BACKPRESSURE NETWORKS WITH COOPERATIVE LINK AUGMENTATION

BY

ULYSSES LEE

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Adviser:

Professor Andrew C. Singer

ABSTRACT

We investigate a cross-layer communication technique which jointly leverages diversity gain from cooperative communications relaying and optimal throughput characteristics of backpressure networking. In particular, we address the capacity limitations of backpressure networks within fading environments by the retasking of a cooperating node as a relay with potentially heterogeneous transmission architecture. Retasking a node as a cooperative relay can temporally allocate resources of one particular session within the backpressure network onto another session, thereby allowing for a more flexible physical (PHY) layer for traffic load balancing. With this, we derive a scheduling method that ensures timely delivery of information in networks without predetermined infrastructure.

Within this thesis, we propose the architecture of an amplify-and-forward relay for cooperative communications and derive the performance of multiple cooperative nodes utilizing this architecture. We also propose a suitable medium access control (MAC) layer facilitating the scheduling and decision making of nearby relay node candidates. The proposed architecture may be of potential interest for emerging device-to-device (D2D) and swarming mesh networks.

To my friends and family, who have never given up on me in even the most trying of times, and to Todd, who supported and encouraged me to achieve the fullest of my potential.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis addresses work in the use of cooperative communications at the physical (PHY) layer protocol in high throughput mobile ad-hoc networks (MANETs). In particular, we consider a slow fading environment heavily loaded with multiple sessions of traffic. This type of scenario is common when considering wireless data links over vehicular communications in airborne or ground-based cluttered urban environments. The design and implementation of such a protocol may provide insight for the development of future airborne swarm communications and other commercial use.

1.1 Motivation

In the 2010s, both the commercial and personal use of unmanned aerial systems (UAS) has increased in the United States. By 2013, the Federal Aviation Administration (FAA) had announced a selection of six test sites tasked to conduct studies on the feasibility of UASs for commercial applications. The same year, Google announced Project Loon, an ambitious project to provide connectivity and data worldwide through the use of inexpensive but intricately designed balloons with base station capabilities. Amazon announced its intentions with Amazon Prime Air, an autonomous flight delivery system that can allow customers to receive their packages in under an hour.

Air-to-air wireless communications and networking is still a widely investigated and open problem in terms of reliable high-throughput data delivery. Due to the impracticality of using reliable communication methods through excessively long wires between moving or airborne nodes, nodes must rely on high power and advanced physical layer techniques to compensate for their less advantageous wireless transmission medias. Furthermore, the uncertainty of the exact location of nodes and event-based policies such as air

space deconfliction suggest that airborne nodes cannot rely on known infrastructure, such as cellular base stations or ground relay points, as supporting backbones for heavy data traffic applications. Cognitive radio, a rapidly autonomously reconfigurable radio that can exploit available wireless spectrum based on its situational awareness, is a flexible candidate technology in alleviating current problems in airborne wireless communications.

This work in particular addresses amplify-and-forward cooperative diversity in backpressure MANETs. Cooperative diversity is a PHY and media access control (MAC) layer (Open Systems Interconnection Layers 1 and 2) transmission technique that utilizes spatially separated cooperating nodes to provide redundant paths in a wireless fading environment. The bit error rate (BER) and outage probability of source-to-destination wireless transmissions improve dramatically as the number of nodes participating increases. Backpressure routing is a dynamic routing algorithm that routes information by prioritizing nodes with maximized difference in locally queued traffic. It can be shown that networks of nodes that follow a distributed control rule prioritizing backpressure also follow an optimal throughput policy that can support the maximum set of inbound network data traffic [1, Tassiulas and Ephremedes]. Although cooperative diversity is treated as a PHY-MAC layer protocol and backpressure routing algorithms have been investigated as cross-layer situationally aware network protocols, the techniques cannot be coupled in a modular fashion. Specifically, MAC protocols that use cooperative diversity in the literature do not consider the current status of the network, nor do they address the possibility of assigning relay nodes for link augmentation as needed by a network undergoing heavy traffic. To our knowledge, there is no framework that investigates the benefits of cooperative diversity in backpressure networks, which we investigate in this thesis.

1.2 Background

1.2.1 Cooperative Communications

The goal of modern wireless communication involves relocating information from point A (which we will refer to as a *transmitter* or *source*) to point B (which we will refer to as a *receiver* or *destination*) without the use of a solid

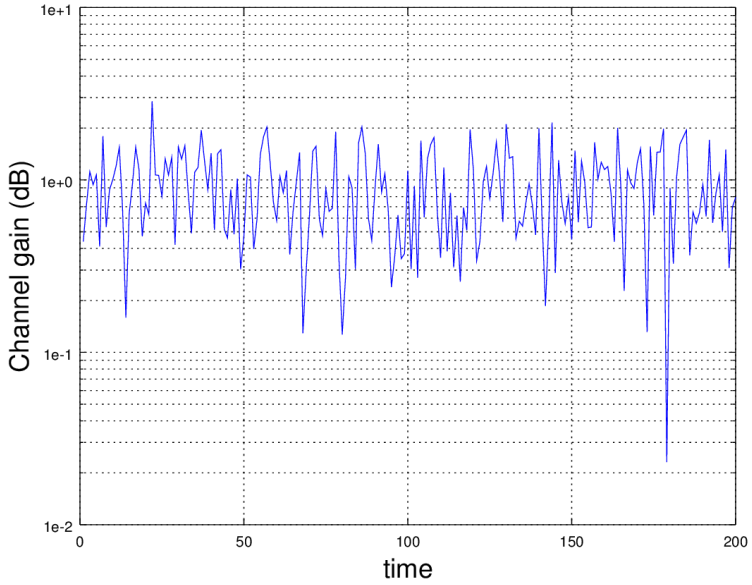

Figure 1.1: Instantaneous Channel Gain Under Rayleigh Fading Conditions

physical medium. Common means of transmitting this signal involve propagating an electromagnetic wave or optical beam in a pattern agreed on by the source and destination a priori. Unfortunately, environment variables such as scatterers, blockages, and interferers may constructively or destructively interfere with signal transmissions. Communication engineers characterize for the cumulative effect of these environment variables as a random event, known as *fading*. Fading is mathematically integrated into the design models of engineers as a random phenomenon affecting the gain and linearity properties of a wireless signal.

An example of the instantaneous channel gain under Rayleigh fading conditions is shown in Figure 1.1. At times, the channel may experience deep fade (strong destructive interference leading to a sharp drop in gain magnitude), resulting in failure and disruption of communications. The quality of communication links experiencing effects from fading channels can be quantified with metrics such as BER and outage probability (probability that the channel gain drops below a threshold value necessary to maintain information exchange at a minimal requirement). One approach in the design of wireless communication systems is to focus on the reduction of these quantities.

One of the most common methods to mitigate the effects of fading is by the use of diversity, or sending information redundantly over different channels in

either space, frequency or time. Spatial diversity treats fading by transmitting or receiving from multiple spatially separated antennas. If the antennas are sufficiently separated, redundant signals will traverse different paths and the probability of multiple signal paths simultaneously experiencing deep fade diminishes with increasing number of antennas. Frequency diversity treats fading by transmitting over multiple or a wider set of frequencies. If some frequencies selected do not experience fading during transmission, the original message can be recovered by message fragments that were received unobstructed. Frequency diversity works exceptionally well in narrow-band interference and frequency-selective fading conditions. Common applications of frequency diversity include sub-carrier interleaving (like in orthogonal frequency division multiplexing) and spread spectrum techniques. Time diversity treats fading by sending redundant signals at different times, or by interleaving (randomizing message bits with a known pseudo-random sequence). By interleaving, the error effects of deep fade are spread over different message blocks and are not significant enough to lose entire messages. These principles are commonly used in today's multi-input-multi-output (MIMO) communication systems.

In order for diversity techniques to be effective, the channel in which the redundant signal copies traverse should be independent (or satisfy some "rich scattering"-like environment). For airborne vehicular communications, independence of paths for spatial diversity typically becomes an invalid assumption because nodes are far from each other (because of air space deconfliction), and signal paths become strongly correlated. Consider a situation where there are 2 airborne nodes separated by hundreds of miles. Each airborne node is equipped with multiple antennas to facilitate spatial diversity. From the local perspective of any antenna on the receiving node, all of the antennas on the transmitting node seem almost coincident . Consequently, all of the signal paths will undergo the same path attenuation and fading effects. The receiver would obtain multiple near-identical copies of a faded signal rather than multiple diversity copies of a faded signal and not be able to exploit the full diversity gain of all elements. Cooperative diversity has been proposed as a practical solution to provide spatial diversity in these situations.

Cooperative diversity is the concept that a node can temporarily act as an antenna element for a source transmission. A source transmission is first

broadcast to the physical medium, with the destination receiving a copy of this signal. At a later stage, nearby nodes that received a copy of the broadcast forward redundant copies to the destination as well. Upon reception of the multiple copies at the destination, the receiver coherently combines information symbols using weights determined by the quality of signal paths before decoding the intended message bits. This method differs from network layer relaying techniques as the destination receives multiple symbol-level copies of the signal and does not designate specific routing paths for the end-to-end delivery of messages. For this reason, cooperative diversity has also been considered virtual MIMO or beyond line-of-sight MIMO.

Common methods of cooperative communications involve decode-and-forward or amplify-and-forward. As the name implies, decode-and-forward relays require common transceiver architecture to decode relayed messages, which may have unintended security consequences and potential for error propagation upon incorrect decoding of messages. Contrarily, amplify-and-forward relays only require transceiver architecture to receive samples, amplify and retransmit relayed messages. This makes amplify-and-forward relaying an excellent candidate for heterogeneous node network design. In this thesis, we consider amplify-and-forward relaying as a means for cooperation.

