relays. This is due to the fact that with an increased relay density in the network, there will be a higher probability that a source node finds a relay at or near the optimal relay position corresponding to that source node. While the "maximum success probability" criterion will be able to select the relay at the optimal (or near optimal) location, the "nearest neighbor criterion" will always pick the closest relay to the source node.

Next we consider the network stability region under the ideal case of no collisions between secondary and relay nodes. In order to be able to visualize the

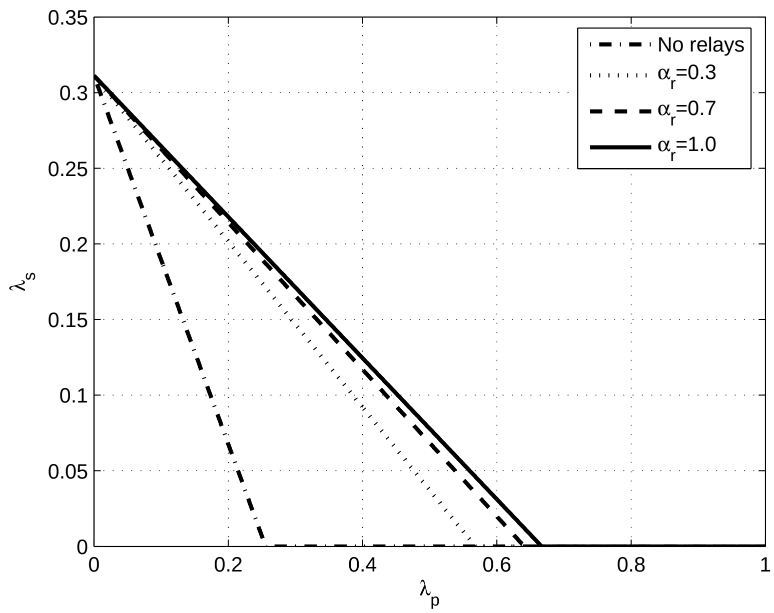

Figure 3.4: System stability region with and without cooperation for different values of α_r.

network's stability region, which has in general $M_p + M_s$ dimensions, we plot the maximum aggregate primary arrival rate $\lambda_p = \sum_i \lambda_i^p$, for $i \in \mathcal{M}_p$, against the maximum aggregate primary arrival rate $\lambda_s = \sum_i \lambda_i^s$, for $i \in \mathcal{M}_s$.

Fig. 3.4 depicts the stability region of the system composed of the primary, relays, and secondary nodes. The system has $M_p = 20$ primary nodes, $M_r = 10$ relay nodes, and $M_s = 10$ secondary nodes. The benefits of cooperation for both primary and secondary networks are significant as illustrated. For instance, at $\lambda_p = 0.2$ we observe a 350% increase in the secondary throughput. Moreover, it is noted that both networks benefit from increasing the fraction of idle time slots assigned to relays for cooperation. On one hand, the primary network benefits from that increase since it will get better service from relays, which in turn increases primary nodes service rates, thus, the network can have an extended stability region

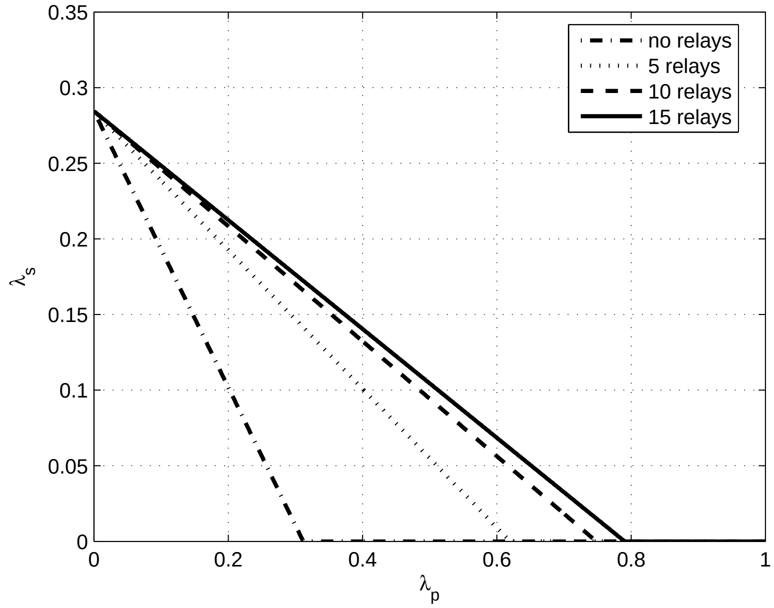

Figure 3.5: Stability region with different number of relays. $M_p = 20$ primary nodes and $M_s = 10$ secondary nodes.

by sustaining higher arrival rates. On the other hand, secondary nodes benefit from assigning a higher fraction of idle resources to relays, since this results in both primary nodes and relays having higher service rates, thus, higher probability of empty queues. With higher probability of empty queues, secondary nodes will have an increased number of idle time slots to transmit their packets. So in conclusion, under the current scenario of no interference between relays and secondary nodes, it is beneficial to both primary and secondary networks to assign all idle resource to cooperation.

Fig. 3.5 depicts the stability region of the system comprised of the primary, relays and secondary nodes queues for $\alpha_r = 1$. It is clear that increasing the number of relays in the network leads to significant improvement in the overall stability region and not only affects primary nodes stability; e.g., for $\lambda_p = 0.25$ we

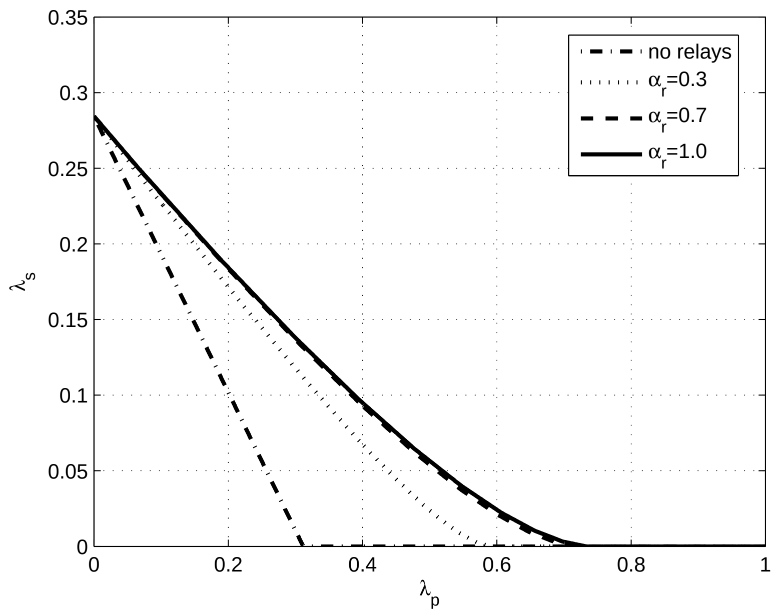

Figure 3.6: System stability region with and without cooperation in case of colliding relay and secondary transmissions for different values of α_r.

see around 300% increase in secondary throughput when using 10 or 15 relays. This is due to the fact that although it is apparent that increasing the number of relays will use more and more of the idle time slots and hence decrease secondary nodes' chance to access the channel, the relays help primary nodes empty their queues at a faster rate, and hence provide the secondary nodes with more opportunities to access the channel. We conclude that, with the secondary nodes able to detect both primary and relay transmissions, it is always better to have the maximum number possible of relays, and assign all free resources to cooperation since this will maximize both primary and secondary networks throughput.

Finally, we consider the worst case scenario in which secondary nodes cannot detect relay transmissions, hence always colliding with them. Fig. 3.6 depicts the stability region for a system with $M_p = 20$ primary users, $M_s = 10$ secondary

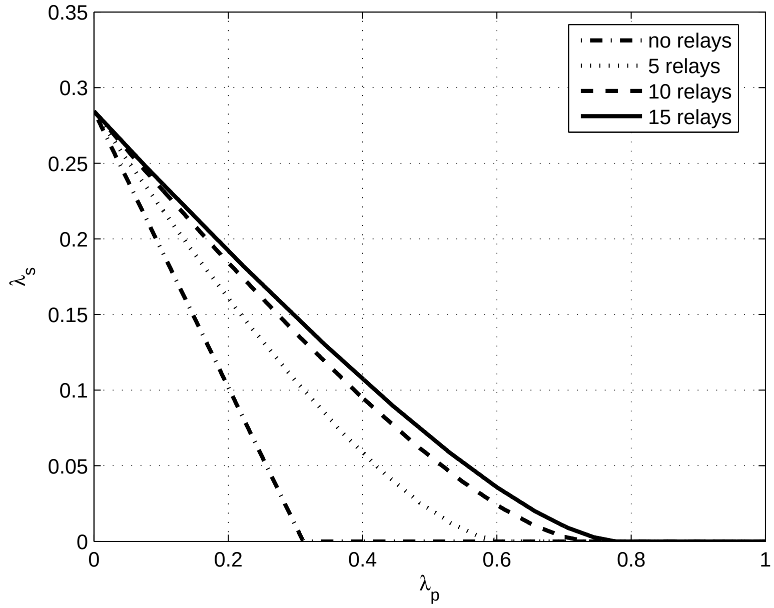

Figure 3.7: System stability region with and without cooperation in case of colliding relay and secondary transmissions for different number of relays.

users, and $M_r = 10$ relays nodes, for different values of α_r, and for different values of relays in Fig. 3.7. Despite the fact that relay and secondary nodes are competing for idle channel resources, significant improvements in the stability regions of both primary and secondary network are observed. It is noted that, as in the case without collisions, both networks benefit from assigning more resources for cooperation, or increasing the number of relays in the primary network, although this increase is apparently increasing the probability of collision. This is mainly because the gains of cooperation on the service rates of (3.34), (4.30),(3.39), and (4.32) exceed the degradation caused by collisions. These gains are even more significant for higher primary arrival rate, where secondary nodes achieve much higher throughput even in the case of increased interference.

To illustrate how the cooperation gains outweigh the degradation due to colli-

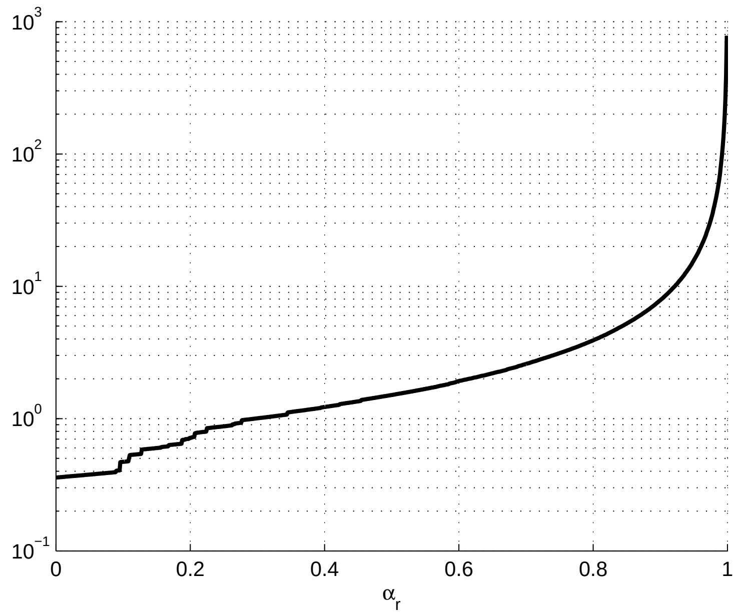

Figure 3.8: Ratio of $\left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right)$ to $(1 - \alpha_r)$ as function of α_r.

sions, Fig. 3.8 depicts the ratio between the terms $\left(1 - \sum_{i \in \mathcal{M}_p} \frac{\lambda_i^p}{P_i}\right)$ and $(1 - \alpha_r)$ that are the essence of the competition between these gains and degradation in performance. It can be seen that below $\alpha_r = 0.3$ this ratio is less than 1, therefore, the result is a degradation in performance, since relays are not providing enough help to the primary network, and at the same time causing collisions with secondary transmissions. For $\alpha_r > 0.3$, cooperation gains start to outweigh collision losses, and we start to notice throughput increase in both primary and secondary networks. Finally, it is noted that the ratio in question is monotonically increasing in α_r, which agrees with the conclusions drawn earlier that it is always better to assign all the idle resources to cooperation.

Chapter 4

Joint Design of Spectrum Sensing and Channel Access

In Chapter 3, opportunistic access to idle channel resources was considered. Specifically, it was shown that idle resources in a wireless communications network could be used for two groups of cognitive radio systems. The first system is composed of a group of cooperative relays that have the ability to sense the medium for transmission opportunities and use these opportunities to offer diversity to primary nodes, or the owners of the channel resources. The second system is comprised of a group of secondary nodes that are interested in using the idle channel resource to transmit their own data packets. Furthermore, the problem of sharing the under-utilized channel resources between those two systems was thoroughly investigated.

To identify transmission opportunities, cognitive radio nodes resort to sensing the wireless channel to detect the presence or absence of primary activity. The fundamental assumption in Chapter 3, and in much of the work on cognitive radios and dynamic spectrum sharing [58–61], is that the effects of the sensing mechanism on the performance of the channel access mechanism is negligible. In other words, the assumption is made that secondary nodes have perfect knowledge of whether primary nodes are active or not.

The above assumption allows the researcher to deal with the main two problems of dynamic spectrum sharing and cognitive radios separately. The first problem is spectrum access coordination between different secondary users while limiting the level of interference to the primary system. In the literature, this problem has been addressed on a negotiating/pricing basis [62–67] or an opportunistic basis [68], [69]. The second problem is the design of highly sensitive detectors to accurately detect the presence or absence of transmissions from primary users.

This chapter will focus on the effects of spectrum sensing errors on the performance of cognitive radio networks. While the issue of spectrum sensing errors has been investigated at the physical layer [54, 70–73], cognitive multiple access design in the presence of sensing errors has received little attention. This chapter will deal with that latter aspect of the problem; specifically, we try to answer the questions: How does the spectrum sensing errors affect the performance of the cognitive radio network from a multiple access protocol design point of view? And, how can the joint design of spectrum sensing and access mechanisms mitigate the negative effects of sensing errors?

To answer the questions posed above, this chapter starts by studying the effects of channel sensing errors on the performance of the multiple access layer. This is achieved through a queueing theoretical analysis of the stability regions of both primary and secondary networks. The stability region is characterized for different operating points on the receiver operating characteristic (ROC) of the energy detector based spectrum sensor. Results reveal a significant reduction in the stability region of both networks due to sensing errors.

To mitigate the negative effects of sensing errors, this chapter proposes a novel joint design of the spectrum sensing and access mechanisms. The design is based on the observation that, in a binary hypothesis testing problem, the value of the test statistic could be used as a measure of how confident we are in the test outcome. The further the value of the test statistic is from the decision threshold, the more confident we are that the decision is correct. Therefore, instead of using the hard decisions of the spectrum sensor to decide whether to access the channel or not, a secondary user can have different access probabilities for different values of the test statistic. For instance, the access probability could be higher for the values of the test statistic further away from the decision threshold, and vice versa. Using this technique, one can set the target false alarm probability as low as possible for the secondary nodes not to overlook spectrum opportunities. At the same time a low probability of collision with primary users could be maintained since the access probability can be set to an arbitrarily low value near the decision threshold, which is not the case with conventional designs, since lowering the false alarm probability results in an increased probability of missed detections, hence increased probability of collision.

4.1 System Model

As in the previous chapter, we consider the uplink of a TDMA cellular network as the primary network. The primary network consists of M_p source nodes numbered $1, 2, ..., M_p$ communicating with a base station (BS) d_p. A secondary network, consisting of M_s nodes numbered $1, 2, ..., M_s$, tries to exploit the unutilized channel resources to communicate their own data packets using slotted ALOHA as their multiple access protocol. Let $\mathcal{M}_p = 1, 2, ..., M_p$ denote the set of primary nodes,

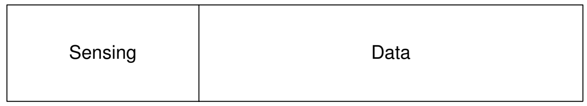

Figure 4.1: Time slot structure, showing the sensing period used by secondary users.

and $\mathcal{M}_s = 1, 2, ..., M_s$ denote the set of secondary nodes. .

Secondary users independently exploit instantaneous spectrum opportunities in the channel (in the form of idle time slots in this model). At the beginning of each slot, a secondary user with data to transmit senses the channel. A spectrum sensor is used to detect the state of the medium (idle or busy). Based on the sensing outcomes, the secondary user decides whether to access the channel or not. At the end of the slot, the receiver acknowledges each successful transmission. The basic slot structure is illustrated in Fig. 4.1.

4.1.1 Channel Model

The wireless channel between a node and its destination is modeled as a Rayleigh flat fading channel with additive white Gaussian noise. The signal received at a receiving node j from a transmitting node i at time t can be modeled as

$$y_{ij}^t = \sqrt{G_i \rho_{ij}^{-\gamma}} h_{ij}^t x_i^t + n_j^t, \tag{4.1}$$

where G_i is the transmitting power, assumed to be the same for all nodes, ρ_{ij} denotes the distance between the two nodes, γ the path loss exponent, h_{ij}^t is the channel fading coefficient between nodes i and j at time t and is modeled as an i.i.d zero mean, circularly symmetric complex gaussian random process with unit variance. The term x_i^t denotes the transmitted signal which has an average unit power and is assumed to be drawn from a constant modulus constellation with zero mean (M-ary PSK for instance). The i.i.d additive white Gaussian noise processes n_j^t have zero mean and variance N_0. Since the arrivals, the channel gains, and the additive noise processes are all assumed stationary, we can drop the index t without loss of generality.

As in the previous chapters, success and failure of packet reception are characterized by outage events and outage probabilities,

$$O_{ij} \triangleq \left\{ h_{ij} : \frac{|h_{ij}|^2 \rho_{ij}^{-\gamma} G_i}{N_0} \leq \beta \right\}, \tag{4.2}$$

and

$$\Pr\{O_{ij}\} = P_{ij}^o = 1 - \exp\left(-\frac{\beta N_0 \rho_{ij}^{\gamma}}{G_i} \right). \tag{4.3}$$

Furthermore, we assume that whenever there is a collision between a primary transmission and a secondary transmission, or between two or more secondary transmissions, all the packets involved are lost.