1.2.2 Backpressure Routing

In some scenarios, wireless point-to-point communications subject to resource and physical constraints may not be a feasible or practical solution to deliver messages. The natural extension of point-to-point communication links considers *networking* for the end-to-end delivery of information. Networking, as a science, was developed to handle multi-user, multi-commodity end-to-end traffic delivery when point-to-point delivery is infeasible. Communication networks, which emphasize the task of end-to-end data delivery, have historically been viewed as a set of modular problems by models such as the Open Systems Interconnection (OSI) 7-layered network protocol stack shown in Figure 1.2. Each layer of the protocol stack plays a role in information delivery independent of the other layers, and layers are separated with the encapsulation/decapsulation of messages (appending the output data structure of a layer to a lower layer with header bits of meta-information and removing

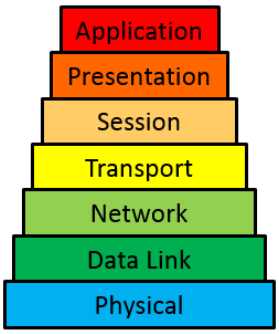

Figure 1.2: OSI 7 Layer Protocol Stack

header bits when messages are passed to a higher layer, respectively). The functional responsibilities of each layer are as follows:

- **Application**: End-user host protocols and interfaces
- **Presentation**: Translation and encoding of data between application and networking services
- **Session**: Management of continuous exchanges of information between multiple nodes
- **Transport**: Reliability and ordering of end-to-end message deliveries
- **Network**: Management of routes and addressing of different nodes
- **Data Link**: Reliability and scheduling of point-to-point messaging
- **Physical**: Transmission and reception of raw messages over physical media

One of the problems in network design is the establishment of end-to-end routes. In a route, messages are sent from one node to another (designated as *hops*) until they reach their intended destination. As the geographic span of a network increases, identifying the best routes for data delivery becomes more difficult. Furthermore, wireless networks must consider the delivery of messages with unreliable links subject to fading and outages. *Queueing networks* provide a mathematical and protocol framework for handling messages when certain links in a network experience outage or are unavailable due to channel access scheduling. In queueing networks, nodes unable to immediately forward their information to a next hop store their packets

(network layer units of messages) onto a local queue. The long term growth or overflow of any queue is an indicator that data traffic is not being directed appropriately. Routing protocols in queueing networks must then satisfy the requirement of stabilizing the queues in the network in addition to providing end-to-end routes for data delivery.

In 1989 [1, Tassiulas and Ephremedes] determined a routing and scheduling policy that could stabilize the largest set of packet generation rates in a network. Furthermore, it was proven that if there existed any routing and scheduling policy that could stabilize queues with a traffic generation vector in a fixed topology, their policy could stabilize the queues as well (their policy dominates over all other policies). The main idea of this optimal policy is to prioritize scheduling to alleviate packets with the highest queue backlog. Consider the scenario where packets were consistently sourced from a location and sinked at some destination within a connected network (packets are not queued at the sink). Nodes will direct traffic to neighboring hops if the queues of these hops are less than that of their own. Since the queue of the destination is always empty, queue backlogs will decrease based on proximity to the destination. This buildup of packets away from the destination generates a "potential" that influences next-hop routing decisions for large networks. Packets will move from high potential nodes to the lowest potential neighboring nodes, analogous to the pressure flow dynamics of fluid mechanical systems. This principal became known as *Backpressure Routing*. As described, backpressure routing incorporates aspects from the network layer 3 (routing) and network layer 2 (scheduling). With this discovery, it became apparent that *cross-layer* protocols (networking protocols that jointly optimized across different layers) were potentially more flexible and responsive to different environments and situations.

1.3 Related Work

1.3.1 Cooperative Communication

The fundamental work on diversity combining of signals is summarized in a comprehensive survey by [2, Brennan]. The pioneering concept for cooperative relaying was published by [3, Cover and Gamal], where the achiev-

able information theoretic capacity had been determined for general relay channels. [4, Laneman et al.] considered the performance benefits of coherently combining signals from both source-destination and relayed exchanges and provided information theoretic capacity bounds for amplify-and-forward and decode-and-forward cooperative diversity methods. In [5, Hasna and Alouini], the authors provide analytic methods to determine the BER and outage probability of links in the form of Bessel functions and Gaussian hypergeometric functions. [6, Su et al.] provides a simple expression for the channel gain probability density function of amplify-and-forward cooperative communications and uses it to determine closed form expressions for the symbol-error-rate of M-ary phase shift keying (MPSK) and quadrature amplitude modulation (QAM) signals. [7, Ibrahim et al.] considers the dynamic reallocation of power and selection of relays so that the capacity is maximized per power expenditure in cooperative communications.

Many of these previously mentioned works focus on the signal processing and information theoretic capacity of cooperative communications, but do not consider network aspects of nodes. In the investigated settings, a node is allocated long-term solely for the purpose of relaying, and the exchange of messages to schedule, retask or coordinate is unnecessary. The first works to consider network aspects of cooperative communication consider the requirements to coordinate a wireless node as a relay. [8, Khan and Karl] investigates messaging requirements needed for MAC protocols to begin cooperative transmission and provides a survey of different MAC schedulers that can coordinate cooperation within a wireless network. [9, Ding and Uysal] considers the signal processing properties and capacity gains for multiple amplify-and-forward relays in orthogonal frequency division multiplexing (OFDM) scenarios; however, they do not consider the allocation of channels due to traffic statistics (a network layer problem).

1.3.2 Backpressure Routing

The premier work on backpressure routing was published in [1, Tassiulas and Ephremedes]. The authors had modeled the state of the network queue vectors as a Markovian random process and applied Lyapunov control theory to guarantee the long-term finite-boundedness of queues under the backpressure

policy in a fixed and connected topology. The large admittance of exogenous traffic rates for the backpressure policy led it to be recognized as the throughput optimal scheduling policy. Since then, different authors have attempted to use the optimal throughput characteristics in maximizing end-to-end data delivery. [10, Ding et al.] formulates a convex optimization problem that maximizes the capacity-differential backpressure product. A distributed solution to this problem decouples the tasks of scheduling and multicarrier power allocation and is shown to outperform networks with static routes or fixed power allocation. [11, Liu et al.] extends the principle of backpressure routing by applying first order dynamic updates to *virtual queues*. By including information on the rate of queue differences, backpressure networks are shown to be more responsive to changes in topology and require less time to converge to a steady state policy.

[12, Neely] considers the performance of a backpressure network with receiver diversity. In the DIVersity BAckpressure Routing (DIVBAR) algorithm, each source sends messages to multiple relays and the destination based on differential backpressure. Upon unsuccessful packet reception at the destination, the responsibility for forwarding messages is passed onto the relay nodes. Because potential relays act only upon unsuccessful packet reception at the destination, cooperation is *reactive* and is not suitable for variable rate encoding transmission schemes. [13, Yeh and Berry] considers the scheduling of transmissions between direct source-destination pairs and cooperative decode-and-forward through Gaussian relay channels. Both of these methods perform decode-and-forward instead of amplify-and-forward, which necessitates common decoding architectures.

1.4 Cooperative Link Augmentation in Backpressure Networks

In this work, we consider a bottom-up (PHY-to-Network) approach to study the effects of cooperative communications in backpressure networks. In Chapter 2, we present a signal model for cooperative communications amplify-and-forward protocol for a binary phase shift keying (BPSK) signal as a physical layer model for our point-to-point transmissions. We show that with maximum ratio coherent combining of multiple diversity paths, the

BER of a cooperative communications amplify-and-forward link outperforms conventional single diversity point-to-point transmission. Using the moment generating function method, we also see that multiple relay diversity links can incrementally improve link quality with diminishing returns. Using the ergodic capacity of a channel achieved through channel-state waterfilling, we derive the conditional capacity gain to a single diversity link if a remote node were to assist as a relay. In Chapter 3, we consider the effect of cooperation in backpressure queueing network dynamics. We look at the ability of a MAC protocol to schedule multiple transmissions as the allowable capacity region of a network and determine how the capacity region changes with cooperative link augmentation. By considering the current state of network backlog and conditional capacity gain to scheduled transmissions, we determine a threshold rule for when a node should forgo its own backlog transmissions and participate as a relay. Consequently, we present an appropriate MAC protocol and message frame structure to facilitate cooperation. We show that under this new MAC protocol, the network queue can be stabilized for larger traffic generation rates not previously accommodated by backpressure networks without cooperation.

CHAPTER 2

COOPERATIVE COMMUNICATIONS

Cooperative communications is a class of physical layer communication techniques that exploit the inherent broadcast nature of wireless transmissions in enabling spatial diversity. During wireless transmissions, a signal from the source node is sent to a destination. Nearby cooperative nodes receive a copy of this signal and can decide to discard their copy or relay this signal to the destination. Assuming the signal transmissions paths to the destination are independent, multiple copies of the source signal are received and coherently combined to produce a signal with higher signal-to-noise ratio (SNR). The intended message from the signal is computed after the receiver decides it has accumulated enough copies of the signal to decode with sufficiently low probability of error. For this reason, such cooperative communication techniques are commonly referred to as energy accumulation techniques. We explore cooperative communication techniques as a physical layer solution to combat fading in heterogeneous node networks.

Different forms of cooperative communications have been studied extensively in literature. The capacity of different cooperative communications techniques has been studied in [4, Laneman et al.]. Relay selection techniques have been studied in [7, Ibrahim et al.]. Selection relaying and incremental relaying have been studied from a networking perspective and can provide intuition on which nodes can be allocated to relay in a cooperative fashion and how much performance improvement is expected.

Relaying techniques can be described as amplify-and-forward and decode-and-forward. In decode-and-forward techniques, the received message at the relay is decoded and retransmitted. This type of "hard" decision has been known to cause error propagation, where if the relay decodes the message incorrectly, it is more likely to cause bit errors within the combined message at the receiver. In amplify-and-forward techniques, the received message at the relay is normalized in power and retransmitted. Amplify-and-forward tech-

niques have the advantage of not requiring a relay node to have message decoding architecture and consequently is a strong candidate for heterogeneous cooperative transmissions. In this chapter, we provide the system model and metric considerations necessary for considering amplify-and-forward in differential backpressure networks.

2.1 Signal Model and Transmission Scheme

We assume the signal model shown in Figure 2.1 for our amplify-and-forward transmission scheme. Suppose we have a source S, destination D, and relay R. After coordination, S intends to transmit a signal to D with R as a relay. Henceforth, we will use the notation (SR), (SD), and (RD) to represent wireless channels between nodes. (We declare (SRD) the overall effect of a 2-hop "virtual" channel.) All channels are subject to Rayleigh fading and channel coefficients can be modeled as independent, zero-mean circularly symmetric Gaussian distributed random variables $(h_{SD}, h_{SR}, h_{RD}) \sim (\mathbb{C}N(0, \sigma_{SD}^2), \mathbb{C}N(0, \sigma_{SR}^2), \mathbb{C}N(0, \sigma_{RD}^2))$. We assume the fading coefficients will not change within the duration of a transmitted signal; that is, each individual symbol transmission period is shorter than the coherence time of the channel. Noise at different locations are assumed to be independent identically distributed complex Gaussian random variables denoted as $n_{ij} \sim \mathbb{C}N(0, N_0)$, respectively. For simplicity, we normalize the noise power N_0 to 1 and scale the corresponding channel gain accordingly. We denote the transmission power from the source as P_S, and wish to keep the average relay transmission power P_R. Assuming Binary Phase Shift Key (BPSK) signalling, the transmission of information symbol $x \in \{-1, 1\}$, our system can be modeled as:

$$y_{SD} = h_{SD}\sqrt{P_S}x + n_{SD} \tag{2.1}$$

$$y_{SR} = h_{SR}\sqrt{P_S}x + n_{SR} \tag{2.2}$$

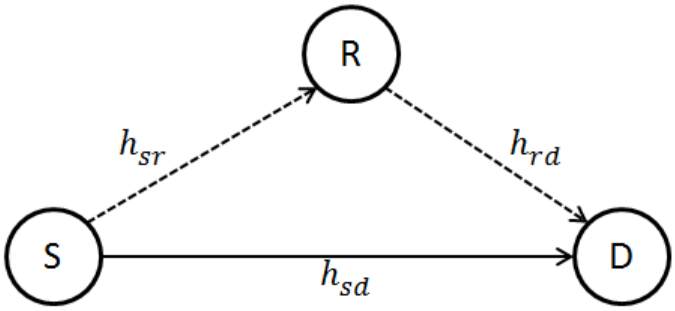

Figure 2.1: Cooperative Communication model

$$\begin{aligned} y_{RD} &= h_{RD}\sqrt{\frac{P_R}{P_S}}\frac{h_{SR}^*}{\|h_{SR}\|^2}y_{SR} + n_{RD} \\ &= h_{RD}\sqrt{P_R}x + \sqrt{\frac{P_R}{P_S}}\frac{h_{RD}}{h_{SR}}n_{SR} + n_{RD} \end{aligned} \tag{2.3}$$

where channel amplitude and phase equalization is executed at each receiving stage. (This scheme assumes globally known channel fading conditions and thus serves as an upper bound for performance of links from a network perspective.) The signal copies at the receiver can be linearly combined as

$$y_\Sigma = \alpha_{SD}\frac{h_{SD}^*}{\|h_{SD}\|}y_{SD} + \alpha_{RD}\frac{h_{RD}^*}{\|h_{RD}\|}y_{RD} \tag{2.4}$$

We wish to determine the appropriate combining coefficients α_{SD} and α_{RD} that would maximize the SNR of the combined signal y_Σ. If we define γ_Σ as the SNR of y_Σ, then we can express γ_Σ as

$$\gamma_\Sigma = \frac{\left(\alpha_{SD}\sqrt{P_S}|h_{SD}| + \alpha_{RD}\sqrt{P_R}|h_{RD}|\right)^2}{\left(\alpha_{SD}^2 + \alpha_{RD}^2\left[\frac{P_R}{P_S}\frac{|h_{RD}|}{|h_{SD}|} + 1\right]\right)N_0} = \frac{\left|\sum_{i=1}^2 \alpha_i\sqrt{P_i}h_i\right|^2}{\sum_{i=1}^2 \alpha_i^2\tilde{n_i}^2} \tag{2.5}$$

with the understanding that $i = 1$ corresponds to properties of the signal from the source, $i = 2$ corresponds to properties of the signal from the relay and $\tilde{n_i}$ represents the root mean squared of the cumulant noise on i. We use the Cauchy-Schwarz inequality to maximize the above expression. For any two finite vectors $\mathbf{a}$, $\mathbf{b}$ of length n in an inner product space

$$\left|\sum_{i=1}^{n} a_i^* b_i\right|^2 \leq \sum_{i=1}^{n} |a_i|^2 \sum_{i=1}^{n} |b_i|^2 \tag{2.6}$$

with equality if and only if $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent, that is $\mathbf{a} = K\mathbf{b}$ for some non-zero scalar K. Allowing the numerator to represent the left-hand side of the inequality and the denominator to represent one of the summations on the right-hand side, we can determine the coefficients α_{SD} and α_{RD}. Let $a_i^* b_i = \alpha_i \sqrt{P_i}\,|h_i|$ and $b_i = \alpha_i \tilde{n_i}$. Then (2.5) can be represented in the Cauchy-Schwarz inequality as

$$\frac{\left|\sum_{i=1}^{2} \alpha_i \sqrt{P_i} h_i\right|^2}{\sum_{i=1}^{2} \alpha_i^2 \tilde{n_i}^2} \leq \sum_{i=1}^{2} \frac{P_i\,|h_i|^2}{\tilde{n_i}^2} \tag{2.7}$$

where the right-hand side is the sum of SNRs received from individual diversity paths. The above expression can be met with equality only if $\alpha_i \tilde{n_i} = K\frac{\sqrt{P_i}|h_i|}{\tilde{n_i}}$. Thus we can determine the combining coefficients as

$$\alpha_{SD} = \frac{\sqrt{P_S}\,|h_{SD}|}{N_0} \tag{2.8}$$

$$\alpha_{RD} = \frac{\sqrt{P_R}\,|h_{RD}|}{\left(\frac{P_S|h_{SD}|^2}{P_R|h_{RD}|^2} + 1\right) N_0} \tag{2.9}$$

The coefficients α_{SD} and α_{RD} are commonly referred to as maximum ratio combining coefficients because they maximize the overall SNR of the combined signal [2, Brennan]. Effectively, a_{SD} and a_{RD} are weighted proportional to each signal amplitude and inversely proportional to noise power (similar to a reliability factor, where stronger signals are counted as more reliable). In (2.7), it is shown that the weighted sum of received signals from different diversity paths using these maximum ratio combining coefficients will result in a cumulant signal with SNR that is the sum of individual SNRs of each diversity branch. If we define γ_{ij} as the SNR of link (i, j), then we have

$$\gamma_\Sigma = \gamma_{SD} + \gamma_{SRD} \tag{2.10}$$

A maximum likelihood decoding decision can be made at the receiver on the combined signal to recover the information bit

$$p(X=1|Y_\Sigma) \underset{\hat{x}=-1}{\overset{\hat{x}=1}{\gtrless}} p(X=-1|Y_\Sigma) \Longrightarrow \hat{x} = sign\left(\mathbb{R}(Y_\Sigma)\right) \tag{2.11}$$

2.2 Average Bit Error Rate Performance of Cooperative Communication Amplify-and-Forward with Maximum Ratio Combining

In this section, we evaluate the average bit error rate (BER) of a cooperative communication amplify-and-forward link. The average BER can be calculated exactly by the moment generating function (MGF) method. The probability of error of BPSK signal is given by

$$P_\epsilon = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2\gamma}}^{\infty} e^{\frac{-t^2}{2}} dt = Q\left(\sqrt{2\gamma}\right) \tag{2.12}$$

where γ is the instantaneous SNR of the received signal and the Q-function is the complementary cumulative distribution function of the standard Gaussian with zero mean and unit variance, defined as

$$Q\left(z\right) = \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{\frac{-t^2}{2}} dt \tag{2.13}$$

Then the average BER can be calculated as

$$\bar{P}_\epsilon = \int_0^\infty Q\left(\sqrt{2\gamma}\right) p\left(\gamma\right) d\gamma \tag{2.14}$$

$$= \int_0^\infty \int_{\sqrt{2\gamma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt \cdot p\left(\gamma\right) d\gamma \tag{2.15}$$

where $p\left(\gamma\right)$ is the probability that the received signal has an SNR of γ. Because the limits of integration in (2.15) include a random variable, direct integration to solve for the probability of error is usually not feasible. [14, Craig] had derived an alternate representation of the Gaussian Q-function as

$$Q\left(z\right) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{z^2}{2\sin^2\theta}\right) d\theta \qquad z \geq 0 \tag{2.16}$$

which is valid as long as the argument of the Q-function is positive. Plugging (2.16) into (2.14), we get a new expression for the average probability of error

$$\bar{P}_\epsilon = \int_0^\infty Q\left(\sqrt{2\gamma}\right) p\left(\gamma\right) d\gamma \tag{2.17}$$

$$= \int_0^\infty \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{2\gamma}{2\sin^2\theta}\right) d\theta p\left(\gamma\right) d\gamma \tag{2.18}$$

$$= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \int_0^\infty \exp\left(-\frac{\gamma}{\sin^2\theta}\right) p\left(\gamma\right) d\gamma d\theta \tag{2.19}$$

The inner integral in (2.19) is the MGF of the SNR γ. The MGF of a continuous random variable X is defined as

$$M_X(s) = \mathbb{E}\left[e^{sX}\right] = \int e^{sx} p_X\left(x\right) dx \tag{2.20}$$

where $\mathbb{E}\left[f\left(x\right)\right]$ is the expectation of $f\left(x\right)$ with respect to the distribution of X, and the integral is taken over the support of X. With this, (2.19) can be expressed in the insightful form of

$$\bar{P}_\epsilon = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} M_\gamma\left(-\frac{1}{\sin^2\theta}\right) d\theta \tag{2.21}$$

The average BER $\bar{p}_e$ of a BPSK signal can be evaluated by first computing the moment generating function of the SNR M_{γ_Σ}. By independence of channels (SD) and (SRD), the overall MGF can be evaluated as the product of the MGFs of the individual channels $M_{\gamma_\Sigma} = M_{\gamma_{SD}} M_{\gamma_{SRD}}$. The SNR $\gamma_{SD} = \frac{P_S |h_{SD}|^2}{N_0}$ is a function of power output P_S and the exponentially distributed random variable $|h_{SD}|^2$. An exponential random variable X with mean λ has the following density and distribution functions

$$p_X\left(x\right) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \cdot U\left(x\right) \qquad P_X\left(x\right) = 1 - e^{-\frac{x}{\lambda}} \tag{2.22}$$

where $U\left(\cdot\right)$ is the unit step function. Note that exponential random variables are closed under positive scaling, so γ_{SD} is an exponential random variable with mean $P_S \sigma_{SD}^2$. To see this, consider the distribution function of a random variable $Z = KX$ for some positive scalar K and exponential random variable X with mean λ. Then $\frac{Z}{K} = X$ and

$$P_Z\left(z\right) = P_{KX}\left(z\right) = P\left(KX \leq z\right) = P\left(X \leq \frac{z}{K}\right) = 1 - e^{\frac{z}{K\lambda}} \tag{2.23}$$

which is the distribution of an exponential random variable with mean $K\lambda$. For simplicity, let $P_S\sigma_{SD}^2 = \beta$, $P_S\sigma_{SR}^2 = \beta_1$, and $P_R\sigma_{RD}^2 = \beta_2$. Then γ_{SD} has density function and MGF:

$$p_{\gamma_{SD}}(\gamma) = \frac{1}{\beta} e^{-\frac{\gamma}{\beta}} \cdot U(\gamma) \tag{2.24}$$

$$M_{\gamma_{SD}}(s) = \mathbb{E}\left[e^{s\gamma_{SD}}\right] = \int_{-\infty}^{\infty} e^{s\gamma} p_{\gamma_{SD}}(\gamma) d\gamma = \frac{\frac{1}{\beta}}{\frac{1}{\beta} - s} \tag{2.25}$$

To determine the SNR of the 2-hop relay link (SRD), we consider the expanded signal model in (2.3)

$$y_{RD} = h_{RD}\sqrt{P_R}x + \sqrt{\frac{P_R}{P_S}}\frac{h_{RD}}{h_{SR}} n_{SR} + n_{RD} \tag{2.26}$$