4.1.2 Queuing Model

Here we adopt the same queuing model used in the previous chapter. Each primary and secondary node has an infinite buffer for storing fixed length packets (see Fig. 4.2). The channel is slotted in time and a slot duration equals the packet transmission time. The arrivals at the i^{th} primary node's queue ($i \in \mathcal{M}_p$), and the j^{th} secondary node's queue ($i \in \mathcal{M}_s$) are Bernoulli random variables, i.i.d from slot to slot with mean λ_i^p and λ_j^s, respectively. Hence, the vector $\Lambda = [\lambda_1^p, ..., \lambda_{M_p}^p, \lambda_1^s, ..., \lambda_{M_s}^s]$ denotes the average arrival rates. Arrival processes are assumed to be independent from one node to another.

Primary users access the channel by dividing the channel resources, time in this case, among them; hence, each node is allocated a fraction of the time. Let

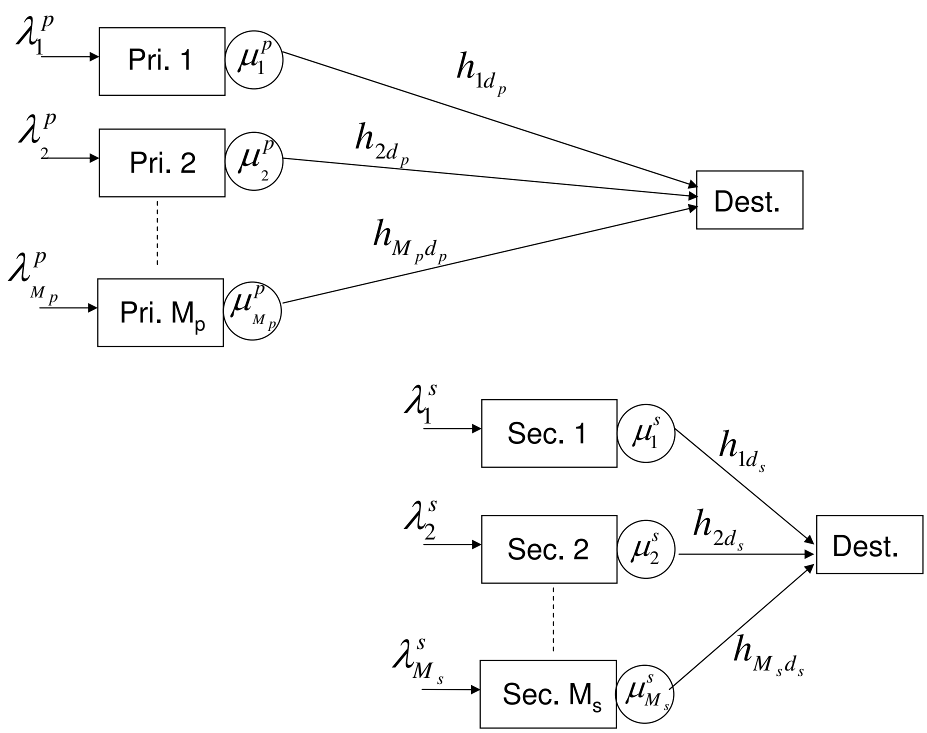

Figure 4.2: Network queuing and channel model

$\mathbf{\Omega_p} = [\omega_1^p, \omega_2^p, ..., \omega_{M_p}^p]$ denote a resource-sharing vector, where $\omega_i^p \geq 0$ is the fraction of time allocated to node $i \in \mathcal{M}_p$, or it can represent the probability that node i is allocated the whole time slot [52]. The set of all feasible resource-sharing vectors is specified as follows

$$\mathcal{F}_p = \left\{ \mathbf{\Omega_p} = (\omega_1^p, \omega_2^p, ..., \omega_{M_p}^p) \in \Re^{+M_p} : \sum_{i \in \mathcal{M}_p} \omega_i^p \leq 1 \right\}. \tag{4.4}$$

In this chapter, as in the previous chapter, the stability of the network's queues will be used as the fundamental performance measure. Recall that if the arrival and service processes of a queueing system are strictly stationary, then one can apply Loynes's theorem to check for stability conditions [53]. This theorem states that if the arrival process and the service process of a queueing system are strictly stationary, and the average arrival rate is less than the average service rate, then the queue is stable; if the average arrival rate is greater than the average service rate then the queue is unstable.

4.2 Effect of Sensing Errors on the Performance

4.2.1 Spectrum Sensing

Spectrum sensing is an essential functionality of cognitive radios, since the devices need to reliably detect weak primary signals of possibly unknown types [54]. In general, spectrum sensing techniques can be classified into three categories: energy detection [74], matched filter coherent detection [75], and cyclostationary feature detection [76]. While these classic signal detection techniques are well known, detecting primary transmitters in a dynamic wireless environment with noise uncertainty, shadowing, and fading is a challenging problem as articulated in [70]. To improve detection accuracy, cooperative spectrum sensing has been proposed [70], [73], [77]. The basic idea is to overcome shadowing and multipath fading by allowing neighboring secondary users to exchange sensing information through a dedicated control channel. In [78], two decision-combining approaches were studied: hard decision with the AND logic operation and soft decision using the likelihood ratio test [79]. It was shown that the soft decision combination of spectrum sensing results yields gains over hard decision combining. In [80], the authors exploited the fact that summing signals from two secondary users can increase the signal-to-noise ratio (SNR) and detection reliability if the signals are correlated.

Since non-coherent energy detection is simple and is able to locate spectrum occupancy information quickly, it will be adopted in our study of the effect of sensing errors on cognitive radios performance, and as the basis for our proposed joint design technique. Detection of the presence of the i^{th} primary user by the j^{th} secondary user can be formulated as a binary hypothesis test as follows,

$$\begin{aligned} \mathcal{H}_0 &: y_{ij}^t = n_j \\ \mathcal{H}_1 &: y_{ij}^t = \sqrt{G_i \rho_{ij}^{-\gamma}} h_{ij}^t x_i + n_j. \end{aligned} \tag{4.5}$$

The null hypothesis $\mathcal{H}_0$ represents the absence of the primary user, hence a transmission opportunity to the secondary user. And the alternative hypothesis $\mathcal{H}_1$ represents a transmitting primary user.

The performance of the spectrum sensor is characterized by the two types of errors and their probabilities, (i) false alarms having probability α, (ii) and missed detections having probability β,

$$\alpha \triangleq \Pr\{\text{decide } \mathcal{H}_1|\mathcal{H}_0 \text{ is true}\}, \tag{4.6}$$

$$\beta \triangleq \Pr\{\text{decide } \mathcal{H}_0|\mathcal{H}_1 \text{ is true}\}. \tag{4.7}$$

The false alarm type of errors where an idle channel is erroneously detected as busy does not incur performance degradation on the primary system, but lowers the potential channel utilization of secondary users. On the other hand, the missed detection events, where a secondary users fails to detect a primary transmission, will result in a collision between primary and secondary transmissions. Therefore, miss detection events will negatively impact the performance of the primary system.

With the assumption that secondary users do not have prior knowledge of primary activity patterns, the probability of miss detection β could be minimized subject to the constraint that the probability of false alarm is no larger than a given value α using the optimal Neyman-Pearson (NP) detector [75].

From the received signal model of (4.1), it follows that under hypothesis $\mathcal{H}_0$ the received signal y_{ij} is a complex Gaussian random variable with zero mean and variance $\sigma_0^2 = N_0$, and under hypothesis $\mathcal{H}_1$, y_{ij} is a complex Gaussian random variable with zero mean and variance

$$\sigma_1^2 = G_i \rho_{ij}^{-\gamma} + N_0.$$

Therefore, the likelihood ratio test for the optimal NP detector can be written as follows,

$$\Lambda(y_{ij}) = \frac{\Pr\{y_{ij}|\mathcal{H}_1\}}{\Pr\{y_{ij}|\mathcal{H}_0\}} = \frac{\frac{1}{\pi\sigma_1^2}e^{-\frac{||y_{ij}||^2}{\sigma_1^2}}}{\frac{1}{\pi\sigma_0^2}e^{-\frac{||y_{ij}||^2}{\sigma_0^2}}}$$
$$= \frac{\sigma_0^2}{\sigma_1^2}e^{-||y_{ij}||^2\left[\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right]} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \eta', \tag{4.8}$$

which can be simplified to

$$||y_{ij}||^2 \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \frac{\eta' - \log\frac{\sigma_0^2}{\sigma_1^2}}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}} = \eta. \tag{4.9}$$

From (4.9), the spectrum sensing problem has been reduced to a simple comparison of the received signal energy $||y_{ij}||^2$ to a threshold η. The optimum threshold could then be calculated through the constraint on the false alarm probability. We first note that, from the received signal model of (4.1), $||y_{ij}||^2$ is exponentially distributed with parameter $1/2\sigma_1^2$ and $1/2\sigma_0^2$, under $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively. Therefore, the false alarm probability is

$$\alpha = \Pr\{||y_{ij}||^2 > \eta|\mathcal{H}_0\} = e^{-\frac{\eta}{2\sigma_0^2}}. \tag{4.10}$$

From which

$$\eta = -2\sigma_0^2\log(\alpha). \tag{4.11}$$

Finally, the probability of misdetection is

$$\beta = \Pr\{||y_{ij}||^2 < \eta|\mathcal{H}_1\} = 1 - e^{-\frac{\eta}{2\sigma_1^2}} = 1 - e^{-\frac{\sigma_0^2\log(\alpha)}{\sigma_1^2}}. \tag{4.12}$$

It is noted that in the design above, the spectrum sensor has based its detection on a single sample of the received signal. Increasing the number of samples will of course increase the reliability of the sensing process. However, we limited ourselves to this design for the purpose of mathematical tractability as it will be clear later.

4.2.2 Performance Analysis

To analyze the effect of sensing errors on the cognitive radio system, we adopt the stability regions of the primary and secondary networks as the performance measure. We will consider a primary network with M_p users and assume all users have the same arrival rate λ_p. The secondary network has M_s users all having the same arrival rate λ_s. We further assume that within each network channels are also symmetric. In other words, all primary users share the same channel statistics, and all secondary users share the same channel statistics. We will begin by characterizing the stability region for the ideal system with no sensing errors to form a base for comparison.

System with Perfect Sensing

We start by characterizing the stability region for the primary system of queues. Since the primary network employs TDMA as a multiple access protocol, it follows directly from Loynes's theorem, that the stability condition for primary network's stability is

$$\lambda_p < \mu_p = \frac{1 - P_{pd}^o}{M_p}, \tag{4.13}$$

where P_{pd}^o is the outage probability of the link between any primary user and its destination, and the division by M_p accounts for the fact that the channel is divided among M_p users.

For the secondary network, we recall that secondary users employ slotted ALOHA to share idle time slots among themselves. Therefore, when an idle time slot is detected, a node with non-empty queues will try to transmit the packet at the head of its queue with access probability p_s. We note that, because of the possible collisions between secondary transmissions, secondary users' queues are interacting. In other words, the service rate of a given queue is dependent on the state of all other queues, whether they are empty or not.

To study the stability of the interacting system of queues consisting of secondary users queues, we make use of the dominant system approach we used in the previous chapter. We define the dominant system as follows:

- Arrivals at each queue in the dominant system are the same as in the original system.
- Time slots assigned to primary node $i \in \mathcal{M}_p$ are identical in both systems.
- The outcomes of the "coin tossing" (that determines transmission attempts of relay and secondary nodes) in every slot are the same.
- Channel realizations for both systems are identical.
- The noise generated at the receiving ends of both systems is identical.
- In the dominant system, secondary nodes attempt to transmit dummy packets when their queues are empty.

The service process of a secondary node depends on the idle time slots unused by the primary users. Therefore, the service process of the k^{th} secondary user can be modeled as

$$Y_k^t = \sum_{i \in \mathcal{M}_p} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t = 0\} \, \overline{O_{kd}^t} \bigcap P_s \bigcap_{l \in \mathcal{M}_s \backslash k} \{\overline{P_s}\} \right], \tag{4.14}$$

where A_i^t denotes the event that slot t is assigned to primary user i, $\{Q_i^t = 0\}$ denotes the event that this user's queue is empty, i.e., the node has no packet to transmit, and according to Little's theorem [55] it has probability $(1 - \lambda^p/\mu_p)$. From the condition if (4.13), $\mu_p = (1 - P_{pd}^o)/M_p$. Event P_s is the event that a secondary node has permission to transmit, which has probability p_s. Therefore, the event $\bigcap P_s \bigcap_{l\in\mathcal{M}_s\backslash k} \{\overline{P_s}\}$ is that only one secondary node is transmitting in the current time slot; otherwise a collision will occur and all packets involved will be lost. Finally, $\overline{O_{kd}^t}$ denotes the event that the k^{th} secondary node link to its destination is not in outage.

Assuming that primary queues are stable, then they offer stationary empty slots. Also the channel statistics are stationary; hence, the secondary service process is stationary. The average secondary service rate is then given by

$$\mu^s = E[Y_k^t] = \left(1 - \frac{\lambda^p M_p}{1-P_{pd}^o}\right) p_s(1-p_s)^{M_s-1}(1-P_{sd}^o), \tag{4.15}$$

where P_{sd}^o is the outage probability of the link between any primary user and its destination. Because of the symmetry assumption made above, this average service rate is equal for all secondary users.

From (4.15), it can be easily shown that the optimum value for the secondary access probability is

$$p_s = \frac{1}{M_s}.$$

Using Loyne's theorem along with (4.15), and from (4.13), the stability region of the system defined by the primary and secondary users can be written as

$$\mathcal{R} = \Bigg\{ (\lambda_p, \lambda_s) \in R^{+2} : \lambda_p < \frac{1-P_{pd}^o}{M_p}, \quad \lambda_s < \left(1 - \frac{\lambda^p M_p}{1-P_{pd}^o}\right) \frac{1}{M_s}(1-\frac{1}{M_s})^{M_s-1}(1-P_{sd}^o) \Bigg\}. \tag{4.16}$$

It is clear from (4.13) and (4.15) that the primary network stability is completely independent from the secondary network operation. The secondary network stability is dependent on the primary network through the condition of empty primary queues for secondary queues service to take place.

System with Non-Perfect Sensing

In the case of non-perfect sensing, the events of miss detection will result in simultaneous primary and secondary transmissions leading to collisions and data loss. Because of these collision events, primary and secondary queues are now interacting. To analyze this interacting system of queues, we will resort to the dominant system of the previous section.

Under the dominant system in which secondary users attempt to transmit dummy packets if their queues are empty, the service process of the i^{th} primary user can be defined as follows,

$$Y_i^t = \left[A_i^t \bigcap \overline{O_{id}^t} \bigcap_{l \in \mathcal{M}_s} \left\{ \overline{\mathcal{B} \bigcap P_s} \right\} \right], \tag{4.17}$$

where $\mathcal{B}$ is the event of a miss detection. Therefore, $\overline{\mathcal{B} \bigcap P_s}$ is the complement of the event that a secondary user miss detects primary activity and has access permission, hence causing a collision with the primary user. From the definition of the service process, it follows that the average primary service rate is defined as

$$\mu_p = E[Y_i^t] = \frac{1 - P_{pd}^o}{M_p} (1 - \beta p_s)^{M_s}. \tag{4.18}$$

Similarly, the the service process of the k^{th} secondary user can be written as

$$Y_k^t = \sum_{i \in \mathcal{M}_p} \left[A_i^t \bigcap \{Q_i^t = 0\} \bigcap \overline{O_{kd}^t} \bigcap \overline{\mathcal{A}} \bigcap P_s \bigcap_{l \in \mathcal{M}_s \setminus k} \left\{ \mathcal{A} \bigcup \overline{P_s} \right\} \right], \tag{4.19}$$

where $\mathcal{A}$ is the event of false alarm. The above definition of the secondary service process accounts for the fact that, for a secondary user to gain uncontested access to an idle time slot, it should correctly identify the slot as idle and have access permission. At the same time for all other secondary users not to access that slot, they either do not have access permission or they detect the time slot as busy.

From this definition, the average secondary service rate is then

$$\mu_s = E[Y_k^t] = \left(1 - \frac{\lambda^p M_p}{(1-P_{pd}^o)(1-\beta p_s)^{M_s}}\right)(1-\alpha)p_s(1-(1-\alpha)p_s)^{M_s-1}(1-P_{sd}^o). \tag{4.20}$$

Using Loyne's theorem along with (4.18), and (4.20), the stability region for a given secondary access probability p_s of the system defined by the primary and secondary users can be written as

$$\mathcal{R}(p_s) = \Bigg\{ (\lambda_p, \lambda_s) \in R^{+2} : \lambda_p < \frac{1-P_{pd}^o}{M_p}(1-\beta p_s)^{M_s}, \\ \lambda_s < \left(1 - \frac{\lambda^p M_p}{(1-P_{pd}^o)(1-\beta p_s)^{M_s}}\right)(1-\alpha)p_s(1-(1-\alpha)p_s)^{M_s-1}(1-P_{sd}^o) \Bigg\}, \tag{4.21}$$

and the maximum stability region can be determined by taking the union over all possible values of p_s as follows,

$$\mathcal{R} = \bigcup_{p_s \in [0,1]} \{\mathcal{R}(p_s)\}, \tag{4.22}$$

which can be simply written as the following constrained optimization problem

$$\max_{p_s \in [0,1]} \mu_s = \left(1 - \frac{\lambda^p M_p}{(1-P_{pd}^o)(1-\beta p_s)^{M_s}}\right)(1-\alpha)p_s(1-(1-\alpha)p_s)^{M_s-1}(1-P_{sd}^o), \\ \text{s.t.} \quad \lambda_p < \frac{1-P_{pd}^o}{M_p}(1-\beta p_s)^{M_s}, \tag{4.23}$$

which requires a 1-D search and can be solved using standard methods [81].

Degradation in performance due to sensing errors is clear from (4.18) and (4.20). It is seen that the average primary service rate is a monotonically decreasing function of the misdetection probability β. Therefore, in order not to severely degrade primary performance, where such a degradation contradicts the principle that presence of the secondary system should be transparent to the primary system, spectrum sensors should be designed with the lowest possible β. Moreover, with lower primary service rate, the channel will busy with higher probability, which negatively affects secondary users, since there will be no enough idle time slots for them to use. But, decreasing β comes at the expense of a higher false alarm rate α, which from (4.20) will degrade the performance of secondary users.

4.2.3 Numerical Results

To see how non-perfect spectrum sensing affects the stability region of the system of primary and secondary users, we consider a network with $M_p = 2$ primary users and $M_s = 2$ secondary users. Distance between primary users and their destination is set to 100m, distance between secondary users and their destination is also 100m, and distance between primary and secondary users is 150m. SNR threshold is 25dB, transmit power is 100mW, path loss exponent $\gamma = 3.7$, and $N_0 = 10^{-11}$.

Fig. 4.3 depicts the ROC for the used spectrum sensor. It can be seen that it has moderate performance. Fig. 4.4 compares the stability region of the system with perfect sensing and the system with non-perfect sensing for different values of the false alarm rate α. The negative effect of sensing errors on the stability of both primary and secondary users is clearly seen. For instance, for a primary arrival rate of $\lambda_p = 0.1$, the maximum stable secondary throughput suffered a 74% reduction. Furthermore, above a primary arrival rate of $\lambda_p = 0.15$ no secondary user can exist in the system; otherwise the whole system of queues will become unstable.

From Fig. 4.4 we see that by allowing the false alarm rate to increase in the

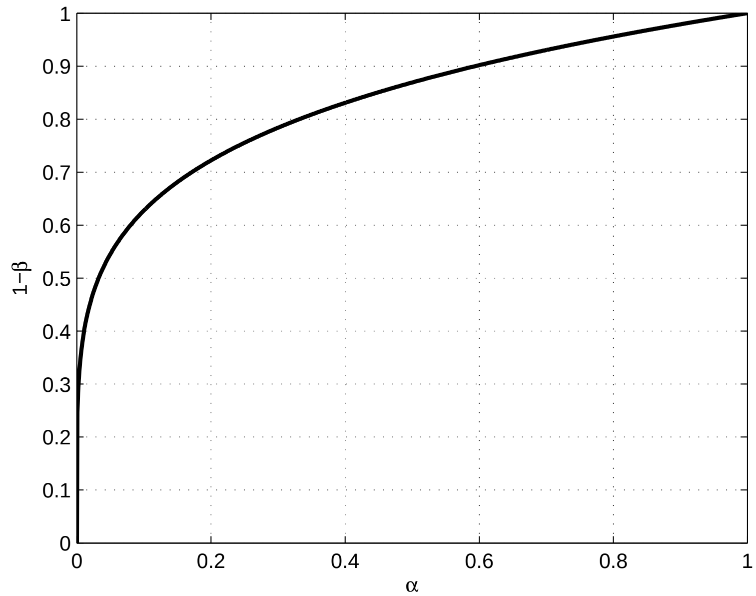

Figure 4.3: ROC for the spectrum sensor in use.

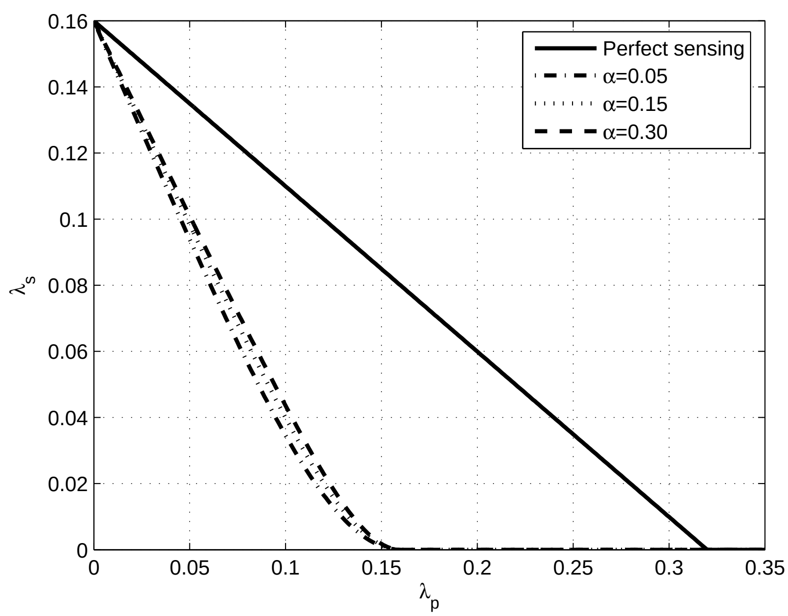

Figure 4.4: Effect of sensing errors on system stability.

detector design, a very slight improvement in secondary throughput is noticed. This is mainly because of the reduction in missed detection probability associated with the increase in the false alarm probability. By reducing the missed detection probability, primary users will have better service rates, hence higher probability of having empty queues and idle time slots. It is noted from (4.20), that the increase in false alarm rate and reduction in missed detection probability are affecting secondary throughput in opposite directions. However, the results of Fig. 4.4 indicates that the gains of reducing the missed detection probability outweigh the degradation due to increased false alarm rate.

4.3 Joint Design of Sensing and Access Mechanisms

In the previous section the detrimental effects of the errors in spectrum sensing were characterized. One of the main causes of these effects is that secondary users base their channel access decisions solely on the outcomes of the spectrum sensor without taking into consideration the possibility that those outcomes incorrect.

For the secondary users to have better channel access decisions, it is necessary to find a method with which they can assess the reliability of the spectrum sensor outcomes. Here we propose the use of the decision statistic $||y_{ps}||^2$ used by the energy detector as a measure for the reliability of the spectrum sensor decisions.

The reasoning behind the use of the value of the decision statistic is that under hypothesis $\mathcal{H}_0$, the value of $||y_{ps}||^2$ has a much higher probability of being closer to zero and far away from the threshold, as can be seen in Fig. 4.5 depicting the CDF of $||y_{ps}||^2$ under both hypotheses. Therefore, the lower the value of $||y_{ps}||^2$,

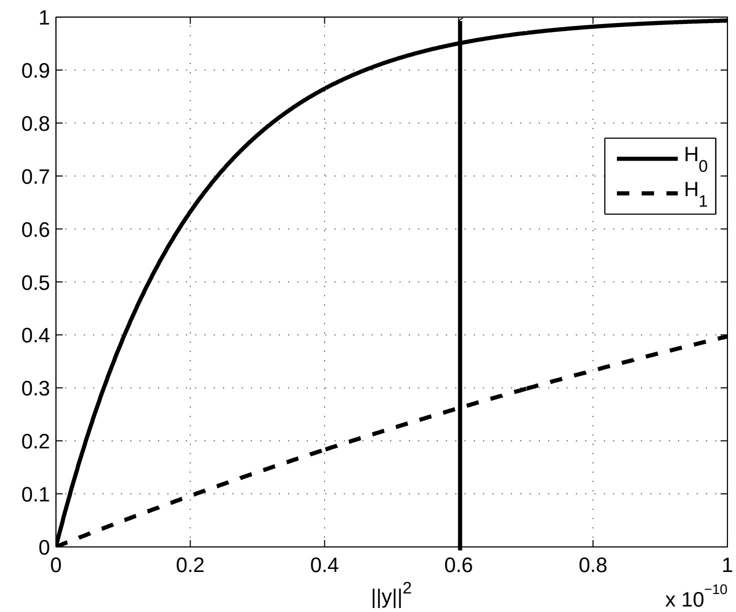

Figure 4.5: CDF of the decision statistic under both hypotheses. The vertical line marks the position of the decision threshold.

the more likely hypothesis $\mathcal{H}_0$ is true, and the more reliable the decision is. On the other hand, as the value of the decision statistic approaches the decision threshold it is more or less equally likely that it is resulting from either one of the hypotheses. Therefore, the closer the value of $||y_{ps}||^2$ is to the decision threshold, the less reliable the outcome of the spectrum sensor is.

In order to exploit the reliability measure established above in taking channel

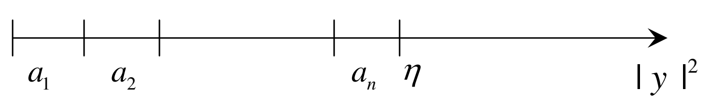

Figure 4.6: Division of the interval $[0, \eta]$ into subintervals and the associated access probabilities.

access decisions, we propose the following scheme for channel access;

- The interval $[0, \eta]$ is divided into n subintervals as shown in Fig. 4.6.
- For each subinterval $i \in [1, n]$, assign an access probability a_i.
- Whenever the decision statistic falls in the i^{th} interval, secondary user will access the channel with the associated access probability.
- In the case when $||y_{ps}||^2 > \eta$, secondary user does not access the channel.

This scheme will enable us to have higher access probabilities for the subintervals closer to zero, since in these subintervals there is a very low probability of colliding with primary transmissions. Moreover, assign lower probabilities to the subintervals close to the decision threshold, where there is a higher risk of collisions.

It should be noted that under the proposed scheme, the decision threshold η is not necessarily chosen according to the Neyman-Pearson detector design.

4.3.1 Design Methodology

In this chapter, we have considered in the stable throughput region of both primary and secondary networks as the main measure of performance. In this section we will use the definition of the stability regions as the main design criteria of our proposed channel access scheme.

Since the cognitive principle is based on the idea that the presence of the secondary system should be "transparent" to the primary, and since we are interested in the stable throughput of primary and secondary networks, we define the secondary system "transparency" as not affecting primary stability. In other words, for a given stable primary system with arrival rate λ_p, secondary activity will be considered transparent if the primary system maintains its stability in the presence of the secondary system. Therefore, the main design criteria for the secondary access scheme will be to maximize its own throughput under the constraint that primary stability is not affected. This design criteria can be formulated as the following constrained optimization problem

$$\max_{a_i, i\in[1,n]} \mu_s, \\ \text{s.t.} \quad \lambda_p < \mu_p. \tag{4.24}$$

To solve the optimization problem of (4.24), we start by calculating the average primary service rate μ_s under the proposed secondary access scheme. As in the previous section, and since collisions between primary and secondary transmission are inevitable, the group of primary and secondary queues are interacting. Therefore, to decouple this interaction, we resort to the dominant system defined in the previous section in which secondary users attempt to transmit dummy packets if their queues are empty. Under this system, the service process of the j^{th} primary user will still be given by (4.17). The difference will be in the definition of the event that a secondary user's transmission collides with primary transmission, $\{\mathcal{B} \bigcap P_s\}$. Here the events of missed detection and channel access can no longer be separated as in (4.17) and (4.18). The probability of that joint event is now

$$\Pr\left\{\mathcal{B} \bigcap P_s\right\} = p_s^1 = \sum_{i\in[1,n]} p_i^1 a_i, \tag{4.25}$$

where a_i is the access probability associated with subinterval i (see Fig. 4.6), and p_i^1 is the probability that the value of $||y_{ps}||^2$ falls in the i^{th} subinterval when hypothesis $\mathcal{H}_1$ is true (primary user exists in the channel), which from the signal model of (4.1) is calculated as

$$p_i^1 = \exp\left(\frac{(i-1)\eta}{2n\sigma_1^2}\right) - \exp\left(\frac{i\eta}{2n\sigma_1^2}\right). \tag{4.26}$$

Similarly, we define the probability that a secondary user accesses the channel when hypothesis $\mathcal{H}_0$ is true as

$$p_s^0 = \sum_{i\in[1,n]} p_i^0 a_i, \tag{4.27}$$

where

$$p_i^0 = \exp\left(\frac{(i-1)\eta}{2n\sigma_0^2}\right) - \exp\left(\frac{i\eta}{2n\sigma_0^2}\right). \tag{4.28}$$

Therefore, the average primary service rate is given by

$$\mu_p = \frac{1-P_{pd}^o}{M_p}\left(1-\sum_{i\in[1,n]} p_i^1 a_i\right)^{M_s}. \tag{4.29}$$

The secondary users' service process is still given by (4.19), with the event that the secondary user has access to the channel under hypothesis $\mathcal{H}_0$, $\left\{\overline{\mathcal{A}}\bigcap P_s\right\}$, now having probability p_s^0 defined in (4.27). Therefore, the average secondary service rate can be written as

$$\mu_s = \left(1-\frac{\lambda^p M_p}{\frac{1-P_{pd}^o}{M_p}\left(1-\sum_{i\in[1,n]} p_i^1 a_i\right)^{M_s}}\right)\left(1-\sum_{i\in[1,n]} p_i^0 a_i\right)^{M_s-1}(1-P_{sd}^o)\sum_{i\in[1,n]} p_i^0 a_i. \tag{4.30}$$

From (4.29) and (4.30), it can be easily seen that the optimization problem of (4.24) is a non-convex problem, which renders the task of secondary users in designing their access strategy very complex, and not guaranteed to yield the global optimal solution if any.

In order to simplify the optimization problem, and convert it into a more tractable problem to solve, expressions for primary and secondary service rates (4.29) and (4.30) will be simplified and bounded as follows.

First, we approximate the expressions (4.29) and (4.30) by expanding the two terms $\left(1-\sum_{i\in[1,n]} p_i^1 a_i\right)^{M_s}$ and $\left(1-\sum_{i\in[1,n]} p_i^0 a_i\right)^{M_s-1}$ and only retaining the first order terms. Therefore, (4.29) and (4.30) are now

$$\mu_p \approx \frac{1-P_{pd}^o}{M_p}\left(1-M_s\sum_{i\in[1,n]} p_i^1 a_i\right), \tag{4.31}$$

and

$$\mu_s \approx \left(1-\frac{\lambda^p M_p}{\frac{1-P_{pd}^o}{M_p}\left(1-M_s\sum_{i\in[1,n]} p_i^1 a_i\right)}\right)\left(1-(M_s-1)\sum_{i\in[1,n]} p_i^0 a_i\right) \times(1-P_{sd}^o)\sum_{i\in[1,n]} p_i^0 a_i. \tag{4.32}$$

The approximation in (4.31) has transformed the constraint of the optimization problem into a linear constraint, but the approximate objective function in (4.32) is still non-convex. To further simplify the objective function we note that, for a stable primary system, the average departure rate from primary queues is equal to their average arrival rate. Therefore, the average number of idle time slots seen by secondary users will always be the same regardless of the actual primary service rate as long as the system is stable.

From the above observation we can conclude that; for all sets of secondary access probabilities that satisfy the optimization constraint (i.e., maintain primary system stability), the variation of the term, $\left(1-\frac{\lambda^p M_p}{\frac{1-P_{pd}^o}{M_p}\left(1-M_s\sum_{i\in[1,n]} p_i^1 a_i\right)}\right)$, which represents the amount of idle time slots available to the secondary system, is in fact small. Therefore, we can safely drop this term from the objective function since it is almost constant over the feasible set of access probabilities. Now the optimization problem has been transformed to the following convex optimization problem

$$\max_{a_i, i\in[1,n]} \left(1-(M_s-1)\sum_{i\in[1,n]} p_i^0 a_i\right)(1-P_{sd}^o)\sum_{i\in[1,n]} p_i^0 a_i,$$
$$\text{s.t.} \quad \lambda_p < \frac{1-P_{pd}^o}{M_p}\left(1-M_s\sum_{i\in[1,n]} p_i^1 a_i\right), \tag{4.33}$$

which can be solved using traditional optimization techniques [81].

4.3.2 Results and Discussions

Here we compare the performance of the proposed joint design of spectrum sensing and channel access mechanisms with the conventional approach based on the Neyman-Pearson (N-P) detector design. We consider a system with $M_p = 4$ primary users and $M_s = 4$ secondary users.

Fig. 4.7 illustrates the stability regions for the ideal case with no sensing errors, for the N-P based detector, and our joint design scheme with $n = 4$ subintervals, using the same threshold as the one used by the N-P design. Huge improvement in the maximum stable throughput of both primary and secondary networks is observed with our proposed scheme. We see that the range of primary arrival rates for which secondary users can exist in the system without affecting primary stability is now equal to that of the ideal system, compared to only 25% of that value with the conventional design. This significant improvement is mainly because the joint design criteria does not blindly rely on the spectrum sensor to decide when to access the channel.

To get more insight into how the channel access probabilities are selected, Fig. 4.8 depicts the channel access probabilities as a function of primary arrival rate.

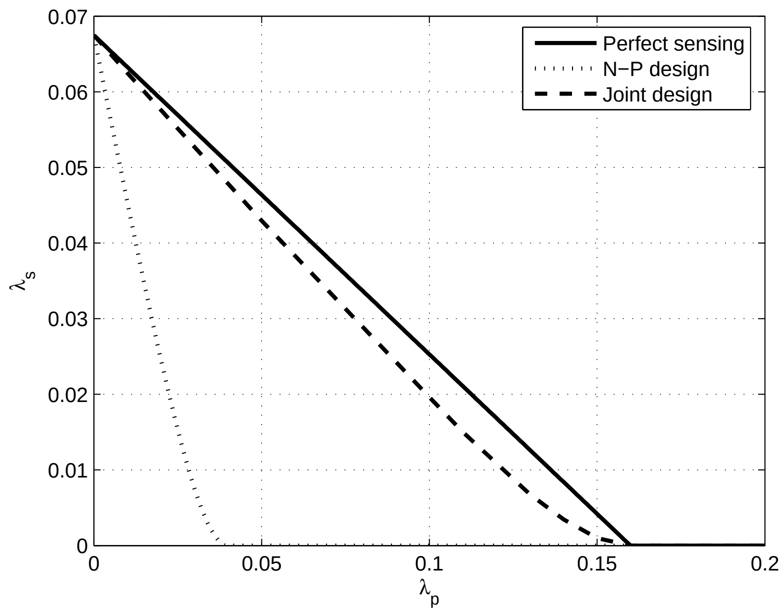

Figure 4.7: Comparison between the joint design and the N-p design.

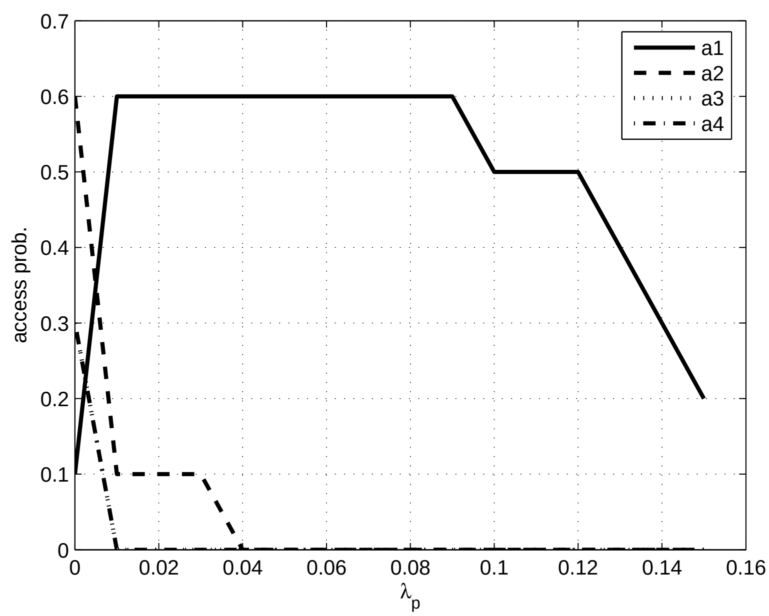

Figure 4.8: Secondary access probabilities.