Then γ_{SRD}:

$$\gamma_{SRD} = \frac{P_R |h_{RD}|^2}{\left(\frac{P_R|h_{RD}|^2}{P_S|h_{SR}|^s} + 1\right) N_0} = \frac{\frac{P_R|h_{RD}|^2}{N_0}\frac{P_S|h_{SR}|^2}{N_0}}{\frac{P_R|h_{RD}|^2}{N_0} + \frac{P_S|h_{SR}|^2}{N_0}} = \mu_H\left(\frac{P_S|h_{SR}|^2}{N_0}, \frac{P_R|h_{RD}|^2}{N_0}\right) \tag{2.27}$$

where $\mu_H(a,b) = \frac{ab}{a+b}$ represents the (half) harmonic mean of its arguments. The MGF of the harmonic mean of two exponential random variables with means β_1 and β_2 is [15, Hasna and Alouini]:

$$M_{\mu H}(s,\beta_1,\beta_2) = \frac{16\frac{1}{\beta_1}\frac{1}{\beta_2}}{3\left(\frac{1}{\beta_1} + \frac{1}{\beta_2} + s + 2\sqrt{\frac{1}{\beta_1\beta_2}}\right)} \left[\frac{4\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right)}{\frac{1}{\beta_1} + \frac{1}{\beta_2} + s + 2\sqrt{\frac{1}{\beta_1\beta 2}}} \cdot {}_2F_1\left(3, \frac{3}{2}; \frac{5}{2}; \frac{\frac{1}{\beta_1} + \frac{1}{\beta_2} + s - 2\sqrt{\frac{1}{\beta_1\beta_2}}}{\frac{1}{\beta_1} + \frac{1}{\beta_2} + s + 2\sqrt{\frac{1}{\beta_1\beta_2}}}\right) + {}_2F_1\left(2, \frac{1}{2}; \frac{5}{2}; \frac{\frac{1}{\beta_1} + \frac{1}{\beta_2} + s - 2\sqrt{\frac{1}{\beta_1\beta_2}}}{\frac{1}{\beta_1} + \frac{1}{\beta_2} + s + 2\sqrt{\frac{1}{\beta_1\beta_2}}}\right)\right] \tag{2.28}$$

$${}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)\Gamma(1+n)} z^n \tag{2.29}$$

where (2.29) is the Gauss hypergeometric series [16, Abramowitz] and $\Gamma(n)$ is the gamma function. The MGF of the combined SNRs from (2.10) can

be determined by multiplying the MGFs of γ_{SD}, and γ_{SRD}. In order to determine the average BER using the MGF, we need to evaluate:

$$\bar{P}_e = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} M_{\gamma_{SD}} \cdot M_{\mu_H} \left(-\frac{1}{\sin^2(\phi)} \right) d\phi \tag{2.30}$$

Numerical integration of (2.30) is shown in Figure 2.2. In this simulation, we compare three scenarios with different resource allocation and receiving methods. A baseline scenario considers when the source can allocate all of its power in a direct transmission over a Rayleigh faded channel (SD). Another scenario considers when the source and relay evenly distribute the total transmission power between themselves for transmission over channels (SR) and (RD); however, the destination ignores any messages it overhears from channel (SD). We label this case as "2-hop Relaying". A third scenario considers when transmission power is evenly distributed between the source and relay, and the destination coherently combines signals from the two diversity paths as described in the previous section. The total transmission power in each scenario is preserved.

We see that even under equal (not necessarily optimal) power allocation between nodes for cooperation, a cooperative communication link performs as well as a direct transmission link in low SNR conditions, and has significant diversity gain at high SNRs. This diversity gain suggests improvements in power consumption and signal quality, and can be even further augmented by additional links. Figure 2.3 shows the analytic performance of half-power allocation schemes for multiple relays (half of the overall power comes from source, and the remaining power is equally distributed over all relays).

2.3 Link Performance Improvement with Amplify and Forward Cooperation

To investigate the rate improvement a cooperative communication scheme would have, we first consider the optimal power allocation for opportunistic transmissions. [17, Goldsmith and Varaiya] proposed that if nodes are allowed to vary their transmission power and rate in response to known channel state information, the achievable capacity subject to some average power constraint is solved by the following convex optimization problem:

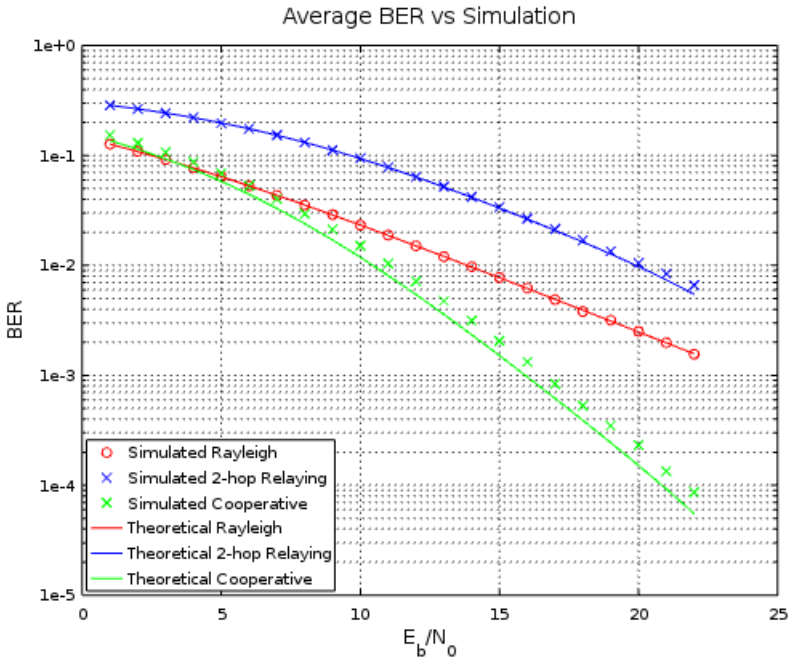

Figure 2.2: $\bar{P}_\epsilon$ in Cooperative Communication ($\sigma^2_{SR} = \sigma^2_{SD} = \sigma^2_{RD} = 1$)

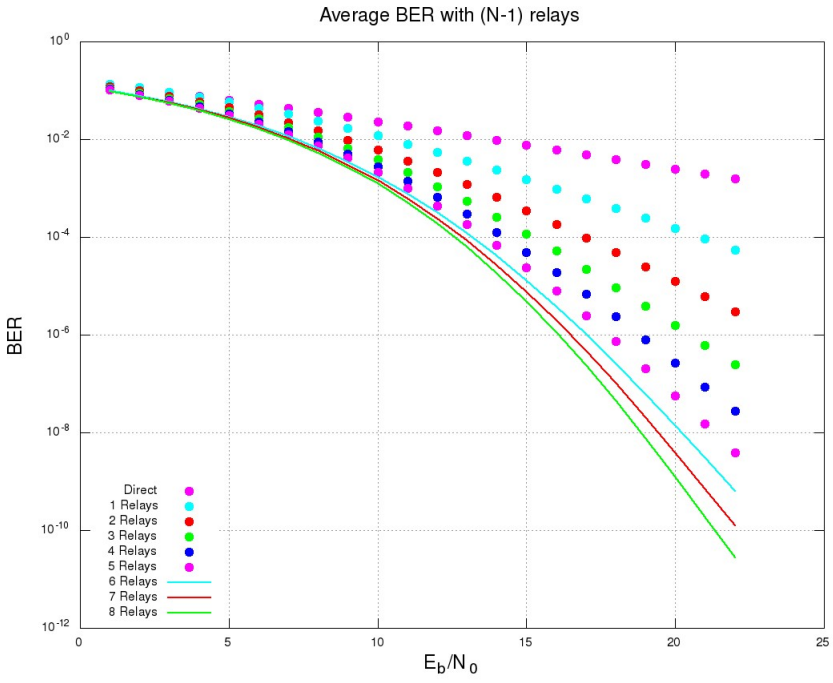

Figure 2.3: $\bar{P}_\epsilon$ in Cooperative Communication with Multiple Relays

$$C^{*}\left(\mathbf{S}\right)=\max_{\mathbf{S}(\gamma):\int \mathbf{S}(\gamma)p(\gamma)d\gamma=P_{S}}\int_{\gamma}\omega\log\left(1+\frac{\mathbf{S}\left(\gamma\right)\gamma}{P_{S}}\right)p\left(\gamma\right)d\gamma \tag{2.31}$$

where P_S is the total average transmission power and $\mathbf{S}\left(\gamma\right)$ is the instantaneous transmission power of the transmitter in response to channel gain γ. Equation (2.31) can be solved by minimizing the following Lagrangian dual function (with $\lambda \geq 0$ as a dual variable):

$$\begin{aligned} L\left(\lambda\right) &= \int_{\gamma}\log\left(1+\frac{\mathbf{S}\left(\gamma\right)\gamma}{P_{S}}\right)p\left(\gamma\right)-\lambda\left(\int_{\gamma}\mathbf{S}\left(\gamma\right)p\left(\gamma\right)-P_{S}\right) \\ &= \int_{\gamma}\left[\log\left(1+\frac{\mathbf{S}\left(\gamma\right)\gamma}{P_{S}}\right)-\lambda\mathbf{S}\left(\gamma\right)\right]p\left(\gamma\right)+\lambda P_{S} \end{aligned} \tag{2.32}$$

$$\frac{\partial L}{\partial\mathbf{S}\left(\gamma\right)}=\left[\frac{\frac{\gamma}{P_{S}}}{1+\frac{\mathbf{S}(\gamma)\gamma}{P_{S}}}-\lambda\right]p\left(\gamma\right)=0\Rightarrow\frac{\mathbf{S}\left(\gamma\right)}{P_{S}}=\frac{1}{\lambda P_{S}}-\frac{1}{\gamma} \tag{2.33}$$

This solution results in a state-based waterfilling technique to allocate power. Starting from $\lambda = 0$, the (2.33) allocates too much power to satisfy the constraint in (2.31). The power constraint is gradually satisfied by increasing λ and results in the optimal rate of information transmissions. It is important to note that this power allocation is a function of the current channel state and it is seemingly agnostic to the channel state distribution function because of the dual problem. [17, Goldsmith and Varaiya] succinctly define the optimal power allocation as

$$\frac{\mathbf{S}\left(\gamma\right)}{P_{S}}=\frac{1}{\gamma_{0}}-\frac{1}{\gamma};\quad\gamma_{0}<\gamma \tag{2.34}$$

for some channel gain threshold γ_0. When $\gamma < \gamma_0$, the source node does not transmit and waits for a better transmission opportunity. The capacity determined by this expression is

$$C^{*}=\int_{\gamma}\omega\log\left(1+\left(\frac{1}{\gamma_{0}}-\frac{1}{\gamma}\right)\gamma\right)p\left(\gamma\right)d\gamma=\int_{\gamma_{0}}^{\infty}\omega\log\left(\frac{\gamma}{\gamma_{0}}\right)p\left(\gamma\right)d\gamma \tag{2.35}$$