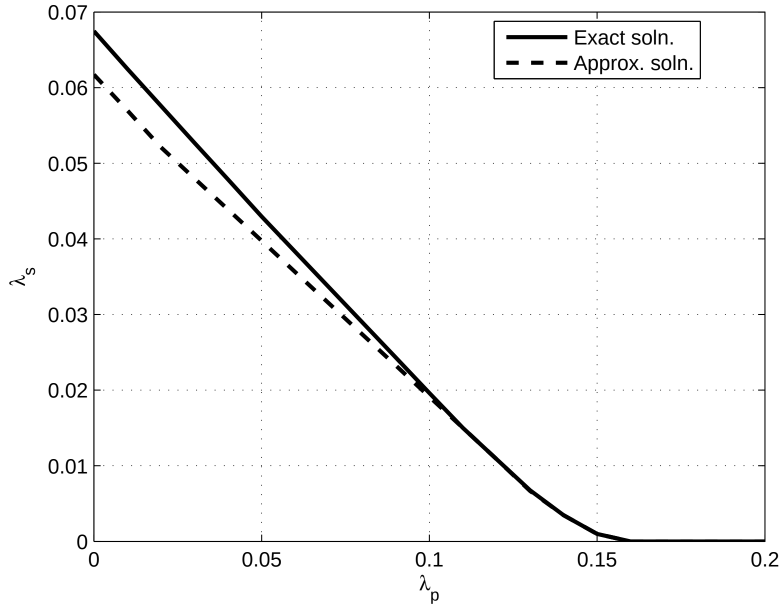

Figure 4.9: Comparison between the solution of the exact and approximate optimization problems.

It is noted that a_1, the access probability for the interval nearest to zero, takes the highest values. This is expected since measurements that land in the corresponding interval have the highest probability of being generated when no primary users are in the channel, hence it is safe that secondary users transmit. As the primary arrival rate increases, all the access probabilities decrease in order to guarantee the stability of primary queues. It is also noted that a_3 and a_4 are overlapping and are zero except at $\lambda_p = 0$, which means that transmitting in the corresponding intervals is the cause of most of the collisions.

Fig. 4.9 compares between the exact solution to the optimization problem of (4.24) (obtained through exhaustive search) and the solution to the simplified problem defined in (4.33). The two solutions are closely matching. We note some deviation (maximum deviation is less than 8%) for low values of primary arrival rate. For low arrival rates, the imposed stability condition is rather easy to satisfy; therefore, secondary users still have a margin of improvement by considering the exact objective function. For high primary arrival rates the approximate solution coincides with the exact one, revealing that the approximation gets better when the arrival rate increases, since the stringent stability constraint in this case does not leave any chance for improvement.

Chapter 5

Random Access Cooperative Networks

Enabling cooperation in a wireless network will in general introduce additional transmissions over the channel. These additional transmissions are either relay transmissions to forward source nodes packets, or signaling information generally needed in order for the relays to coordinate their actions. In large random access networks without a centralized scheduler as in IEEE 802.11 systems [82], these extra transmissions will increase the number of packet collisions and it is not clear if there is any benefit of using physical layer cooperation in this case. In this chapter, we tackle this problem by studying the possible benefits of cooperative communications from a MAC layer perspective. In particular, we focus on cooperative communications in wireless random access networks based on the IEEE 802.11 protocol.

Specifically, this chapter tries to answer the questions of: How can cooperation be enabled in a random access network without a possible increase in the number of packet collisions? And, since cooperation introduces extra transmissions in the channel, what are the benefits and possible tradeoffs associated with cooperation in this case?

To answer these questions, this chapter starts by proposing a cooperative protocol in which a relay node is deployed to help different network nodes to forward their packets to the access point (AP). Presence of the relay will help improve the communication channel through the spatial diversity it creates. The main challenge is how to minimize the collision probability in order for the relay presence not to degrade the network performance instead of improving it. A relay can achieve this goal by intelligently accessing the wireless medium at times when it is guaranteed that no other node is attempting to access the medium. Analysis of the IEEE 802.11 protocol and the CSMA/CA protocol on which it is based reveals that such time instants exist at the end of each packet transmission. After a packet is transmitted over the channel, regardless of the outcome of this transmission, all network nodes will have to wait for a constant plus a random amount of time before making a transmission attempt. By allowing the relay to deviate from this access mechanism, and to access the channel immediately after each transmission attempt on the channel, the relay is guaranteed to have an uncontested access to the wireless medium. But, there may exist times when no node is transmitting (possibly because of empty queues), and the relay is willing to access the medium to transmit packets remaining in its queue. In this case, the relay will revert to the normal access scheme, in which it will wait for a random amount of time and then try to access the channel.

In a network operating as described above, all nodes's queues are interacting; i.e., as discussed in the previous chapters, the service process of a given queue depends on the state of all other queues (whether they are empty or not). Interaction between queues is mainly because of the possible collisions that occur if more than one node is trying to access the channel at the same instant. And, for the relay, its own ability to access the channel is dependent on other nodes transmissions. In order to capture this queue interaction, and be able to analyze the performance of our protocol, two coupled Markov models are used to describe the operation of the relay and other network nodes. These Markov models are able together to completely describe the dynamics of the network and interactions between different nodes. Moreover, queuing analysis is used to analyze the delay performance of the network. The results presented reveal significant gains in terms of network throughput, delay, and the number of supported nodes, through cooperation and our proposed protocol. Furthermore, it is shown that, by virtue of the protocol design, the collision probability has decreased rather than increased due to extra transmissions on the channel.

Related works that study the impact of cooperation in random access networks are few [20], [21], [22]. In [21], the authors proposed a distributed version of network diversity multiple-access (NDMA) [23] protocol and they provided pairwise error probability analysis to demonstrate the diversity gain. In [20] and [22], the authors presented the notion of utilizing the spatial separation between users in the network to assign cooperating pairs (also groups) to each other. In [22], spread spectrum random access protocols were considered in which nearby inactive users are utilized to gain diversity advantage via cooperation assuming a symmetrical setup where all terminals are statistically identical. However, the previously cited works still focus on physical layer parameters such as the diversity gains achieved and the outage probability. User cooperation in a slotted ALOHA random access network was investigated in [24], where the gains of cooperation on stability region of a network consisting of multiple cooperating pairs is characterized.

5.1 IEEE 802.11 DCF Operation

The distributed coordination function (DCF) is the fundamental medium access mechanism in the IEEE 802.11 protocol [82]. This is a random access method based on the CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) with binary slotted exponential backoff. A node with a packet to transmit invokes the carrier sensing mechanism to determine the busy/idle state of the channel. If the channel is sensed to be idle for a period of time equal to a Distributed Inter-Frame Space (DIFS), the node proceeds with packet transmission. Otherwise, if the channel is sensed to be busy (either immediately or during the DIFS), the node continues monitoring the channel until it is measured idle for a DIFS. The node then defers for a randomly selected backoff interval, initializing its random backoff timer, which is decremented as long as the channel is sensed idle. The backoff timer is frozen when a transmission is detected and is reactivated when the channel is sensed idle again for more than one DIFS interval. This random backoff mechanism constitutes the collision avoidance feature of the protocol. Its goal is to minimize the probability of collision with packets being transmitted by other nodes. In addition, to avoid channel capture, a node must wait a random backoff time between two consecutive new packet transmissions, even if the channel is sensed idle in the DIFS time.

The time immediately following an idle DIFS is slotted, and a node is allowed to transmit only when its backoff timer reaches zero and at the beginning of each slot time. The slot size, σ, is set equal to the time needed at any node to detect the transmission from any other node. It depends on the physical layer, and it

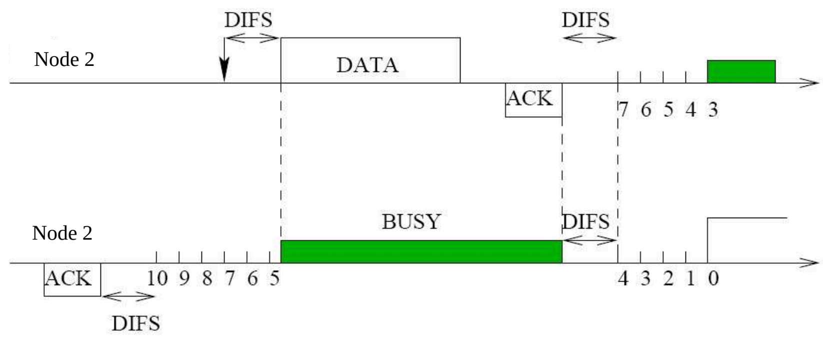

Figure 5.1: DCF basic access mechanism; numbers in figure represent node's backoff timer.

accounts for the propagation delay, and the time needed to detect a busy channel.

The random backoff interval is uniformly chosen in the range $(0, w-1)$. The value w is called the contention window, and depends on the number of transmissions failed for the packet. At the first transmission attempt, w is set equal to a minimum contention window value $CW_{\min}$. After each unsuccessful transmission (due to packet collision or channel error), w is doubled, up to a maximum value $CW_{\max} = 2^m CW_{\min}$, where m is the maximum backoff stage. Once w reaches $CW_{\max}$, it remains at this value until it is reset to $CW_{\min}$ after the successful transmission of a packet.

Fig. 5.1 illustrates the DCF operation. Two nodes 1 and 2 share the same wireless channel. At the end of the packet transmission, node 2 waits for a DIFS and then chooses a backoff time equal to 10, before transmitting the next packet. A packet arrives at node 1 at the time indicated with an arrow in the figure. After DIFS, the packet is transmitted. Note that the transmission of packet 1 occurs in the middle of the slot time corresponding to a backoff value, for node 2, equal to 5. As a consequence of the channel sensed busy, the backoff time is frozen to its value 5, and the backoff counter decrements again only when the channel is sensed idle for a DIFS.

To signal the successful packet reception, an ACK is transmitted by the destination. The ACK is immediately transmitted at the end of the packet, after a period of time called short inter-frame space (SIFS). As the SIFS is shorter than a DIFS, no other station is able to detect the channel idle for DIFS until the end of the ACK. If the transmitting node does not receive the ACK within a specified ACK_Timeout, the packet is assumed to be lost and the node reschedules the packet transmission according to the given backoff rules.

5.2 Cooperation Protocol

Inherent wireless channel fading and transmission errors have a significant impact on the network's performance [83]. In IEEE 802.11 based networks and in wireless networks in general, nodes are unable to detect collisions by hearing their own transmission. Therefore, there is no means to differentiate between a packet loss due to collision from a packet loss due to wireless channel impairments. Because of that, a source node will deal with a wireless channel induced packet loss the same way it deals with a collision induced packet loss. Hence, the contention window is doubled and the node waits for a random amount of time before reattempting transmission.

In case of a collision, the nodes involved in the collision assume that the network is congested; hence, it doubles its contention window size and waits for a longer amount of time before reattempting transmission. This approach is effective in reducing the probability of collision the next time a transmission attempt is made. In case of a channel induced packet loss, the node involved will make the same assumption of a congested channel and unnecessarily invoke the backoff procedure. As a result of invoking the backoff procedure in an non congested channel, the network will suffer from an increased delay and lower achievable throughput [84], [85].

To combat the wireless channel impairments leading to these problems, we propose the deployment of a cooperative relay node into the coverage area of the wireless network. The cooperative relay node will help combat the channel fading through the introduction of spatial diversity into the network. A relay node will help source nodes forward their packets by operating in an incremental decode-and-forward mode [12]. In this mode, in case of a packet loss, the relay first decodes the received packet, and then re-encodes and forwards a regenerated version of the packet to the access point. Like all other network nodes, the relay node make use of the AP's ACK packet to know if a packet is successfully received by the AP.

In case the relay successfully receives a packet, but the AP does not receive that packet (ACK_Timeout occurs), the relay stores that packet in its queue for transmission and sends an ACK packet over the channel to inform other nodes that the packet was received successfully. Upon receiving the relay's ACK packet, the node owning the packet will drop it from its queue and the delivery of that packet is now the relay's responsibility. Because of the relay's ACK packet, the node with a lost packet will not assume that the channel is congested and instead of doubling its contention window it will reset it to $CW_{\min}$.

The challenging part in the design of our cooperation protocol is how the relay gains access to the wireless channel without increasing the number of collisions and hence rendering its existence useless. To answer this question we propose the following relay channel access scheme;

- Following a transmission attempt from one or more source nodes (outcome of the transmission attempt is irrelevant), the relay node attempts to transmit the packet at the head of its queue immediately after the AP ACK, or after the ACK_Timeout

- For the relay not to be totally dependent on other nodes transmission attempts, the relay also maintains a single stage backoff counter with contention window size CW_r

- When the relay's backoff counter reaches zero, it will attempt to transmit the packet at the head of its queue like any other node

- Like any other node, the relay will invoke the backoff procedure after each transmission attempt in order not to capture the channel. The only difference is that the relay has a single backoff stage as opposed to m stages for other nodes

By accessing the channel after each transmission attempt on the channel, the relay has the ability to serve the packets in its queue without causing any collisions. Furthermore, the relay's single backoff stage guarantees that the relay will have access to the channel even if all other nodes' queues are empty (hence, no transmission attempts take place). Moreover, by controlling the relay's contention window CW_r, one can give the relay a higher channel access priority and also control the probability of having a collision between relay transmissions and other nodes transmissions.

5.3 System Model and Analysis

5.3.1 Channel model

We consider a Rayleigh fading channel model, where the signal received at the access point or the relay is modeled as

$$y_{ij} = \sqrt{G r_{ij}^{-\alpha}} h_{ij} x + n_j \tag{5.1}$$

where i is the source index, $j \in \{A, R\}$ is the access point or the relay index, x is the transmitted signal, G is the transmission power, assumed to be the same for all nodes, r_{ij} denotes the distance between source node i and its destination j, γ is the path loss exponent, h_{ij} the channel fading coefficients, modeled as zero-mean complex Gaussian random variables with unit variance, and n_j is an additive noise term at the destination, modeled as zero-mean complex Gaussian random variable with variance N_0. We assume that the channel coefficients are constant for the duration of the transmission of one packet. In this chapter, we will only consider the case of a symmetric network, where all the inter-users channels are assumed to be statistically identical.

Success and failure of packet reception is characterized by outage events and outage probabilities. As discussed in previous chapters, the probability of outage is given by,

$$P_{ij}^{out} = Pr\left\{ \mid h_{ij} \mid^2 < \frac{\beta N_0 r_{ij}^{\gamma}}{G} \right\} = 1 - \exp\left(-\frac{\beta N_0 r_{ij}^{\gamma}}{G} \right). \tag{5.2}$$

5.3.2 Markov Models

A number of models have been proposed in the literature to study the performance of IEEE 802.11 DCF in the saturated [86–88], unsaturated [89,90] traffic conditions, and in the presence of channel impairments [84,85]. To analyze the performance of our cooperative protocol, we start from the discrete-time Markov model for non-saturated sources developed in [91], and incorporate the channel effects and relay operation into the model. We consider two separate Markov chains, the first chain models source nodes while the second models the relay node.

We assume that the network consists of N contending nodes in addition to the relay node. Each node has an infinite length queue to store packets awaiting for transmission. Each node receives packets from upper layer based on a Poisson arrival process with arrival rate λ_s [1] packets/sec, and fixed packet size L. The queuing model used will be discussed in details in section 5.3.3

Source Nodes

Fig. 5.2 represents the discrete-time Markov chain used to model the operation of source nodes. Each node is modeled by a pair of integers (i, k). The backoff stage i, starts at 0 at the first attempt to transmit a packet and is increased by 1 every time a transmission attempt fails (either due to collision or channel fading error), up to a maximum value m. It is reset after a successful transmission. At any backoff stage $i \in [0, m]$, the backoff counter, k, is initially chosen uniformly between $[0, W_i^s - 1]$, where

$$W_i^s = 2^i W_0^s, \quad 0 \leq i \leq m, \tag{5.3}$$

is the range of the counter, and W_0^s is the parameter $\mathrm{CW}_{\mathrm{min}}$ specified in the IEEE 802.11 standard. The backoff counter is decremented by 1 in each idle time slot of duration σ, and the node transmits when the backoff counter $k = 0$.

[1] The super(sub)script s or r are used to differentiate between source node and relay parameters.

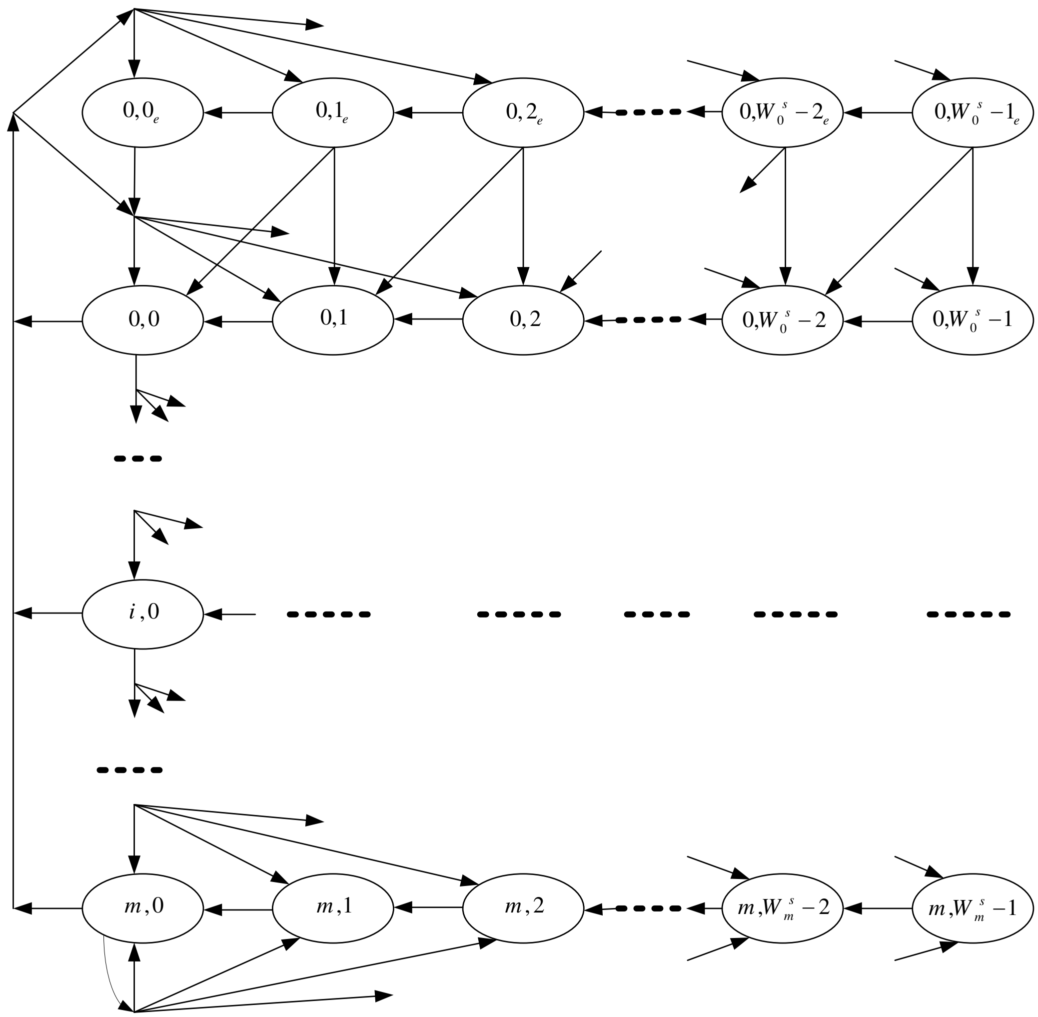

Figure 5.2: Source node's Markov model.