In a Rayleigh fading channel, the capacity expression becomes

$$\begin{aligned} C^* &= \int_{\gamma_0}^{\infty} \omega \log\left(\frac{\gamma}{\gamma_0}\right) \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\gamma} d\gamma \\ &= \int_{\gamma_0}^{\infty} \omega \left(\log(\gamma) - \log(\gamma_0)\right) \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\gamma} d\gamma \\ &= \int_{\gamma_0}^{\infty} \omega \log(\gamma) \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\gamma} d\gamma - \omega \log(\gamma_0) e^{-\frac{1}{\sigma^2}\gamma_0} \\ &= \omega \log(\gamma) e^{-\frac{1}{\sigma^2}\gamma}\Big|_{\gamma=\infty}^{\gamma_0} + \omega \int_{\gamma_0}^{\infty} \frac{1}{\gamma} e^{-\frac{1}{\sigma^2}\gamma} d\gamma - \omega \log(\gamma_0) e^{-\frac{1}{\sigma^2}\gamma_0} \\ &= \omega E_1\left[\frac{1}{\sigma^2}\gamma_0\right] \end{aligned} \tag{2.36}$$

where $E_1[x] = \int_x^{\infty} \frac{1}{t} e^{-t} dt$ is the exponential integral of the first order. The average power expenditure P_S determines the appropriate value for γ_0. For a Rayleigh fading channel

$$\begin{aligned} \int_{\gamma} \mathbf{S}(\gamma) p(\gamma) d\gamma = P_S \Rightarrow 1 &= \int_{\gamma_0}^{\infty} \left(\frac{1}{\gamma_0} - \frac{1}{\gamma}\right) \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\gamma} d\gamma \\ &= \frac{1}{\gamma_0} e^{-\frac{1}{\sigma^2}\gamma_0} - \int_{\gamma_0}^{\infty} \frac{1}{\gamma\sigma^2} e^{-\frac{1}{\sigma^2}\gamma} d\gamma \\ &= \frac{1}{\gamma_0} e^{-\frac{1}{\sigma^2}\gamma_0} - \frac{1}{\sigma^2} E_1\left[\frac{1}{\sigma^2}\gamma_0\right] \end{aligned} \tag{2.37}$$

Since $E_1[x]$ is not an elementary function, γ_0 can be numerically evaluated as the roots of the equation:

$$f(\gamma_0) = \frac{1}{\gamma_0} e^{-\frac{1}{\sigma^2}\gamma_0} - \frac{1}{\sigma^2} E_1\left[\frac{1}{\sigma^2}\gamma_0\right] - 1 \tag{2.38}$$

The plot in Figure 2.4 shows the achievable transmission rate C^* in a Rayleigh fading channel with channel variance $\sigma^2 = 1$. Figure 2.5 shows the corresponding γ_0 transmission decision threshold necessary to satisfy the average power expenditure requirement P_S. Because of the $E_1\left[\frac{1}{\sigma^2}\gamma_0\right]$ term, one might find the bounding approximations of the exponential integral useful [16, Abramowitz].

$$\frac{1}{2} e^{-x} \ln\left(1 + \frac{2}{x}\right) < E_1(x) < e^{-x} \ln\left(1 + \frac{1}{x}\right) \tag{2.39}$$

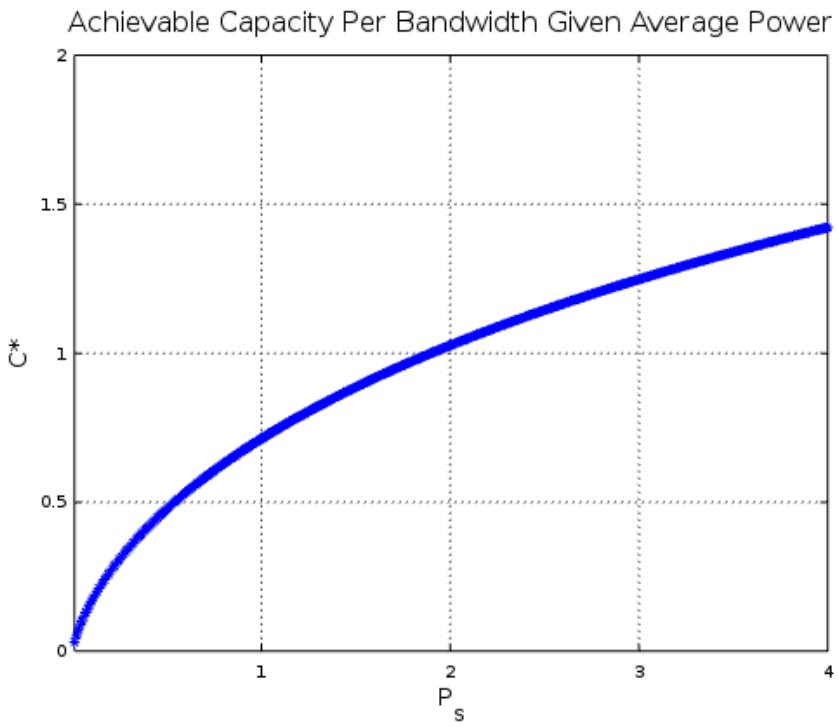

Figure 2.4: Achievable Capacity in a Rayleigh Fading Channel C^*

As σ^2 increases (channel is better on average), so does the value of γ_0 ($\lim_{\sigma^2 \to \infty} \gamma_0 = 1$). The rationale suggests that if the channel has more opportunities for higher gain, the transmitter should wait for these high gain opportunities to transmit while still being able to support its average power expenditure requirement. We consider this protocol "state-based opportunistic transmission" to achieve the most point-to-point throughput given power consumption. Because of the direct relationship between average power expenditure and achievable capacity, we realize that this same protocol determines the minimum power necessary to fulfill a required transmission rate. Consequently, (2.36) can be used to determine γ_0 and the minimum average power necessary to fulfill a capacity requirement in this scenario. (This problem is not a convex problem because different power assignments in response to transmission states do not form a convex set of possible power expenditures).

To achieve the projected capacity performance [17, Goldsmith and Varaiya] proposed the transmitter encoding and receiver decoding architecture in Figure 2.6. With known γ, the transmitter will select an appropriate code and power allocation. The decoder will decode the received message with the

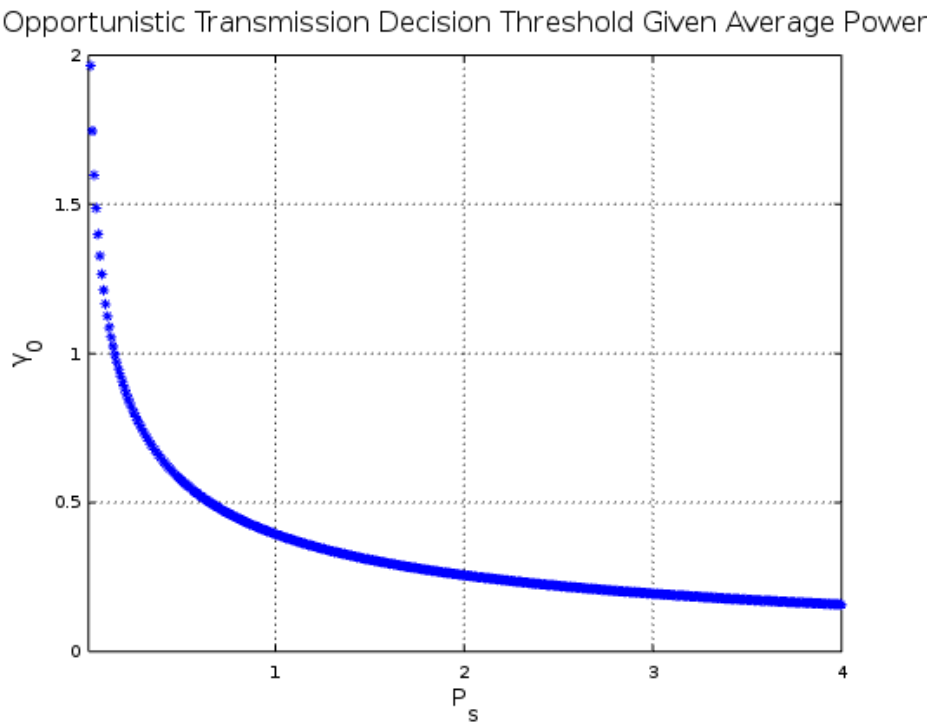

Figure 2.5: Opportunistic Transmission Threshold γ_0

appropriate codebook.

In order to solve (2.35) for a cooperative link, we must determine the density function of the equivalent channel SNR (2.10). The density of the sum of independent random variables can be computed using convolution or with the characteristic function/MGF method. The probability density function of γ_{SRD} is given by [15, Hasna and Alouini]:

$$p_{\gamma_{SRD}}(\gamma) = \frac{2\gamma e^{-\gamma\left(\frac{1}{\beta_1}-\frac{1}{\beta_2}\right)}}{\beta_1\beta_2} \cdot \left[\left(\frac{\beta_1+\beta_2}{\sqrt{\beta_1\beta_2}}\right) K_1\left(\frac{2\gamma}{\sqrt{\beta_1\beta_2}}\right) + 2K_0\left(\frac{2\gamma}{\sqrt{\beta_1\beta_2}}\right)\right] \cdot U(\gamma) \quad (2.40)$$

where K_1 and K_0 represent the first and zeroth order modified Bessel functions of the second kind. Alternatively, [6, Su et al.] provides a probability density function for γ_{SRD} without the use of Bessel functions:

$$p_{\gamma_{SRD}}(\gamma) = \gamma \int_0^1 \frac{1}{\beta_1\beta_2 t^2 (1-t)^2} e^{-\left(\frac{1}{\beta_1(1-t)}+\frac{1}{\beta_2 t}\right)\gamma} dt \cdot U(\gamma) \quad (2.41)$$