States $(0,k)_e$, $k \in [0, W_0^s - 1]$ are introduced to represent the state of the node when it has an empty queue after a successful transmission. Note that $i = 0$ in these states, because if $i > 0$ then a failed transmission should have occurred, so a packet must be waiting.

The fundamental assumption in our model is that, at each transmission attempt, and regardless of the number of retransmission suffered, each packet fails with a constant and independent probability P_f^s or P_f^r, for the source nodes or relay node, respectively [91], [86].

Let τ_s and τ_r be the probability that a source node or the relay transmit in a give slot, respectively. Now we are ready to write the Markov chain's transition probabilities, for $0 \leq i \leq m$

$$\begin{aligned}
&P\{(i,k)|(i,k+1)\} = P_i, && 0 \leq k \leq W_i^s - 2 \\
&P\{(0,k)|(i,0)\} = \frac{(1-q_s)(1-P_f^s)}{W_0^s}, && 0 \leq k \leq W_0^s - 1 \\
&P\{(0,k_e)|(i,0)\} = \frac{q_s(1-P_f^s)}{W_0^s}, && 0 \leq k \leq W_0^s - 1
\end{aligned} \tag{5.4}$$

where q_s is the probability that the node's queue is empty upon a departure (see section 5.3.3). P_i is the probability that the channel is sensed idle by the source node (i.e., all the remaining $N-1$ source nodes and the relay node are not attempting to transmit), and is given by $P_i = (1-\tau_s)^{N-1}(1-\tau_r)$.

P_f^s is the probability of a failed transmission attempt (either due to collision or channel fading), and is given by

$$P_f^s = 1 - (1-\tau_s)^{N-1}(1-\tau_r)((1-P_{sA}^{out}) + (1-P_{sR}^{out}) - (1-P_{sA}^{out})(1-P_{sR}^{out})), \tag{5.5}$$

which is the complement of the probability of the event that there is no collision, and that either the access point or the relay have correctly received the packet.

The first equation in (5.4) accounts for the fact that, at the beginning of each idle slot time, the backoff counter is decremented. The second and third equations account for the fact that following a successful packet transmission, the backoff stage i is reset to 0, and thus the backoff is initially uniformly chosen in the range $[0, W_0^s - 1]$.

In case of an unsuccessful transmission at backoff stage $i-1$, the backoff stage is increased, and the new initial backoff counter is initially chosen in the range $[0, W_i^s - 1]$. Once the backoff stage reaches the value m^s, it is not increased in subsequent packet transmissions. Then we have

$$\begin{aligned} P\{i, k|i-1, 0\} &= P_f^s / W_i^s \\ P\{m, k|m, 0\} &= P_f^s / W_m^s. \end{aligned} \quad (5.6)$$

Given that the node's queue is empty and the chain is in state $(0, k)_e$, in case of a packet arrival, the backoff counter is decremented and the chain makes a transition into the $(0, k-1)$ state; otherwise, the chain transits into $(0, (k-1)_e)$. When the backoff timer reaches zero, the node remains in state $(0, 0_e)$ as long as the queue is empty. If a packet arrives, then the node moves into state $(0, k)$, where k is uniformly chosen in the range $[0, W_0^s - 1]$. Therefore we have

$$\begin{aligned} P\{(0, k_e)|0, (k+1)_e\} &= P_i(1-a_i), && 0 \le k \le W_0^s - 2 \\ P\{(0, k)|(0, (k+1)_e)\} &= P_i a_i, && 0 \le k \le W_0^s - 2 \\ P\{(0, k)|(0, 0_e)\} &= (1-P_i)a_b, && \end{aligned} \quad (5.7)$$

where a_i and a_b are the probabilities of at least one packet arrival during an idle or a busy slot, respectively. From the Poisson arrival assumption, these probabilities ar given by

$$a_i = 1 - e^{-\lambda_s \sigma} \tag{5.8}$$

$$a_b = 1 - e^{-\lambda_s T_b}, \tag{5.9}$$

where σ is the idle slot duration, and T_b the busy slot duration. For simplicity we neglect the difference in durations between successful and unsuccessful transmission attempts. Typically, $\sigma = 20\mu s$, and $T_b = 2160.4\mu s$, based on 11 Mbps channel rate and packet size $L = 2312$ octets [82].

We now solve for the stationary distribution of this Markov chain. This will enable us to calculate different network performance measures. Let $\pi_s(i,k)$ denote the stationary probability of being in state (i,k). First, we will find expressions for all the stationary probabilities as a function of $\pi_s(0,0)$. Using balance equations, we have

$$\begin{aligned} \pi_s(i,0) &= P_f^s \pi_s(i-1,0), \quad 0 < i < m \\ (1 - P_f^s)\pi_s(m,0) &= \pi_s(m-1,0) P_f^s, \end{aligned} \tag{5.10}$$

which yields

$$\begin{aligned} \pi_s(i,0) &= (P_f^s)^i \pi_s(0,0) \quad 0 < i < m \\ \pi_s(m,0) &= \frac{(P_f^s)^m}{1 - P_f^s} \pi(0,0). \end{aligned} \tag{5.11}$$

It can be easily shown that, for $0 \le i \le m$ and $0 \le k \le W_i^s - 1$

$$\pi_s(i,k) = \frac{W_i^s - k}{P_i W_i^s} \pi_s(i,0). \tag{5.12}$$

Transitions into state $(0, (W_0^s - 1)_e)$ occur from state $i, 0$ if the node's queue is empty following a successful transmission, therefore

$$
\begin{aligned}
(P_i + (1 - P_i)a_b)\pi_s(0, (W_0^s - 1)_e) &= \frac{q_s(1 - P_f^s)}{W_0^s} \sum_{i=0}^{m} \pi_s(i, 0) \\
&= \pi_s(0, 0)\frac{q_s}{W_0^s}, \qquad (5.13)
\end{aligned}
$$

where we used the fact that

$$
\sum_{i=0}^{m} \pi_s(i, 0) = \frac{\pi_s(0, 0)}{(1 - P_f^s)}. \qquad (5.14)
$$

We then have for $0 < k < W_0^s - 1$,

$$
(P_i + (1 - P_i)a_b)\pi_s(0, k_e) = \frac{q_s(1 - P_f^s)}{W_0^s} \sum_{i=0}^{m} \pi_s(i, 0) + P_i(1 - a_i)\pi_s(0, k + 1)_e, \qquad (5.15)
$$

with $(P_i + (1 - P_i)a_b)$ on the left hand side replaced by $(P_i a_i + (1 - P_i)a_b))$ if $k = 0$. Straightforward recursion leads to the following expressions for $\pi_s(0, k)_e$ in terms of $\pi_s(0, 0)$,

$$
\pi_s(0, k)_e = \pi_s(0, 0)\frac{q_s}{W_0^s(P_i + (1 - P_i)a_b)} \frac{1 - c^{W_0^s - k}}{1 - c}, \qquad 0 < k \le W_0^s - 1 \qquad (5.16)
$$

$$
\pi_s(0, 0)_e = \pi_s(0, 0)\frac{q_s}{W_0^s(P_i a_i + (1 - P_i)a_b)} \left(1 + \frac{P_i(1 - a_i)}{P_i + (1 - P_i)a_b} \frac{1 - c^{W_0^s - 1}}{1 - c}\right), \qquad (5.17)
$$

where

$$
c = \frac{P_i(1 - a_i)}{P_i + (1 - P_i)a_b}.
$$

Finally, we express the probabilities $\pi_s(0, k)$ in terms of $\pi_s(0, 0)$. Using balance equations, we have

$$
\begin{aligned}
P_i\pi_s(0, k) =& \frac{(1 - P_f^s)(1 - q_s)}{W_0^s} \sum_{i=1}^{m} \pi_s(i, 0) + (1 - Pi)a_b\pi_s(0, k)_e + P_i a_i \pi(0, k + 1)_e \\
&+ P_i\pi(0, k + 1) + \frac{(1 - P_i)a_b + P_i a_i}{W_0^s}\pi(0, 0)_e, \qquad (5.18)
\end{aligned}
$$

from which we have the following generalization for $0 < k < W_0^s - 1$,

$$\pi_s(0,k) = \frac{W_0^s - k}{P_i W_0^s}\left[(1-q_s)\pi_s(0,0) + ((1-P_i)a_b + P_i a_i)\pi_s(0,0)_e\right] + \sum_{l=k+1}^{W_0^s-1}(P_i a_i + (1-P_i)a_b)\pi_s(0,l)_e + (1-P_i)a_b\pi_s(0,k)_e. \tag{5.19}$$

Through equations (5.11), (5.12), (5.16) and (5.19) all the steady state probabilities are expressed in terms of $\pi_s(0,0)$. Imposing the normalization condition

$$\sum_{i=0}^{m}\sum_{k=0}^{W_i^s-1}\pi_s(i,k) + \sum_{k=0}^{W_0^s-1}\pi_s(0,k)_e = 1, \tag{5.20}$$

we can calculate $\pi_s(0,0)$, hence, all the steady state probabilities.

Finally, since a node will make a transmission attempt in a given slot time if the Markov chain is in state $\pi_s(i,0)$ for $i \in [0,m]$, then, τ_s, the probability that a source node makes a transmission attempt in a given slot time can be expressed as

$$\tau_s = \sum_{i=0}^{m}\pi_s(i,0). \tag{5.21}$$

Relay Node

A relay node operating as described in section 5.2 will be modeled using the Markov chain model of Fig. 5.3. The model has a single backoff stage represented by states $k \in [0, W^r - 1]$. The backoff counter is uniformly chosen in that range, and the relay makes a transmission attempt when in state 0. The relay node makes a transition to state e if its queue becomes empty after a successful transmission. Finally, the chain is in state t when the relay is attempting to transmit following a busy channel.

Again we have the assumption that, at each transmission attempt, and regardless of the number of retransmission suffered, each packet fails with a constant

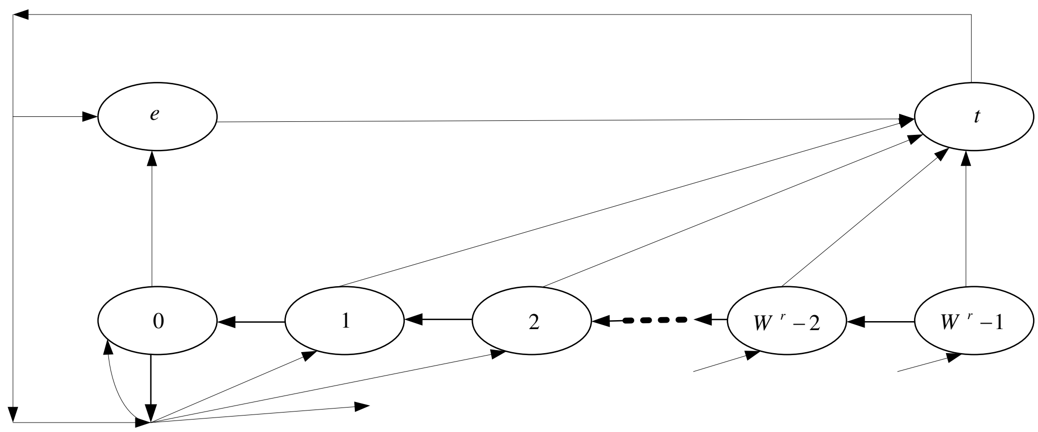

Figure 5.3: Relay node's Markov model.

and independent probability P_f^s or P_f^r, for the source nodes or relay node, respectively [91], [86].

Now we are ready to write the Markov chain's transition probabilities. At the beginning of each idle slot time, the backoff counter is decremented, so

$$P\{k|k+1\} = P_i, \quad 0 \le k \le W^r - 1 \tag{5.22}$$

where P_i is the probability that the channel is sensed idle by the relay node (i.e., all N source nodes are not attempting to transmit), and is given by $P_i = (1-\tau_s)^N$.

Since the relay Markov chain has a single backoff stage, following an unsuccessful transmission attempt or a successful attempt that leaves the relay queue non-empty, the backoff counter is initially uniformly chosen in the range $[0, W^r-1]$, so

$$\begin{aligned} P\{k|0\} &= \frac{(1-q_r)(1-P_f^r)}{W^r} + \frac{P_f^r}{W^r}, \qquad 0 \le k \le W^r - 1 \\ P\{k|t\} &= \frac{(1-q_r)(1-P_f'^r)}{W^r} + \frac{P_f'^r}{W^r}, \qquad 0 \le k \le W^r - 1 \end{aligned} \tag{5.23}$$

where q_r is the probability that a departing packet will leave the relay queue empty. P_f^r is the probability of a failed relay transmission attempt out of state

0 (failure due to collision or channel error), and P'^r_f is the probability of a failed transmission attempt out of state t (failure can be caused only by channel errors). These probabilities are given by

$$P^r_f = 1 - (1-\tau_s)^N (1 - P^{out}_{RA}) \tag{5.24}$$

$$P'^r_f = (1 - P^{out}_{RA}). \tag{5.25}$$

A successful transmission that leaves the relay queue empty leads to a transition to state e, so we have

$$\begin{aligned} P\{e|0\} &= q_s(1 - P^r_f) \\ P\{e|t\} &= q_s(1 - P'^r_f). \end{aligned} \tag{5.26}$$

Transitions into state t occur when the relay attempts to transmit a packet immediately after any transmission attempt on the channel, so

$$\begin{aligned} P\{t|e\} &= N\tau_s(1-\tau_s)^{N-1} P^{out}_{sAP}(1 - P^{out}_{sR}) = a \\ P\{t|k\} &= 1 - P_i, \quad 0 < k \le W^r \end{aligned} \tag{5.27}$$

The first equation accounts for the case when the relay queue is empty and that a source node transmission leads to a packet arrival at the relay, so that the relay will immediately forward that packet. The second equation is for the case when the relay queue is not empty; thus, the relay will transmit after a busy period regardless of the reception state (success or failure) of the packet transmitted during that period.

Let $\pi_r(k)$ denote the stationary probability of being in state (k). We now solve for the stationary distribution of this Markov chain. First, we will find expressions for all the stationary probabilities as a function of $\pi_r(0)$. Using balance equations, we have

$$\pi_r(k) = \frac{P_f^r + (1-P_f^r)(1-q_r)}{W^r}\pi_r(0) + \frac{P_f'^r + (1-P_f'^r)(1-q_r)}{W^r}\pi_r(t) + P_i\pi_r(k+1). \tag{5.28}$$

Straightforward recursion leads to the following expressions for $\pi_r(k)$ in terms of $\pi_r(0)$ and $\pi_r(t)$,

$$\pi_r(k) = \left[\frac{P_f^r + (1-P_f^r)(1-q_r)}{W^r}\pi_r(0) + \frac{P_f'^r + (1-P_f'^r)(1-q_r)}{W^r}\pi_r(t)\right]\frac{1-P_i^{W^r-k}}{1-P_i}, \quad 0 < k \leq W^r - 1 \tag{5.29}$$

Applying the balance equation on state e, we have

$$a\pi(e) = (1-P_f^r)q_r\pi_r(0) + (1-P_f'^r)q_r\pi_r(t). \tag{5.30}$$

Finally, for state t, we have

$$\pi_r(t) = a\pi_r(e) + (1-P_i)\sum_{k=0}^{W^r-1}\pi_r(k). \tag{5.31}$$

Substituting (5.29) and (5.30) in (5.31), and letting

$$y = \left[\frac{P_f^r + (1-P_f^r)(1-q_r)}{W^r}\pi_r(0) + \frac{P_f'^r + (1-P_f'^r)(1-q_r)}{W^r}\pi_r(t)\right]\sum_{k=0}^{W^r-1}\pi_r(k), \tag{5.32}$$

we have

$$\pi_r(t) = \frac{W^r q_r(1-P_f^r) + (1-P_i)y\left(P_f^r + (1-P_f^r)(1-q_r)\right)}{W^r - W^r q_r(1-P_f'^r) - (1-P_i)y\left(P_f'^r + (1-P_f'^r)(1-q_r)\right)}\pi_r(0). \tag{5.33}$$

Using equations (5.29), (5.30) and (5.33) all the steady state probabilities can be expressed in terms of $\pi_r(0)$. Imposing the normalization condition

$$\sum_{k=0}^{W^r-1}\pi_r(k) + \pi_r(e) + \pi_r(t) = 1, \tag{5.34}$$

we can calculate $\pi_r(0)$, hence, all the steady state probabilities.

Finally, the probability, τ_r, that the relay node makes a transmission attempt in a given slot is given by

$$\tau_r = \pi_r(0). \tag{5.35}$$

Note that transmissions out of state t are not included in this probability calculation, since these transmission events are not initiated by backoff counter timeout, and hence cannot result in collisions with source node transmissions.

5.3.3 Queuing Model

We assumed that each node receives packets from the upper layer based on a Poisson arrival process with arrival rate λ_s packets/sec, and fixed packet size L. Packet processing at each node can be seen as a single server with service rate μ_s, which depends on the channel access mechanism, the interaction between different nodes, and the channel statistics. Therefore, node queues can be modeled as M/G/1 queues.

The relay node can also be seen as a single server system. However, in the case of the relay, both the arrival rate, λ_r, and the service rate, μ_r, are dependent on the channel access mechanism, interaction between different nodes and the relay, channel statistics, and the number of source nodes and their arrival rates. For mathematical tractability, we will model the relay as an M/G/1 queuing system. Simulation results will later show that this is in fact a good approximation to the behavior of relay queue.