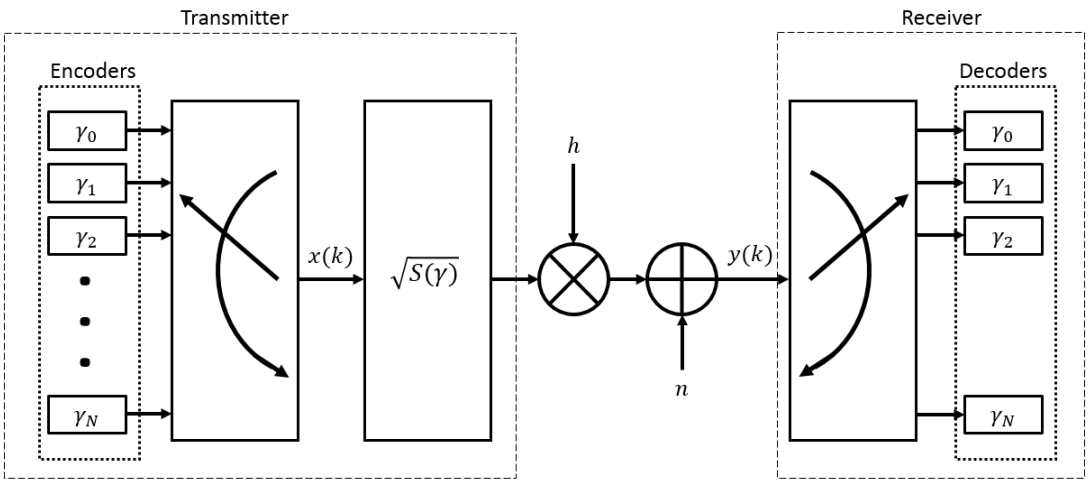

Figure 2.6: Channel State Based Encoding/Decoding Architecture

Then the density function of the equivalent cooperative channel SNR γ_Σ can be determined with the convolution of the two density functions (2.41) and (2.24)

$$\begin{aligned} p_{\gamma_\Sigma}(\gamma) &= \left(p_{\gamma_{SD}} * p_{\gamma_{SRD}}\right)(\gamma) \\ &= \int_{-\infty}^{\infty} p_{\gamma_{SD}}(\gamma-\tau)\, p_{\gamma_{SRD}}(\tau)\, d\tau \\ &= \int_{-\infty}^{\infty}\int_{0}^{1} \frac{1}{\beta} e^{-\frac{\gamma-\tau}{\beta}} \cdot \frac{\tau e^{-\left(\frac{1}{\beta_1(1-t)}+\frac{1}{\beta_2 t}\right)\tau}}{\beta_1\beta_2 t^2(1-t)^2} dt \cdot U(\gamma-\tau)\, U(\tau) d\tau \\ &= \int_{0}^{\gamma} \tau e^{-\frac{\gamma}{\beta}} \int_{0}^{1} \frac{e^{-\left(\frac{1}{\beta_1(1-t)}+\frac{1}{\beta_2 t}-\frac{1}{\beta}\right)\tau}}{\beta\beta_1\beta_2 t^2(1-t)^2} dt \cdot d\tau \end{aligned} \tag{2.42}$$

Instead of evaluating the above integral, we consider a *reactive protocol* where a relay decides to cooperate given some known information about the current channel states, which we assume can be exchanged via handshake messages. In that case, we can consider the expected gain a relay can provide if it chooses to cooperate, given channel states. Without cooperation, the source destination link at one instance of channel state γ_{SD} can provide

$$C_{old} = \omega \log\left(\frac{\gamma_{SD}}{\gamma_0}\right) \tag{2.43}$$

Suppose by overhearing the source message to initiate transmission and the

destination's reply (typically a request-to-send and clear-to-send, or RTS/CTS), the relay can determine γ_{SD} and γ_{SR}. At this point, it can consciously make the determination whether it would be beneficial to aid the source's transmission. With the known channel SNRs γ_{SD} and γ_{RD}, and the mean of the channel SNR from the relay to destination ($P_R\sigma_{RD}^2$), the relay can determine an expected transmission rate gain if it functioned temporarily as a relay. The expected transmission rate gain can be calculated as:

$$\begin{aligned}
\mathbb{E}\left[\Delta C|\gamma_{SD},\gamma_{SR}\right] &= \int_0^\infty \omega \log\left(\frac{\gamma_{SD}+\frac{\gamma_{SR}\cdot\gamma}{\gamma_{SR}+\gamma}}{\gamma_0}\right) p_{\gamma_{RD}}(\gamma)\, d\gamma - C_{old} \\
&= \int_0^\infty \omega \log\left(\frac{\gamma_{SD}+\frac{\gamma_{SR}\cdot\gamma}{\gamma_{SR}+\gamma}}{\gamma_0}\right) \frac{1}{\beta_2} e^{-\frac{\gamma}{\beta_2}} d\gamma - C_{old} \\
&= \int_0^\infty \omega \left[\log\left[\gamma_{SD}\left(1+\frac{1}{\gamma_{SD}}\cdot\frac{\gamma_{SR}\cdot\gamma}{\gamma_{SR}+\gamma}\right)\right] - \log(\gamma_0)\right] \frac{1}{\beta_2} e^{-\frac{\gamma}{\beta_2}} d\gamma - C_{old} \\
&= \int_0^\infty \omega \left[\log(\gamma_{SD}) + \log\left(1+\frac{1}{\gamma_{SD}}\cdot\frac{\gamma_{SR}\cdot\gamma}{\gamma_{SR}+\gamma}\right) \frac{1}{\beta_2} e^{-\frac{\gamma}{\beta_2}} d\gamma - \omega\log(\gamma_{SD})\right] \\
&= \int_0^\infty \omega \log\left(1+\frac{1}{\gamma_{SD}}\cdot\frac{\gamma_{SR}\cdot\gamma}{\gamma_{SR}+\gamma}\right) \frac{1}{\beta_2} e^{-\frac{\gamma}{\beta_2}} d\gamma
\end{aligned} \tag{2.44}$$

which can be evaluated using integration by parts:

$$\int_a^b u dv = (u\cdot v)\Big|_a^b - \int_a^b v du \tag{2.45}$$

$$\begin{aligned}
u &= \log\left(1+\frac{1}{\gamma_{SD}}\cdot\frac{\gamma_{SR}\cdot\gamma}{\gamma_{SR}+\gamma}\right) \qquad dv = \frac{1}{\beta_2} e^{-\frac{\gamma}{\beta_2}} d\gamma \\
du &= \frac{\frac{\gamma_{SR}[\gamma_{SD}(\gamma_{SR}+\gamma)]-\gamma_{SD}(\gamma_{SR}\gamma)}{[\gamma_{SD}(\gamma_{SR}+\gamma)]^2}}{\left[1+\frac{1}{\gamma_{SD}}\cdot\frac{\gamma_{SR}\cdot\gamma}{\gamma_{SR}+\gamma}\right]} d\gamma \qquad v = -e^{-\frac{\gamma}{\beta_2}}
\end{aligned} \tag{2.46}$$

The term du can be further simplified by partial fraction decomposition

$$du = \frac{\frac{\gamma_{SR}[\gamma_{SD}(\gamma_{SR}+\gamma)]-\gamma_{SD}(\gamma_{SR}\gamma)}{[\gamma_{SD}(\gamma_{SR}+\gamma)]^2}}{\left[1+\frac{1}{\gamma_{SD}}\cdot\frac{\gamma_{SR}\cdot\gamma}{\gamma_{SR}+\gamma}\right]}d\gamma = \frac{\gamma_{SR}^2}{(\gamma+\gamma_{SR})\left[(\gamma_{SR}+\gamma_{SD})\,\gamma+\gamma_{SR}\cdot\gamma_{SD}\right]}d\gamma \tag{2.47}$$

$$= \frac{\gamma_{SR}^2/\left(\gamma_{SR}+\gamma_{SD}\right)}{(\gamma+\gamma_{SR})\left(\gamma+\frac{\gamma_{SR}\cdot\gamma_{SD}}{\gamma_{SR}+\gamma_{SD}}\right)}d\gamma = \left[\frac{-1}{(\gamma+\gamma_{SR})}+\frac{1}{\left(\gamma+\frac{\gamma_{SR}\cdot\gamma_{SD}}{\gamma_{SR}+\gamma_{SD}}\right)}\right]d\gamma$$

It is interesting to note that the harmonic mean of γ_{SR} and γ_{SD} appears as a root in the partial fraction decomposition in (2.47). Plugging in (2.46) and (2.47) to evaluate (2.44), we get:

$$\mathbb{E}\left[\Delta C^*|\gamma_{SD},\gamma_{SR}\right] = \omega\left(e^{\frac{\mu_H(\gamma_{SR},\gamma_{SD})}{\beta_2}}E_1\left[\frac{\mu_H\left(\gamma_{SR},\gamma_{SD}\right)}{\beta_2}\right] - e^{\frac{\gamma_{SR}}{\beta_2}}E_1\left[\frac{\gamma_{SR}}{\beta_2}\right]\right) \tag{2.48}$$

To understand (2.48), we consider the following function and its derivative:

$$f\left(x\right) = e^x E_1\left[x\right] \qquad\qquad f'\left(x\right) = e^x\left[E_1\left[x\right] - \frac{e^{-x}}{x}\right] \tag{2.49}$$

The derivative of this function can be determined by applications of the chain rule and the second fundamental theorem of calculus. The expression within the square brackets of (2.49) can be shown to be non-positive, specifically, $\frac{e^{-x}}{x}$ is an upper bound of $E_1\left[x\right]$ [16, Abramowitz, pg 229, eq 5.1.19]. Consequently, the term in the parenthesis of (2.48) cannot be negative, because the μ_H is upper bounded by the minimum if its arguments. The function $e^x E_1\left[x\right]$ is plotted in Figure 2.7. If γ_{SD} and $\mu_H\left(\gamma_{SR},\gamma_{RD}\right)$ were plotted on the abscissa, the difference between the corresponding function values would represent expected capacity gain per unit bandwidth for a source to destination link if a node supported as a relay.

We conclude with the statement that if a relay decides to engage in cooperative transmission, the cooperation cannot reduce the average capacity (an intuitive result). In the next chapter, we investigate when a node should decide to cooperate as a relay.