For any queueing system with single (in contrast to batch) arrivals and departures, the queue length seen by an arriving customer is equal to its length left by a departure customer. Furthermore, using the PASTA (Poisson Arrivals See

Time Averages) property associated with Poisson arrivals, the queue length at an arbitrary time equals the queue length seen by an arriving customer. This enables us to calculate the probability that a node queue is empty as follows [92]:

$$q_s = 1 - \frac{\lambda_s}{\mu_s}. \tag{5.36}$$

Similarly, for the relay queue,

$$q_r = 1 - \frac{\lambda_r}{\mu_r}. \tag{5.37}$$

These equations hold only when the queues are stable. For an irreducible and aperiodic Markov chain with countable number of states, the chain is stable if and only if there is a positive probability for the queue being empty, i.e., $\lim_{t\to\infty} \Pr\{Q_i(t) = 0\} > 0$. (As discussed in Chapter 3.) If the arrival and departure processes of a queuing system are strictly stationary, then one can apply Loynes' theorem to check for stability conditions [53]. This theorem states that, if the arrival and departure processes of a queuing system are strictly stationary, and the average arrival rate is less than the average departure rate, then the queue is stable; if the average arrival rate is greater than the average departure rate, then the queue is unstable.

In the following we will calculate the service and arrival rates for the different network nodes.

Source Nodes' Service Rate

The service time, S_s, of a packet is defined as the interval between the time the packet comes to the head of the transmission queue and the time the packet is acknowledged for correct reception (reception by either the AP or the relay node). We will use the probability generating function (PGF) [93] to characterize the discrete probability distribution of the service time, S_s.

Time spent in the backoff counter decrements constitutes the first component of a packet's service time. For the backoff counter to decrement by 1 (i.e., the Markov chain of Fig. 5.2 makes a transition from state (i, k) to state $(i, k-1)$, a node will, in general, spend j busy slots and a single idle slot (at which transition occurs) at each step of the counter. We should also note that, because of the way the relay accesses the channel, there will be two types of busy periods; (i) a period of duration T_b, if the relay does not access the channel immediately after a source node transmission. This occurs if the relay queue is empty, and the source node transmission does not result in a packet arrival at the relay (i.e., transmission resulted in a collision, a successful reception at the AP, or a failure to reach both the AP and the relay). From the point of view of the source node of interest, this event has the probability

$$P_{b1} = \pi_r(e)\left[1 - (1-\tau_s)^{N-1} - (N-1)\tau_s(1-\tau_s)^{N-2}P_{sA}^{out}(1-P_{sR}^{out})\right]. \tag{5.38}$$

(ii) a busy period of duration $2T_b$, if the relay accesses the channel immediately after a source node transmission (either the relay queue is not empty, or the source node transmission resulted in an arrival at the relay). A given source node will observe this event with probability

$$P_{b2} = \pi_r(e)(N-1)\tau_s(1-\tau_s)^{N-2}P_{sA}^{out}(1-P_{sR}^{out}) + \left[1-(1-\tau_s)^{N-1}\right]\sum_{k=1}^{W^r-1}\pi_r(k). \tag{5.39}$$

The PGF characterizing the distribution of the time spend at each counter step could now be written as

$$\begin{aligned} F(z) &= P_i z^{\sigma}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}(P_{b1}z^{T_b})^j(P_{b2}z^{2T_b})^{i-j} \\ &= \frac{P_i z^{\sigma}}{1-P_{b1}z^{T_b}-P_{b2}z^{2T_b}}. \end{aligned} \tag{5.40}$$

The next component is the time spent in a backoff stage i before making a transmission attempt (i.e., before the backoff counter reaches zero). At stage i the counter is initialized uniformly in the range $k \in [0, W_i^s - 1]$. Therefore the distribution of the time spent in stage i is characterized by the PGF,

$$F_i(z) = \sum_{k=0}^{W_i^s - 1} \frac{F^k(z)}{W_i^s}. \tag{5.41}$$

Finally, the PGF for the service time S_s can be written as

$$\begin{aligned} G_s(z) &= (1 - P_f^s) z^{T_b} \left[\sum_{i=0}^{m-1} (P_f^s z^{T_b})^i \prod_{j=0}^{i} F_j(z) + (P_f^s z^{T_b})^m \prod_{j=0}^{m} F_j(z) \sum_{i=0}^{\infty} (P_f^s z^{T_b})^i F_m^i(z) \right] \\ &= (1 - P_f^s) z^{T_b} \left[\sum_{i=0}^{m-1} (P_f^s z^{T_b})^i \prod_{j=0}^{i} F_j(z) + \frac{(P_f^s z^{T_b})^m \prod_{j=0}^{m} F_j(z)}{1 - P_f^s z^{T_b} F_m^i(z)} \right], \end{aligned} \tag{5.42}$$

which accounts for the busy slot in which the packet is successfully delivered, the possible number of failures a packet encounters (hence, the number of backoff stages it goes through), and finally, the amount of time probably spent at the maximum backoff stage m.

The service rate can then be calculated by differentiating $G_s(z)$ and setting $z = 1$,

$$\mu_s^{-1} = E[S_s] = \left. \frac{dG_s(z)}{dz} \right|_{z=1}. \tag{5.43}$$

Relay Node Arrival Rate

The time, A_r, between packet arrivals to the relay queue is composed of the following components; (i) idle periods in which no node (source or relay) is transmitting. These periods have a length σ and probability

$$P_i = (1 - \tau_s)^N (1 - \tau_r). \tag{5.44}$$

(ii) Busy periods of duration T_b, which occur if the relay queue is empty and the transmission attempt does not result in an arrival at the relay. This occurs with probability

$$P_{b1} = \pi_r(e)\left[1 - (1-\tau_s)^{N-1} - (N-1)\tau_s(1-\tau_s)^{N-2}P_{sA}^{out}(1-P_{sR}^{out})\right]. \qquad (5.45)$$

(iii) Busy periods of duration $2T_b$ not resulting in a relay arrival, which occur if the relay queue is not empty when a source node makes a transmission attempt. This has a probability

$$P_{b2} = \left[1 - (1-\tau_s)^{N-1} - (N-1)\tau_s(1-\tau_s)^{N-2}P_{sA}^{out}(1-P_{sR}^{out})\right]\sum_{k=1}^{W^r-1}\pi_r(k). \quad (5.46)$$

(iv) Finally, a busy period during which a packet enters the relay queue. This will always have a duration T_b, and has a probability

$$P_a = \left[\pi_r(e) + \sum_{k=1}^{W^r-1}\pi_r(k)\right](N-1)\tau_s(1-\tau_s)^{N-2}P_{sA}^{out}(1-P_{sR}^{out}). \qquad (5.47)$$

Given the above mentioned probabilities, we can write the PGF of A_r as follows:

$$G_a(z) = P_a z^{T_b}\sum_{i=1}^{\infty}\sum_{j=1}^{i}\sum_{k=0}^{i-j}\frac{i!}{j!k!(i-j-k)!}(P_i z^{\sigma})^j(P_{b1}z^{T_b})^k(P_{b2}z^{2T_b})^{i-j-k}. \quad (5.48)$$

The arrival rate can then be calculated by differentiating $G_a(z)$ and setting $z = 1$,

$$\lambda_r^{-1} = E[A_r] = \left.\frac{dG_a(z)}{dz}\right|_{z=1}. \qquad (5.49)$$

Relay Node Service Rate

Similar to the way the source nodes' service rate was calculated, we will start the calculation of the relay node service rate by defining the different components that constitute a packet's service time S_r. We note that, as opposed to source nodes, the relay can leave the backoff stage after any source node's transmission attempt on the channel (Fig. 5.3). Therefore, the time the packet at the head of the queue spends in the backoff stage can be split into two components; (i) the time before the backoff counter (initialized uniformly between 0 and $W^r - 1$) reaches 0, which in the relay case is composed only of idle slots. The PGF characterizing the distribution of that time is then given by

$$F_0 = \sum_{k=0}^{W^r-1} \frac{P_i^k z^{k\sigma}}{W^r}. \tag{5.50}$$

(ii) The time spent in the backoff stage before the Markov chain reaches state t, which is composed of a single busy period and a maximum of $W^r - 1$ idle slots. The PGF characterizing the distribution of that time is then given by

$$F_t = (1 - P_i) z^{T_b} \sum_{k=0}^{W^r-2} \frac{P_i^k z^{k\sigma}}{W^r}. \tag{5.51}$$

Finally, let

$$a = N\tau_n (1-\tau_n)^{N-1} P_{sA}^{out} (1 - P_{sR}^{out}), \tag{5.52}$$

be the probability that a packet enters relay queue. The PGF for the service time S_r can be written as

$$G_r(z) = \pi_r(e) a (1 - P_f'^r) z^{T_b} + \left(a P_f'^r z^{T_b} + \sum_{k=0}^{W^r-1} \pi_r(k) \right)$$
$$\times \left[\left(F_0(z)(1-P_f^r) z^{T_b} + F_t(z)(1 - P_f'^r) z^{T_b} \right) \sum_{i=0}^{\infty} \sum_{j=0}^{i} \binom{i}{j} \left(F_0(z) P_f^r z^{T_b} \right)^j \left(F_t(z) P_f'^r z^{T_b} \right)^{i-j} \right], \tag{5.53}$$

which accounts for the case when a packet is immediately served by the relay after it enters the queue (if queue was empty at packet arrival), the possible number of failures a packet encounters getting transmitted from either state 0 or state t, and finally, the periods at which the packet is delivered successfully.

The service rate can then be calculated by differentiating $G_r(z)$ and setting $z = 1$,

$$\mu_r^{-1} = E[S_r] = \left.\frac{dG_r(z)}{dz}\right|_{z=1}. \tag{5.54}$$

5.3.4 Iterative Numerical Solution

In the previous sections, a complete model for the network and cooperation protocol was developed. It was shown that there is an interdependence between the different model parameters; for instance, the two parameters τ_s and τ_r are calculated from the stationary distributions of the source and relay nodes' Markov chains, respectively. At the same time, the stationary distributions themselves depend on τ_s and τ_r through the transition probabilities. The same applies for the source and relay service rates and the relay arrival rate. Moreover, Markov chains' stationary distributions for the source and relay nodes are inherently interdependent because of the network operation and the above mentioned interdependencies.

To solve for the model parameters given the interdependencies mentioned above, we note that all parameters are expressed as functions of the stationary distributions of the source and relay nodes' Markov chains. We make use of the following property: for an ergodic irreducible Markov chain, let $\boldsymbol{\pi}$ be the stationary distribution vector, $\boldsymbol{\pi}_0$ an arbitrary initial vector, and $\mathbf{P}$ the transition probability matrix of the Markov chain. Then

$$\lim_{L\to\infty} \boldsymbol{\pi}_0 \times \underbrace{\boldsymbol{P} \times \boldsymbol{P} \times \cdots \times \boldsymbol{P}}_{L} = \boldsymbol{\pi}, \tag{5.55}$$

independent of the value of the initial vector $\boldsymbol{\pi}_0$.

Therefore, we solve our model using the following iterative approach;

1. Initialize the stationary distribution vectors $\boldsymbol{\pi}_s$ and $\boldsymbol{\pi}_r$ for each chain with some initial values
2. Calculate different model parameters
3. Update the stationary distribution vectors based on the calculated model parameters
4. goto step 2

Numerical results reveal that this approach has very good convergence rates.

5.3.5 Throughput

Let S be the normalized network throughput, defined as the fraction of time the channel is used to successfully transmit payload bits to the AP, which can be expressed as

$$S = \frac{P_s \cdot T_p}{T_s,} \tag{5.56}$$

where P_s is the probability of a successful transmission to the AP (by source or relay nodes), and T_p is the time to transmit the payload part of a packet. (Of course this is less than T_b, the total transmission time of a packet including the headers and the AP ACK packet.) T_s is the expected slot duration.

To calculate the probability P_s, we identify the events that result in a successful packet delivery to the AP, which are: (i) If the relay queue is empty, a source node successfully transmits a packet to the AP, or that packet fails to reach the AP but was successfully received and forwarded by the relay. This event has a probability

$$P_s^1 = \pi_r(e)\left[N\tau_s(1-\tau_s)^{N-1}\left((1-P_{sA}^{out}) + P_{sA}^{out}(1-P_{sR}^{out})(1-P_{RA}^{out})\right)\right]. \tag{5.57}$$

(ii) If the relay queue is not empty, a source node transmission fails to reach the AP (due to fading or collision), and the relay successfully transmits the packet at the head of its queue to the AP. This occurs with probability

$$P_s^2 = \left[1 - (1-\tau_s)^N - N\tau_s(1-\tau_s)^{N-1}(1-P_{sA}^{out})\right] \sum_{k=0}^{W^r} \pi_r(k). \tag{5.58}$$

(iii) If the relay queue is not empty and both source node transmission and the immediately following relay node transmission were successful. This occurs with probability

$$P_s^3 = N\tau_s(1-\tau_s)^{N-1}(1-P_{sA}^{out})(1-P_{RA}^{out}) \sum_{k=0}^{W^r} \pi_r(k). \tag{5.59}$$

(iv) The relay succeeds in transmitting a packet when its backoff counter reaches 0, which has a probability

$$P_s^4 = \tau_r(1-\tau_s)^N(1-P_{RA}^{out}). \tag{5.60}$$

Finally, the probability $P_s = P_s^1 + P_s^2 + 2P_s^3 + P_s^4$. The factor of 2 before P_s^3 accounts for the fact that the associated event results in the successful delivery of two packets to the AP.

The average length of a randomly chosen slot time is given by

$$\begin{aligned} T_s = (1-\tau_s)^N(1-\tau_r)\sigma + \big[\tau_r + \pi_r(e)\big(1-(1-\tau_s)^N \\ -N\tau_s(1-\tau_s)^{N-1}P_{sA}^{out}(1-P_{sR}^{out})\big)\big]\, T_b + 2\pi_r(t)T_b, \end{aligned} \tag{5.61}$$

which accounts for the idle slots, busy slots in which the relay transmits or a source node transmission is not followed by the relay transmission (when the relay queue is empty and no arrivals occur during source transmission), and busy slots in which a relay transmission follows source transmissions.

Based on an 11 Mbps transmission rate, and payload of length $L = 2312$ octets, typical slot durations are $\sigma = 20\mu s$, $T_p = 1681.5\mu s$, and $T_b = 2160.4\mu s$.

Delay

In our cooperation protocol, a packet can encounter two queuing delays; the first in the source node's queue and the second in the relay's queue. If a packet successfully transmitted by a source node goes to the AP, then this packet is not stored on the relay's queue. Let P_a denote the probability of this event. Then the total delay encountered by a packet can be modeled as

$$D = \begin{cases} D_s, & w.p.\ P_a \\ D_s + D_r, & w.p.\ 1 - P_a \end{cases} \tag{5.62}$$

where D_s and D_r are the queuing delays in the source and relay queues, respectively. We can elaborate more on (5.62) as follows. For a given packet in the source node's queue, if the first successful transmission for this packet is to the AP, then the delay encountered by this packet is only the queuing delay in the source node's queue. On the other hand, if the first successful transmission for this packet is not to the AP, then the packet will encounter a queuing delay in the source node's queue in addition to the queuing delay in the relay's queue.

First, we find the queuing delay in either the source node or the relay queue, as both are modeled as M/G/1 queues, with the difference being in the average arrival and departure rates. For an M/G/1 queue, the mean waiting time in queue is given by the Pollaczek-Kinchin formula [92],

$$\mathrm{E}[W_i] = \frac{\lambda_i \mathrm{E}[S_i^2]}{2(1 - \lambda_i/\mu_i)}, \tag{5.63}$$

where $i \in (s, r)$, λ_i is the average arrival rate, μ_i the average service rate, and S_i the service time. From the mean waiting time, one immediately gets the mean queuing delay as

$$D_i = \mathrm{E}[W_i] + \mathrm{E}[S_i]. \tag{5.64}$$

The probability P_a, that, for any packet, the first successful transmission form the source node's queue is to the AP is given by

$$P_a = \frac{1 - P_{sA}^{out}}{(1 - P_{sA}^{out}) + (1 - P_{sR}^{out}) - (1 - P_{sA}^{out})(1 - P_{sR}^{out})}. \tag{5.65}$$

From (5.62) and (5.65), the average delay is thus given by

$$D = D_s + (1 - P_a)D_r. \tag{5.66}$$

5.4 Results and discussions

We compare the performance of the cooperative protocol and the CSMA/CA protocol without cooperation. We set the SNR threshold $\beta = 15$ dB and the path loss exponent $\gamma = 3.7$. The distance between any node and AP is 120 m, and between any node and the relay 70 m, and between relay and AP 50 m. Transmission power is $100mW$, and noise variance $N_0 = 10^{-11}$. Source node's initial contention window $W_0^s = 32$ with $m = 5$ backoff stages, and relay node's contention window size $W_0^r = 32$. To validate the analytical model used of this chapter, we compare its result with an event-driven custom simulation program that we developed using Matlab. The simulator closely follows the 802.11 protocol details for each transmitting node as well as the relay.

Convergence behavior of the iterative solution of the model is demonstrated in Fig. 5.4, where we have plotted the difference in the value of $\pi_s(0,0)$ between iterations for a network with $N = 10$ nodes and arrival rate $\lambda_s = 25$ packets/sec. It is clear that the iterative approach converges to the model's solution in about 16 iterations for both the CSMA/CA and our cooperative protocol models. Since the cooperative model is more complex and involves the solution of two Markov

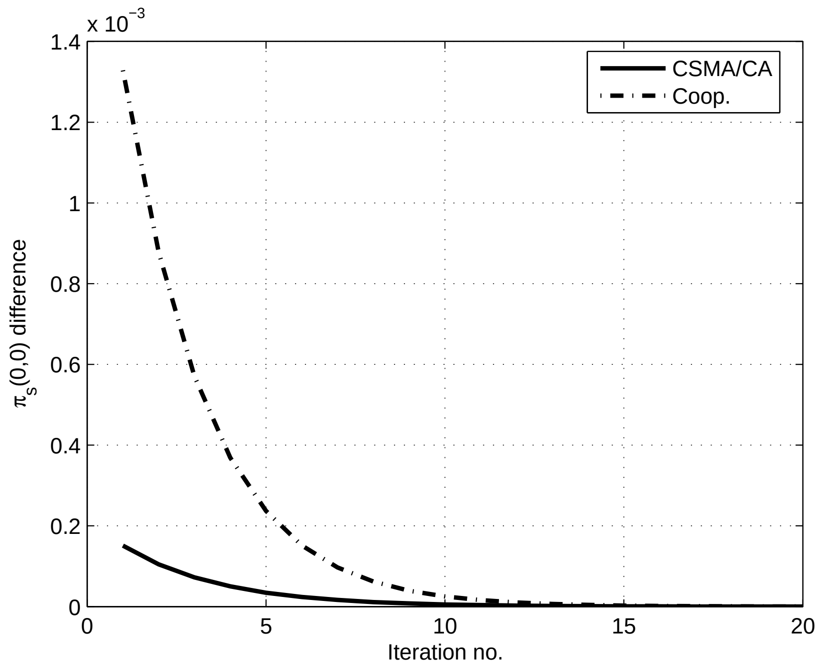

Figure 5.4: Difference in the value of $\pi_s(0,0)$ between iterations vs number of iterations.