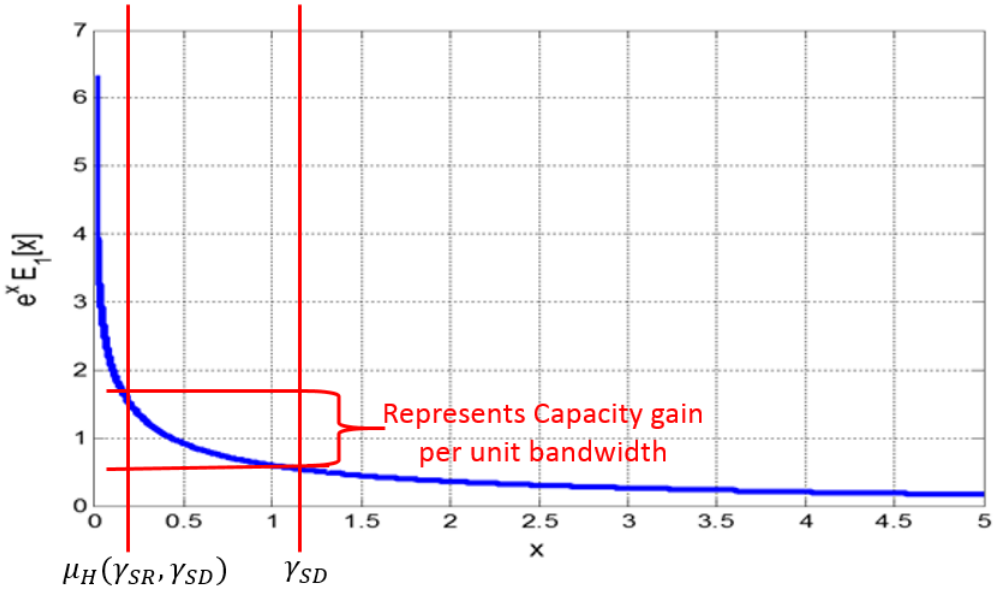

Figure 2.7: Function $e^x E_1[x]$ Used to Determine Potential Increase in Capacity if Cooperatively Communicating

CHAPTER 3

BACKPRESSURE ROUTING

In the previous chapter, we have demonstrated how a cooperating node can act as an amplify-and-forward relay and augment single hop source-destination transmissions. On top of that, because the relay does not decode any information, specialized hardware and forward error correction circuitry are not necessary; virtually any node that can sample and retransmit on common frequencies can assist as a relay. Furthermore, by using adaptive rate transmissions, we have shown that the augmented BER and SINR performance can yield an overall rate improvement. In this chapter, we investigate efficient ways of leveraging this rate improvement from a network perspective. This study leads to the development of an efficient distributed medium access control (MAC) protocol and a cross-layer throughput optimal implementation with cooperation.

3.1 Queueing Networks and Differential Backpressure

We consider a network $\mathbf{G}$ of nodes $\mathbf{N}$ and links $\mathbf{E}$ denoted $\mathbf{G} = (\mathbf{N}, \mathbf{E})$. Each node in the network has the ability to queue packets that it has not had the opportunity to transmit yet. Packets intended for different destinations are placed into different queues; for example, packets locally queued at n_i intended for n_j will be placed into q_i^j. Let $q_i^j(t)$ represent the non-negative number of packets at q_i^j at time t. This description allows us to denote node queue processes as $\mathbf{q}_i(t) = \begin{bmatrix} q_i^1(t) & q_i^2(t) & \dots & q_i^{|\mathbf{N}|}(t) \end{bmatrix}^T$ and the network queue process as $\mathbf{Q}(t) = \begin{bmatrix} \mathbf{q}_1(t) & \mathbf{q}_2(t) & \dots & \mathbf{q}_{|\mathbf{N}|}(t) \end{bmatrix}^T$. Packets arriving at their intended destination exit the network instead of queuing at their destination, so $q_i^i(t) = 0 \ \forall n_i \in \mathbf{N}, \forall t \in \mathbf{T}$. The exogenous traffic generated at each node n_i is a random process denoted $\lambda_i^j(t)$. For this thesis, we will consider each λ_i^j as independent wide-sense stationary non-negative integer

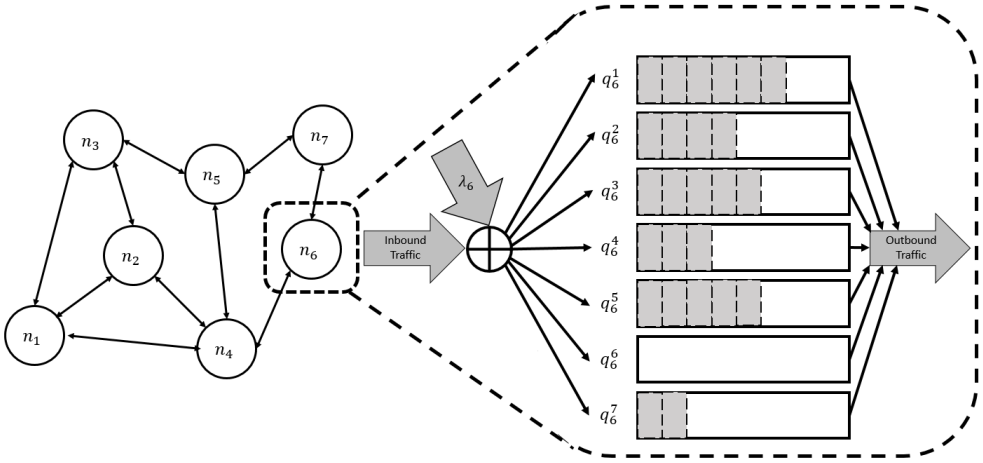

Figure 3.1: Queueing Network Model

random processes with mean $\mathbb{E}\left[\lambda_i^j\right] = \mu_i^j$ and finite variance. We define the exogenous traffic rate matrix Λ as the matrix of expected values of the λ_i^j processes. The breakdown of such a structure is shown in Figure 3.1.

We investigate a single hop network $\mathbf{G}_1$ of 4 nodes shown in Figure 3.2 as it is the most illustrative of our results. In this example, node n_1 generates traffic intended for node n_2 and node n_3 generates traffic intended for node n_4 (we designate nodes n_1 and n_3 as *sources*), while nodes n_2 and n_4 do not generate packets. Furthermore, we impose an additional constraint that nodes n_1 and n_3 cannot decode each other's messages (they cannot queue each other's messages). In this single hop network, the exogenous traffic rate matrix Λ can be sparsely represented as a vector $\Lambda = \left[\mathbb{E}\left[\lambda_1^2\right] \quad \mathbb{E}\left[\lambda_3^4\right]\right]^T$.

Suppose there exists enough available bandwidth in the spectrum so that both nodes n_1 and n_3 can schedule their transmissions without interfering with each other and that they are resource constrained only in their available transceiver capabilities. During each timeslot t, each node n_i can schedule a transmission to n_j and allow for a rate of information transfer $r_{ij}(t)$ limited by the capacity between the two nodes, $c_{ij}(t)$, that is, $r_{ij}(t) \leq c_{ij}(t)$. Then queue length dynamics can be modeled by the following equations

$$q_i^j(t+1) = \max\left(q_i^j(t) - r_{ij}(t), 0\right) + \lambda_i^j(t) \qquad (i,j) \in (1,2), (3,4) \quad (3.1)$$

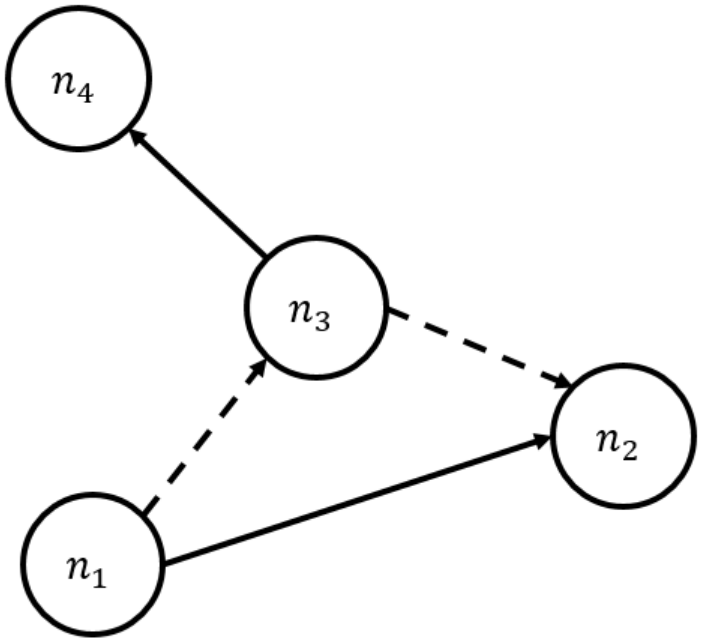

Figure 3.2: Single Hop Network $\mathbf{G}_1$

$$q_i^j(t) = 0 \qquad\qquad i \in 2, 4 \qquad\qquad (3.2)$$

where the $\max(\cdot, 0)$ in (3.1) is because nodes cannot send more information than they have queued. Assuming nodes n_1 and n_3 have non-zero channel access probability only when have non-empty queues, we can define the capacity region of the network Π as the set of admissible exogenous traffic rates that the network can support without any queue growing to infinity at any time. We consider a network that satisfies this condition as *stable* and conversely, the network is *unstable* if any queue goes to infinity at any point in time. The capacity region of a network described like $\mathbf{G}_1$ is an n-dimensional hypercube where the side lengths are the source-destination link capacities and n is the number of sources in the network.

The shape of Π is influenced by different factors, such as channel access probability, power limitations of sources, and available resources. Two examples of capacity regions are shown in Figure 3.3, where the horizontal axis represents the transmission rate from node n_1 to n_2, while the vertical axis represents the transmission rate from node n_3 to n_4. The shaded region represents the network capacity region Π. In Figure 3.3a, we assume that n_1 and

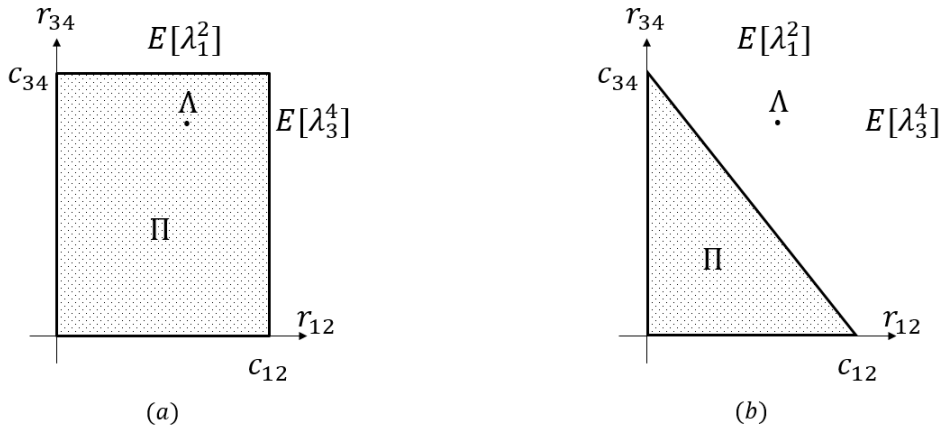

Figure 3.3: Capacity Regions of $\mathbf{G}_1$ with Scheduling Constraints

n_3 are fitted with an appropriate MAC and there are sufficient bandwidth resources so that they can both schedule transmissions simultaneously. In that case, Π is rectangularly shaped with each side length limited by their respective capacities, c_{ij}. If the network is environmentally resource constrained or the MAC is designed to only schedule a limited number of users, Π takes a shape similar to that shown in Figure 3.3b, where the network scheduler only has enough bandwidth or time to schedule one transmission between n_1 and n_3, and decides accordingly. If it schedules for n_3 all the time, n_1 will never have access to the channel, so its capacity would be 0 in this situation. A similar reasoning for the time sharing of channel access shows that Π is the convex hull of all schedulable capacity rates at any given time. Figure 3.3b is an example of a network where the queue growth cannot be bounded. It is easy to prove that $\mathbf{G}_1$ is stable if and only if Λ lies in the interior of Π; that is,

$$\mathbb{E}\left[\lambda_1^2\right] \leq c_{12} \tag{3.3}$$

$$\mathbb{E}\left[\lambda_3^4\right] \leq c_{34} \tag{3.4}$$

If either (3.3) or (3.4) does not hold, then $\mathbf{G}_1$ is unstable. To prove this, consider if (3.3) is false. Then if we weaken $\max(\cdot, 0)$ in (3.1), the queue length dynamics become:

$$q_1(t+1) = q_1(t) - r_{12}(t) + \lambda_1(t) \tag{3.5}$$

The queue growth of n_1 can be defined as:

$$\Delta q_1(t) = q_1(t+1) - q_1(t) = \lambda_1(t) - r_{12}(t) \geq \lambda_1(t) - c_{12}(t) \tag{3.6}$$