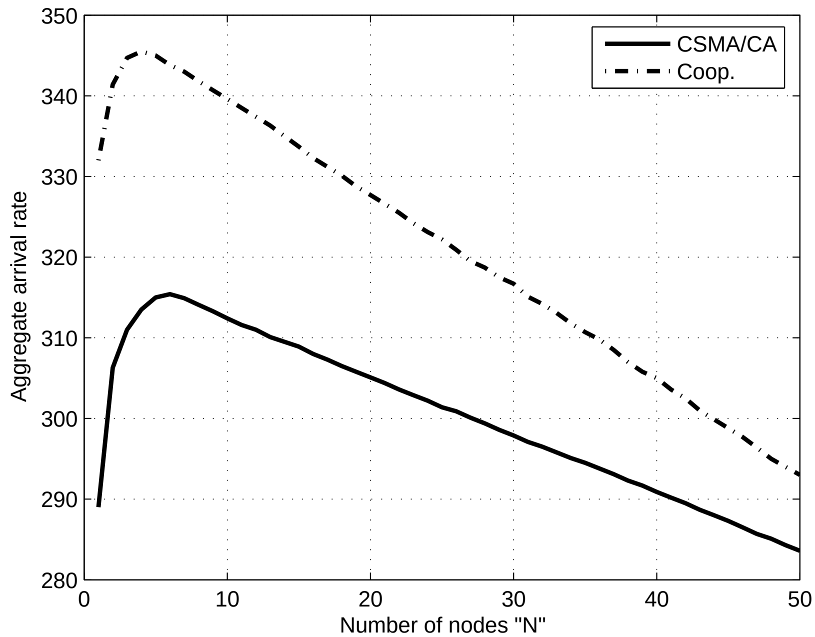

Figure 5.5: Maximum achievable aggregate arrival rate vs number network nodes.

chains, it takes slightly longer to converge than the CSMA/CA model without cooperation.

Fig. 5.5 depicts the maximum aggregate arrival rate supported by the network while maintaining queues stability versus the number of network nodes. We can observe that, for a given number of nodes, the proposed cooperative protocols resulted in a 7% average increase in the maximum supported aggregate arrival rate. This increase is due to the fact that the relay node provides a more reliable path to the AP leading to a higher packet delivery rate. Therefore, source nodes are able to empty their queues at a faster rate, thus freeing the channel for relay access and for additional nodes that the network might accommodate.

The normalized network throughput is depicted in Fig. 5.6 as a function of the number of network nodes for a fixed arrival rate of λ_s. The results show good

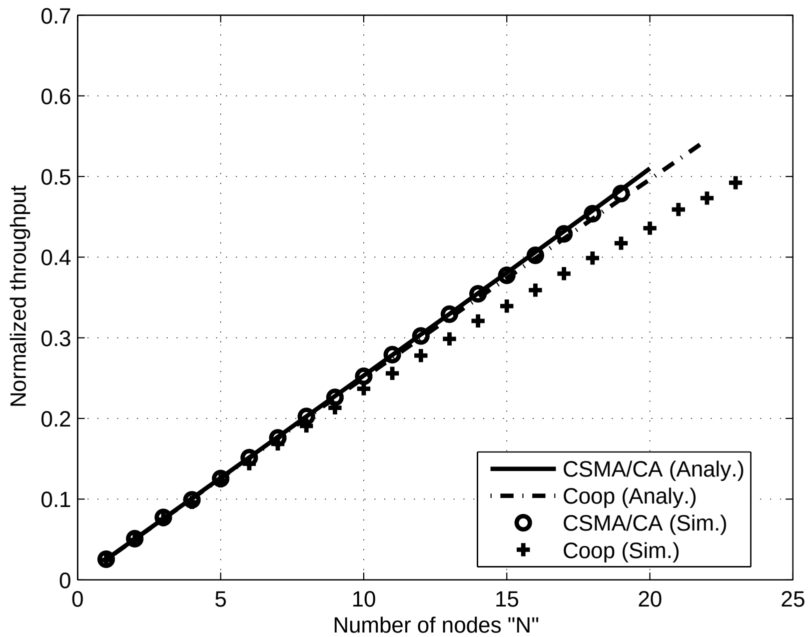

Figure 5.6: Normalized throughput vs. number of nodes for $\lambda_s = 15$.

match between the analytical model and simulation results. The results shown are under the condition that all the network queues are stable. It is noted that the network throughput is almost identical under both the cooperative and non-cooperative scenarios. This can be interpreted as follows; for a stable queue, it is well known that the average departure rate from the queue is equal to its average arrival rate. Therefore, for a given arrival rate, the average number of successfully transmitted packets to the AP will always be the same under both cooperative and non-cooperative protocols, as long as the queues are stable. On the other hand, the number of nodes supported by the network has increased by 10%, from 20 nodes in the non-cooperative case, to 22 nodes in the cooperative case.

Fig. 5.7 shows the delay performance of our cooperative protocol compared to the non-cooperative CSMA/CA protocol. It is clear that our protocol outperforms

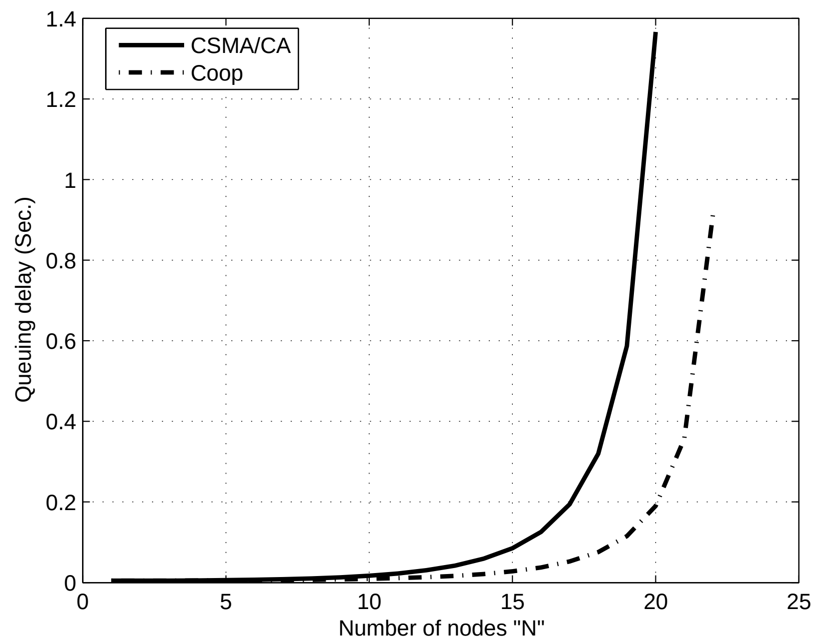

Figure 5.7: Queuing delay vs. number of nodes for $\lambda_s = 15$.

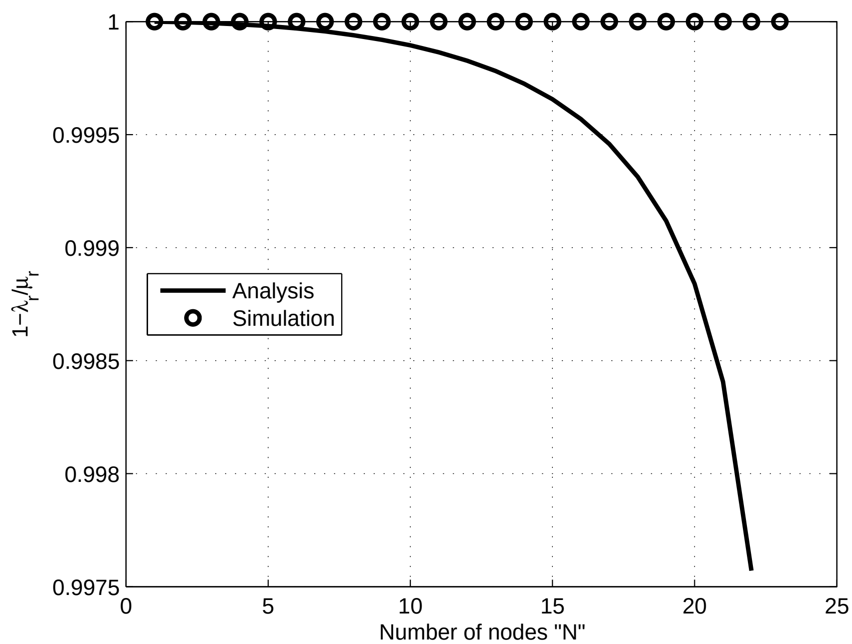

Figure 5.8: Probability that relay queue is empty vs. number of nodes for $\lambda_s = 15$.

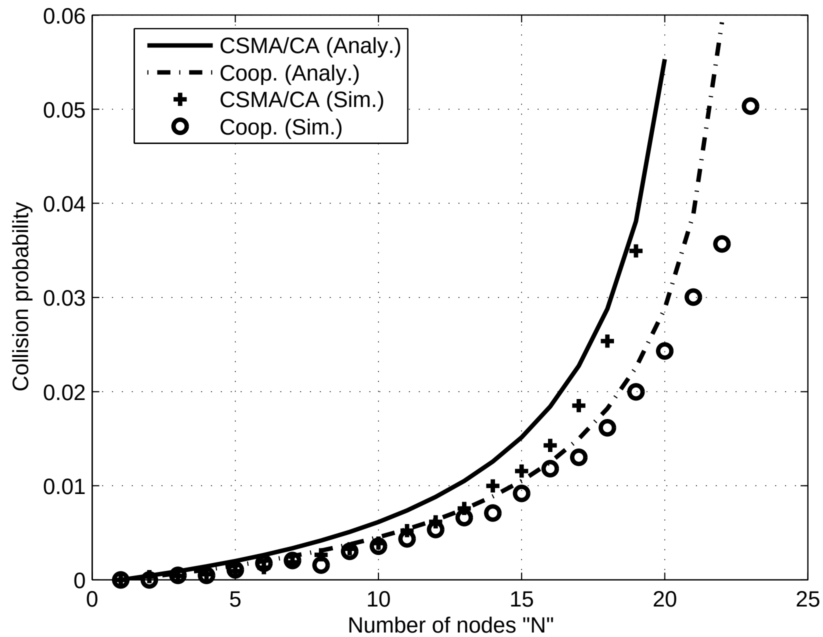

Figure 5.9: Collision probability vs. number of nodes for $\lambda_s = 15$.

CSMA/CA in terms of queuing delay. This is mainly because most of the relay's transmission attempts are made just after source nodes' transmissions, and not by waiting for the backoff counter to reach 0. Therefore, the relay is guaranteed a high degree of uncontested channel access. Moreover, as the network load increases, the average number of source nodes' transmission attempts increase, which offers the relay more channel access opportunities to service its queue that now has a higher arrival rate. To prove this, the quantity $(1 - \lambda_r/\mu_r)$, which from queuing theory is the probability that the relay queue is empty, is plotted in Fig. 5.8. It can be seen that there is less than 1% variation in the probability over the range of supported number of nodes. Much less variation is observed in the simulation results.

Fig. 5.9 compares between the collision probability of CSMA/CA and our cooperative protocol. The results show good match between the analytical model

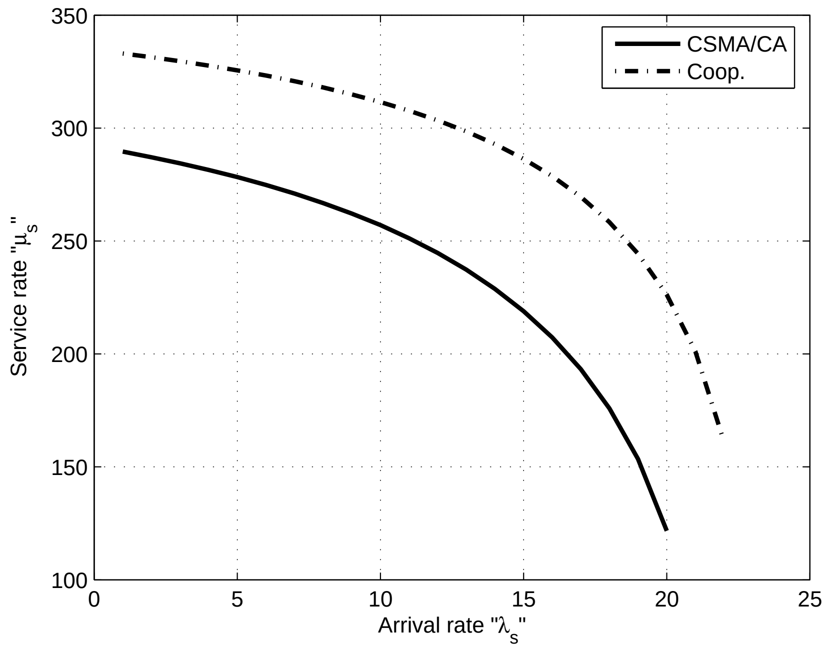

Figure 5.10: Source node's service rate vs. number of nodes for $\lambda_s = 15$.

and simulation results. Another merit of our cooperation protocol and its channel access mechanism is that, the introduction of the relay node in the network does not result in an increased collision probability as it is the case with any random access protocol. We further notice a decrease in the collision probability, which is because of the second path to the AP the relay offers to the network nodes. This second path helps the different nodes empty their queues at a faster rate, hence, nodes do not have to access the channel as often as in the case without cooperation, which reduces the collision probability. Finally, Fig. 5.10 demonstrates the effect of cooperation on how fast nodes' queues get empty by comparing the source node's service rates under both cooperative and non-cooperative protocols. An average increase of 28% is observed in the service rate, which interprets the reduction achieved in the collision probability.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, we have developed and analyzed cooperative communications protocol that have leveraged the benefits of cooperative diversity into different wireless networks MAC layers. In particular, we have developed multiple access cooperative protocols for speech networks, TDMA networks, and CSMA/CA random access networks. All the designed protocols share the ability to identify channel access opportunities and offer diversity to their respective networks without incurring any bandwidth efficiency losses. More specifically, we have addressed the following problems.

In Chapter 2, we have proposed a novel multiple access protocol for cooperative packet speech communications. Cooperation is implemented through a relay that efficiently helps active calls by using resources released by those users undergoing a period of silence in the speech conversations. Through cooperation, the proposed protocol addresses one important problem in some speech communications protocol, namely that wireless channel errors leads to active calls losing their channel reservation, which leads to an increase in medium access contention and a reduction in system capacity. A Markov model describing network dynamics was developed and analyzed in order to characterize our protocol's performance. Results revealed around 80% increase in throughput and a significant decrease in delay in the low SNR regime. The decrease in delay is translated into around 50% decrease in packet dropping probability, which in turn is translated into an improved speech quality.

In Chapter 3, we have tackled the problem of sharing idle channel resources in a TDMA network between cooperative relays and secondary cognitive radios. We have shown that idle time slots resulting from the bursty nature of the source nodes could be effectively used by relay nodes to offer diversity to source nodes. Based on the stability analysis of the network in presence of relays, two relay assignment schemes were developed. With these relay selection schemes, and through the help offered by relays, the TDMA network exhibits significant performance increase in terms of the maximum stable throughput. Next, the problem of sharing idle resource between cooperative relays and secondary cognitive radios was thoroughly investigated. Two different scenarios were considered. The first models the ideal case where relays and cognitive radio nodes are completely aware of each other's action, and hence no collisions between their transmissions could take place. The second scenario models the worst case where collisions between relays and secondary nodes transmission have the highest probability. Using queueing theory, and resorting to dominant system approaches, the stability regions of the primary and secondary networks are characterized for both scenarios. Results reveal that under both scenarios, both primary and secondary networks benefit from cooperation. Even under the worst case scenario, the gain to both networks due to cooperation outweighs the losses that might occur due to the collisions between relays and secondary nodes transmissions.

In Chapter 4, we investigated the question of how spectrum sensing errors affect the performance of a cognitive radio system from a MAC layer perspective. Analytical results reveal severe degradation in terms of throughput for both primary and secondary networks. The conclusion is drawn that separating the design of the spectrum sensing and the channel access mechanisms is suboptimal, and can have detrimental effects on the performance. Based on this conclusion, a joint design of spectrum sensing and channel access mechanisms is proposed and analyzed. The joint design made use of the fact that, in a binary hypothesis testing problem, the value of the test statistic could be used as a measure of how reliable the test outcome is. The proposed scheme bases the selection of the channel access probability on that reliability measure. Therefore, for a decision with higher reliability, the cognitive radio can access the channel with higher probability, and vice versa. Analytical results of the system's performance under the proposed scheme show significant improvements in terms of the throughput of both primary and secondary networks.

Finally, Chapter 5 answers the questions of how to enable cooperation in a random access network and whether the network will benefit cooperation or will the added relay transmissions will lead to an increase in collision probability and hence a decrease in performance. Those questions are answered through the design of a cooperative protocol that is able to identify specific instants in time in which relays can have uncontested access to the channel. Therefore, the network can benefit from cooperation without the risk of increased collision probability. Specifically, in a CSMA/CA based network, after each packet transmission, and irrespective of the outcome of this transmission, all network nodes will have to wait for a constant plus a random amount of time before making a transmission attempt. The proposed protocol enables cooperative relays to access the channel at those specific times where all other nodes defer from transmitting. A detailed analytic model that captures all the possible interactions between different nodes and cooperative relays was built. The model mainly consists of two coupled Markov models used to describe the operation of the relay and other network nodes. These Markov models are able together to completely describe the dynamics of the network and interactions between different nodes. Analytical and simulation results reveal significant gains in terms of throughput and delay performance. Furthermore, it is shown that, by virtue of the protocol design, collision probability has decreased rather than the expected increase due to extra transmissions on the channel.

6.2 Future Work: Multiple Relay Deployment in Single and Multi-Hop Random Access Networks

In Chapter 5 of this thesis we tackled the problem of enabling cooperation in CSMA/CA based random access networks. This work dealt only with the deployment of a single relay in a single-hop network, where all nodes have direct access to the access point.

Still, there exist a multitude of challenging questions regarding cooperation in random access networks that need to be answered. For instance; for the single-hop network with multiple relays

1. How can channel access be coordinated between relays in order to prevent

or minimize the risk of collisions between relay transmissions?

2. How should relays be assigned to different network nodes?

To answer the first question, we will note that if relays utilize the cooperation protocol of Chapter 5, they will attempt to transmit their packets at the end of each transmission attempt on the channel. Using this approach, we are faced with two additional problems. (i) We might run the risk of relays capturing the medium for long times. For example, in case of two relays only, after a transmission from the first relay, the second relay will detect that the channel became idle. According to our protocol, it will identify this as a transmission opportunity and start to transmit. This cycle of relays alternating transmissions might take arbitrarily long time to stop. However, this can be dealt with if relays access the channel with probability as in ALOHA, which will also reduce the number of possible collisions. (ii) The second problem is the risk of possible collisions between relay transmissions. As mentioned earlier this can be dealt with if relays have some probability with which they access the channel. A second possibility is that relays use a second backoff counter used solely for their special access mechanism we developed in Chapter 5.

To answer the second question, which involves relay selection, one might resort to similar relay selection techniques as the ones we proposed in Chapter 3. Of course the selection criteria will depend on the performance measure in question.

Looking at the muti-hop scenario, the list of unanswered questions is rather long. For instance;

1. Relay access to the channel will now be affected by the hidden and exposed nodes problem. Therefore, careful design of the channel access mechanism

is required to make sure relays will not eventually cause a degradation in channel performance.

2. Since in a multi-hop network each node might have different destination with whom it is communicating, relays can play a double role. On one hand they might provide help the same way as in single-hop network. On the other hand they might play a role in the routing process. The question is now how relays can balance between those two requirements and maintain their own queue stability.

3. Relay selection is now much more complex, since a single node might require help from different relays to reach different destinations.

All those are interesting questions that pose challenging research direction for the future.

BIBLIOGRAPHY

[1] J. G. Proakis, *Digital Communications*, McGraw-Hill Inc., 1994.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.

[3] G. J. Foschini and M. Gans, "On the Limits of Wireless Communication in a Fading Environment When Using Mutiple Antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, Mar. 1998.

[4] E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[5] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journ. Sel. Areas in Comm.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[6] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Trans. Info. Theory*, vol. 44, no. 2, pp. 744–765, March 1998.

[7] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-Time Block Coding for Wireless Communications: Performance Results," *IEEE Journ. Sel. Areas in Comm.*, vol. 17, no. 3, pp. 451460, Mar. 1999.

[8] A. Sendonaris, E. Erkip, and B. Aazhang, "Increaing Uplink Capacity via User Cooperation Diversity," in *IEEE Int. Symp. Info. Theory (ISIT)*, Aug. 1998.

[9] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity, Part I: System Description," *IEEE Trans. Comm.*, vol. 51, pp. 1927–1938, Nov. 2003.

[10] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity, Part II: Implementation Aspects and Performance Analysis," *IEEE Trans. Comm.*, vol. 51, pp. 1939–1948, Nov. 2003.

[11] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, "An Efficient Protocol for Realizing Cooperative Diversity in Wireless Networks," in *Proc. IEEE Int. Symp. Info. Theory (ISIT)*, Jun. 2001.

[12] J. N. Laneman, D. N. C. Tse, , and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Trans. Info. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[13] J. N. Laneman and G. W. Wornell, "Distributed Space-Time Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Trans. Info. Theory*, vol. 49, pp. 2415–2525, Oct. 2003.

[14] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Info. Theory*, vol. 25, no. 9, pp. 572–584, September 1979.

[15] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks," *IEEE Trans. Info. Theory*, Dec. 2005.

[16] A. K. Sadek, W. Su, and K. J. R. Liu, "Multinode Cooperative Communications in Wireless Networks," *IEEE Trans. Signal Proc.*, vol. 55, no. 1, pp. 341–355, Jan. 2007.

[17] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "Multihop diversity in wireless relaying channels," *IEEE Trans. Comm.*, vol. 52, no. 10, pp. 1820–1830, October 2004.

[18] W. Su, A. K. Sadek, and K. J. R. Liu, "Cooperative Communications in Wireless Networks: Performance Analysis and Optimum Power Allocation," *Wireless personal Communications*, vol. 44, no. 2, pp. 181–217, Jan. 2008.

[19] A. K. Sadek, K. J. R. Liu, and A. Epheremides, "Cognitive Multiple Access via Cooperation: Protocol Design and Performance Analysis," *IEEE Trans. Info. Theory*, vol. 53, no. 10, pp. 3677–3696, Oct. 2007.

[20] G. Scutari, S. Barbarossa, and D. Ludovici, "Cooperation Diversity in Multihop Wireless Networks Using Opportunistic Driven Multiple Access," in *Proc. IEEE Intl. Workshop on Signal Proc. Advances for Wireless Comm. (SPAWC)*, Rome, Italy, June 2003, pp. 170–174.

[21] R. Lin and A. P. Petropulu, "Cooperative Transmission for Random Access Wireless Networks," in *Proc. Intl. Conf. on Acoust. Speech and Sig. Proc.*, Philadelphia, PA, USA, March 2005, pp. 19–23.

[22] A. Ribeiro, N. D. Sidiropoulos, and G. B. Giannakis, "Achieving Wireline Random Access Throughput in Wireless Networking via User Cooperation," in *Proc. of IEEE Workshop on Signal Proc. Advances in Wireless Comm. (SPAWC)*, New York, NY, June 2005.

[23] M. K. Tsatsanis, R. Zhang, and S. Banerjee, "Network-Assisted Diversity for Random Access Wireless Networks," *IEEE Trans. Signal Proc.*, vol. 48, pp. 702–711, March 2000.

[24] C.-K. Lin. and Y.-W. Peter Hong, "On the Finite-User Stability Region of Slotted ALOHA with Cooperative Users," in *Proc. of the IEEE Intl. Conf. Comm. (ICC)*, Beijing, China, May 2008, pp. 1082–1086.

[25] Amr A. El-Sherif, A. Kwasinski, A. K. Sadek, and K. J. Ray Liu, "Content-Aware Cooperative Multiple Access Protocol for Packet Speech Communications," in *Proc. IEEE Global Telecom. Conf. (GLOBECOM07)*, Washington, DC, Nov. 2007, pp. 2952–2956.

[26] Amr A. El-Sherif, A. Kwasinski, A. K. Sadek, and K. J. Ray Liu, "Content-Aware Multiple Access Protocol for Cooperative Packet Speech Communications," *IEEE Trans. Wireless Comm.*, vol. 8, no. 2, pp. 995–1005, Feb. 2009.

[27] Amr A. El-Sherif, A. K. Sadek, and K. J. Ray Liu, "On Spectrum Sharing in Cooperative Multiple Access Networks," in *IEEE Global Telecom. Conf. (GLOBECOM08)*, Dec. 2008.

[28] Amr A. El-Sherif and K. J. Ray Liu, "Opportunistic Multiple Access for Cognitive Radio Networks," *IEEE Trans. on Communications, Under Review.*

[29] Amr A. El-Sherif and K. J. Ray Liu, "Joint Design of Spectrum Sensing and Channel Access," *In preparation.*

[30] Amr A. El-Sherif and K. J. Ray Liu, "Cooperation in Random Access Networks: Protocol Design and Performance Analysis," *In preparation.*

[31] D. Goodman and S. X. Wei, "Efficiency of Packet Reservation Multiple Access," *IEEE Trans. on Vehicular Technology*, vol. 40, no. 1, pp. 170–176, Feb. 1991.

[32] P. T. Brady, "A Model for On-Off Speech Patterns in Two-Way Conversation," *Bell System Technical Journal*, vol. 48, pp. 2445–2472, Sept. 1969.

[33] G. Wu, K. Mukumoto, and A. Fukuda, "Analysis of an Integrated Voice and Data Transmission System Using Packet Reservation Multiple Access," *IEEE Trans. on Vehicular Technology*, vol. 43, no. 2, pp. 289–297, May 1994.

[34] N. Benvenuto, S. Pupolin, and G. Guidotti, "Performance Evaluation of Multiple Access Spread Spectrum Systems in the Presence of Interference," *IEEE Trans. on Vehicular Technology*, vol. 37, no. 2, pp. 73–77, 1988.

[35] B. Hajek, A. Krishna, and R. O. LaMaire, “On the Capture Probability for a Large Number of Stations,” *IEEE Trans. Comm.*, vol. 45, no. 2, pp. 254–260, Feb. 1997.

[36] M. Zorzi and R. Rao, “Capture and Retransmission Control in Mobile Radio,” *IEEE JSAC*, vol. 12, pp. 1289–1298, Oct. 1994.

[37] M. M. Khan and D. Goodman, “Effects of Channel Impairments on Packet Reservation Multiple Access,” in *Proc. 44th IEEE Veh. Tech. Conf.*, Stockholm, Sweeden, June 1994, pp. 1218–1222.

[38] S. Nanda, D. J. Goodman, and U. Timor, “Performance of PRMA: A Packet Voice Protocol for Cellular Systems,” *IEEE Trans. on Vehicular Technology*, vol. 40, no. 3, pp. 584–598, Aug. 1991.

[39] H. Qi and R. Wyrwas, “Performance Analysis of Joint Voice-Data PRMA over Random Packet Error Channels,” *IEEE Trans. on Vehicular Technology*, vol. 45, no. 2, pp. 332–345, May 1996.

[40] L. Sun and E. C. Ifeachor, “Voice Quality Prediction Models and Their Application in VoIP Networks,” *IEEE Trans. on Multimedia*, vol. 8, no. 4, pp. 809–820, Aug. 2006.

[41] International Telecommunication Union, *Methods for Subjective Determination of Transmission Quality*, 1996.

[42] ITU-T, “Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” 2001.

[43] ETSI/GSM, ,” .

[44] ITU-T, “Recommendation G.107: The E-model, A Computational Model for Use in Transmission Planning,” 2003.

[45] S. Haykin, “Cognitive Radio: Brain-Empowered Wireless Communications,” *IEEE Journ. Sel. Areas in Comm.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[46] B. Tsybakov and W. Mikhailov, “Ergodicity of Slotted ALOHA,” *Probl. Predachii Info.*, vol. 15, pp. 73–87, 1979.

[47] R. Rao and A. Ephremides, “On the Stability of Interacting Queues in a Multi-Access System,” *IEEE Trans. Info. Theory*, vol. 34, pp. 918–930, Sept. 1988.

[48] W. Szpankowski, "Stability Conditions for Some Multiqueue Distributed System: Buffered Random Access Systems," *Adv. Appl. Probab.*, vol. 26, pp. 498–515, 1994.

[49] W. Luo and A. Ephremides, "Stability of N Interacting Queues in Random-Access Systems," *IEEE Trans. Info. Theory*, vol. 45, no. 5, pp. 1579–1587, July 1999.

[50] V. Naware, G. Mergen, and L. Tong, "Stability and Delay of Finite-User Slotted ALOHA with Multipacket Reception," *IEEE Trans. Info. Theory*, vol. 51, no. 7, pp. 2636–2656, Jul. 2005.

[51] J. Luo and A. Ephremmides, "On the Throughput, Capacity and Stability Regions of Random Multiple Access," *IEEE Trans. Info. Theory*, vol. 52, no. 6, pp. 2593–2607, Jun. 2006.

[52] M. Kobayashi, G. Caire, and D. Gesbert, "Impact of Multiple Transmit Antennas in a Queued SDMS/TDMA Downlink," in *Proc. of 6th IEEE Workshop on Signal Proc. Advances in Wireless Comm. (SPAWC)*, New York, NY, June 2005.

[53] R. M. Loynes, "The Stability of a Queue with Non-Independent Interarrival and Service Times," *Proc. Cambridge Philos. Soc.*, pp. 497–520, 1962.

[54] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation Issues in Spectrum Sensing for Cognitive Radio," in *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, Pacific Grove, CA, 2004, pp. 772–776.

[55] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, 2nd edition, 1992.

[56] A. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-Based Wireless Location," *IEEE Signal Proc. Mag.*, vol. 22, no. 4, pp. 24–40, Jul. 2005.

[57] G. Sun, J. Chen, W. Guo, and K. J. Ray Liu, "Signal Processing Techniques in Network-Aided Positioning: A Survey," *IEEE Signal Proc. Mag.*, vol. 22, no. 4, pp. 12–23, Jul. 2005.

[58] P. Papadimitratos, S. Sankaranarayanan, and A. Mishra, "A Bandwidth Sharing Approach to Improve Licensed Spectrum Utilization," *IEEE Comm. Mag.*, vol. 43, pp. 10–14, Dec. 2005.

[59] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework," *IEEE Journ. Selected Areas in Comm.*, vol. 25, pp. 589600, Apr. 2007.

[60] Y. Chen, Q. Zhao, and A. Swami, "Distributed Cognitive MAC for Energy-Constrained Opportunistic Spectrum Access," in *Proc. MILCOM*, Washington, D.C., Oct. 2006, pp. 1–7.

[61] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Optimal Dynamic Spectrum Access via Periodic Channel Sensing," in *Proc. WCNC*, Mar. 2007, pp. 33–37.

[62] L. Cao and H. Zheng, "Distributed Spectrum Allocation via Local Bargaining," in *Proc. IEEE Conf. on Sensor and Ad Hoc Commu. and Networks (SECON)*, Santa Clara, CA, Sep. 2005.

[63] J. Huang, R. A. Berry, and M. L. Honig, "Auction-Based Spectrum Sharing," *ACM/Springer Mobile Networks and Applications Journal (MONET)*, vol. 11, no. 3, pp. 405–418, Jun. 2006.

[64] H. Zheng and C. Peng, "Collaboration and Fairness in Opportunistic Spectrum Access," in *Proc. IEEE Intl. Conf. Comm. (ICC)*, May 2005.

[65] R. Etkin, A. Parekh, and D. Tse, "Spectrum Sharing for Unlicensed Bands," in *Proc. IEEE DySPAN 2005*.

[66] Z. Ji and K. J. R. Liu, "Dynamic Pricing Approach for Spectrum Allocation in Wireless Networks with Selfish Users," in *Proc. IEEE Globecom*, San Francisco, CA, Nov. 2006.

[67] Z. Ji and K. J. R. Liu, "Belief-Assisted Pricing for Dynamic Spectrum Allocation in Wireless Networks with Selfish Users," in *Proc. IEEE SECON 2006*, Reston, VA, Sep. 2006, pp. 119–127.

[68] Y. Xing, R. Chandramouli, S. Mangold, and S. Shankar N, "Dynamic Spectrum Access in Open Spectrum Wireless Networks," *IEEE Journ. Sel. Areas in Comm.*, vol. 24, no. 3, pp. 626–637, Mar. 2006.

[69] S. Keshavamurthy and K. Chandra, "Multiplexing Analysis for Spectrum Sharing," in *Proc. IEEE MILCOM06*, Washington, D.C., Oct. 2006.

[70] A. Sahai, N. Hoven, and R. Tandra, "Some Fundamental Limits on Cognitive Radio," in *Proc. of Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, Oct. 2004.

[71] K. Challapali, S. Mangold, and Z. Zhong, "Spectrum Agile Radio: Detecting Spectrum Opportunities," in *Proc. Int. Symp. Advanced Radio Technologies*, Boulder, CO, Mar. 2004.

[72] B. Wild and K. Ramchandran, "Detecting Primary Receivers for Cognitive Radio Applications," in *Proc. 1st IEEE Symp. New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, pp. 124–130.

[73] A. Ghasemi and E. Sousa, "Collaborative Spectrum Sensing for Opportunistic Access in Fading Environments," in *Proc. 1st IEEE Symp. New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, p. 131136.

[74] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice Hall, Upper Saddle River, NJ, 1998.

[75] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1994.

[76] "A Cyclostationary Feature Detector," in *Proc. 28th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 1994.

[77] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative Sensing Among Cognitive Radio," in *Proc. IEEE ICC*, 2006, pp. 1658–1663.

[78] E. Vistotsky, S. Kuffner, and R. Peterson, "On Collaborative Detection of TV Transmissions in Support of Dynamic Spectrum Sharing," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Baltimore, MD, Nov. 2005.

[79] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed Detection with Multiple Sensors: Part II - Advanced Topics," *Proc. IEEE*, vol. 85, no. 1, pp. 6479, Jan. 1997.

[80] G. Ghurumuruhan and Y. Li, "Agility Improvement Through Cooperative Diversity in Cognitive Radio," in *Proc. IEEE Global Commun. Conf.*, St. Louis, MO, Nov. 2005.

[81] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 2003.

[82] IEEE Computer Society LAN MAN Standards Committee, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11-1999, New York, NY: IEEE, 1999.

[83] P. Chatzimisios, A. C. Boucouvalas, and V. Vitsas, "Performance Analysis of IEEE 802.11 DCF in Presence of Transmission Errors," in *Proc. Intl. Conf. Comm. (ICC)*, June 2004, pp. 3854–3858.

[84] Zoran Hadzi-Velkov and Boris Spasenovski, "On The Capacity of IEEE 802.11 DCF with Capture in Multipath-Faded Channels," in *Proc. Intl. Conf. Comm. (ICC)*, Jul. 2003, pp. 121–126.

[85] Y. Zheng, K. Lu, D. Wu, and Y. Fang, "Performance Analysis of IEEE 802.11 DCF in Imperfect Channels," vol. 55, no. 5, pp. 1648–1656, Sep. 2006.

[86] G. Bianchi, "Performance Analysis of The IEEE 802.11 Distributed Coordination Function," vol. 18, no. 3, pp. 535–547, Mar. 2000.

[87] Y. Tay and K. Chua, "A Capacity Analysis for the IEEE 802.11 MAC Protocol," vol. 7, no. 2, pp. 159–171, Mar. 2001.

[88] Y. Xiao, "Saturation Performance Metrics of the 802.11 MAC," in *Proc. IEEE Veh. Technol. Conf.*, Oct. 2003, vol. 3, pp. 1453–1457.

[89] K. Duffy, D. Malone, and D. J. Leith, "Modelling of 802.11 Distributed Coordination Function in Non-Saturated Condition," vol. 9, no. 8, pp. 715–718, Aug. 2005.

[90] M. Garetto and C. F. Chiasserini, "Performance Analysis of WLANs under Sporadic Traffic," in *Proc. Networking*, New York: Springer-Verlag, May 2005, vol. 3462, pp. 1343–1347.

[91] D. Malone, K. Duffy, and D. Leith, "Modelling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions," vol. 15, no. 1, pp. 159–172, Feb. 2007.

[92] Ronald W. Wolff, *Stochastic Modeling and The Theory of Queues*, Prentice Hall, 1989.

[93] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, Oxford University Press, 2001.