If (3.3) is false, then $\Delta q_1(t)$ is a stochastic process with a positive mean. The process $q_1(t+1)$ can be rewritten as the summation of all instances of $\Delta q_1(t)$ from some initial time 0 to t. Since $q_1(t+1)$ includes the summation of random variables with positive means, the mean of $q_1(t+1)$ is the growing summation of positive terms. Over time, the expected queue length $\mathbb{E}[q_1(t+1)]$ will grow without bound and cannot converge to a finite limit.

$$\lim_{t\to\infty} \mathbb{E}[q_1(t+1)] = q_1(0) + \lim_{t\to\infty} \sum_{\tau=0}^{t} \mathbb{E}[\Delta q_1(\tau)] \quad \text{D.N.E.} \tag{3.7}$$

The approach above inspires a technique called "minimizing Lyapunov Drift" in designing optimal throughput networking strategies, first suggested by [1, Tassiulas and Ephremedes] and then expanded by [18, Neely] and other authors. The idea is to apply Lyapunov control theory to determining a scheduling rule for controlling the growth of queue lengths in a network.

A Lyapunov function is a non-negative monotonic and continuously differentiable scalar function of control variables. Conceptually, it is an *energy function* that can be used to determine a control rule to dissipate energy efficiently. For multivariable systems, the Lyapunov function is usually based on the of a norm or distance of the current system state variables from to a desired operating region. The quadratic function and the weighted quadratic functions are common Lyapunov functions used to determine control rules to drive the control variable x to 0. These functions are shown below:

$$L(x) = x^2 \tag{3.8}$$

$$L(\mathbf{x}) = \sum_i w_i x_i^2 \qquad w_i \geq 0 \quad \forall i \tag{3.9}$$

Manipulations of the weights w_i in conjunction with Little's law can be used for quality of service and application/flow control problems. Suppose we wish to use a Lyapunov function to determine a scheduling rule that keeps the queues in a network bounded. We select the scaled Frobenius norm of our network queue process $\mathbf{Q}(t)$ as a trivial choice of a Lyapunov function.

$$L\left(\mathbf{Q(t)}\right) = \frac{1}{2}\left\|\mathbf{Q}(t)\right\|_F^2 = \frac{1}{2}\sum_j \sum_i \left(q_i^j(t)\right)^2 \tag{3.10}$$

If we can determine a scheduling and rate allocation rule that keeps (3.10) bounded by some quantity, say K, for all time, we obtain a loose bound for the maximum queue length of the network:

$$L\left(\mathbf{Q(t)}\right) \leq K \Rightarrow \max_{i,j} q_i^j(t) \leq \sqrt{2K} \qquad \forall t \in \mathbf{T} \tag{3.11}$$

To determine such a rule, we use (3.1) to understand how $L\left(\mathbf{Q}\left(t\right)\right)$ grows with time. The Lyapunov drift function is defined as

$$\begin{aligned}
\Delta L\left(\mathbf{Q}(t)\right) &= L\left(\mathbf{Q}(t+1)\right) - L\left(\mathbf{Q}(t)\right) \qquad (3.12)\\
&= \frac{1}{2}\sum_j \sum_i \left[\max\left(q_i^j(t) - r_{ij}(t), 0\right) + \lambda_i^j(t)\right]^2 - \frac{1}{2}\sum_j \sum_i \left(q_i^j(t)\right)^2\\
&\leq \frac{1}{2}\sum_j \sum_i \left[\left(r_{ij}\right)^2 - 2q_i^j r_{ij} + 2\left|q_i^j - r_{ij}\right|\lambda_i^j + \left(\lambda_i^j\right)^2\right]\\
&= \frac{1}{2}\sum_j \sum_i \left[\left(r_{ij}\right)^2 + 2\left(q_i^j - r_{ij}\right)\lambda_i^j + \left(\lambda_i^j\right)^2\right] - \sum_j \sum_i q_i^j r_{ij}
\end{aligned}$$

where the terms in (3.12) have been separated into the difference of terms, which will increase and decrease the Lyapunov drift. Note that because packets are queued before they can be transmitted, and a node cannot send more packets than it has in its queue, $r_{ij} \leq q_{ij}$. If we follow the methodology in determining queue stability from (3.7), we are interested in developing a protocol that leads (3.12) to average out to 0 over time, so that

$$\begin{aligned}
\mathbb{E}\left[\Delta L\left(\mathbf{Q}(t)\right)\right] = 0 \Rightarrow \mathbb{E}\left[L\left(\mathbf{Q(t)}\right)\right] &= \mathbb{E}\left[L\left(\mathbf{Q}\left(0\right)\right) + \sum_0^t \Delta L\left(\mathbf{Q}(t)\right)\right] \qquad (3.13)\\
&= L\left(\mathbf{Q}\left(0\right)\right) + \sum_0^t \mathbb{E}\left[\Delta L\left(\mathbf{Q}\left(t\right)\right)\right]\\
&= L\left(\mathbf{Q}\left(0\right)\right)
\end{aligned}$$

In practice, it is very difficult to satisfy the condition $\mathbb{E}\left[\Delta L\left(\mathbf{Q}(t)\right)\right] = 0$ without a priori knowledge of network traffic or instantaneous backlog. Most

designers use known queue information to create a scheduling rule so that $\mathbb{E}\left[\Delta L\left(\mathbf{Q}\right)|\mathbf{Q}\right] \leq 0$ beyond some boundary point which we do not want the queue values to exceed. To see if such a condition can be achieved, we minimize (3.12) given the queue information, which is equivalent to maximizing the rightmost term of (3.12).

$$\underset{\mathbf{R}, r_{ij} \leq q_i^j, c_{ij}}{\operatorname{argmin}} \ \mathbb{E}\left[\Delta\left(\mathbf{Q}\right)|\mathbf{Q}\right] \Rightarrow \underset{\mathbf{R}, r_{ij} \leq q_i^j, c_{ij}}{\operatorname{argmax}} \sum_j \sum_i q_i^j r_{ij} \tag{3.14}$$

The maximization in (3.14) leads to the *backpressure scheduling* algorithm. If the network resource constraints or MAC allowed for at most 1 transmission during each timeslot, it should schedule for the link-session with the highest rate-backlog product. For $\mathbf{G}_1$, it requires the evaluation of

$$c_{12} \cdot q_1^2 \gtrless c_{34} \cdot q_3^4 \tag{3.15}$$

If c_{12} and c_{34} are comparable, then (3.15) simply prioritizes channel access to the node with a higher backlog. Consequently, many distributed MAC schedulers in the literature use only local queue information and that of their 1 hop neighbors [10, Ding et al.]. This process leads to an inherent control rule for queues. If not enough channel access has been allocated to a particular node in a network, its queue will grow, prompting more channel access. MAC schedulers for mobile ad-hoc networks are typically designed to greedily maximize according to (3.14).

3.2 Changing The Capacity Region

The scenario in Figure 3.4 assumes there is enough bandwidth and time to schedule transmissions for both nodes n_1 and n_3. Consider Figure 3.4a, when a particular inbound traffic rate exceeds the capacity of an individual link in the network. If a node could enlist the help of a nearby neighbor with excess resources, we have shown that the point-to-point transmission rate can change, thus the effective capacity region can change to encompass additional traffic rates. Figure 3.4b shows the change in capacity region if node n_3 supports node n_1 some percentage of its transmission cycle. If node n_3 relays for node n_1 all the time, then its effective capacity will reduce to

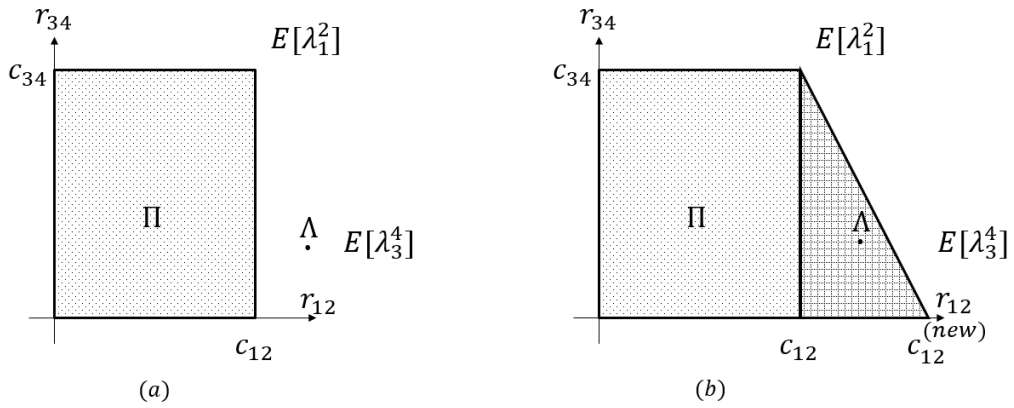

Figure 3.4: $\mathbf{G}_1$ Capacity Regions with Cooperative Transmission

0 and the capacity of n_1 will increase to $c_{12}^{(new)}$. Clearly, if node n_3 has any traffic, it should reserve some time, power and bandwidth to complete its own task. When should node n_3 support node n_1? A reasonable decision should address the following points:

- Node n_3 should not support node n_1 if $\mathbb{E}[\lambda_3^4]$ is at the boundary of Π, i.e. $\mathbb{E}[\lambda_3^4] \approx c_{34}$. While long term averages are difficult to compare practically, this can be inferred if the link activation rate/channel access probability is high. (If a node needs to transmit often to maintain a steady queue, it probably should not be additionally overloaded to support another.)

- Node n_3 should not support node n_1 if it has overtaxed its power constraint. Specifically, if node n_3 had to previously address a low latency or heavy traffic transmission and expended more power than it should have, it lacks the appropriate resources to support node n_1 without being outside of its intended mode of operation.

- Node n_3 should support node n_1 if it will yield a better expected quantifiable result for the network. Since the backpressure network prioritizes minimizing queues, when given a choice, node n_3 should evaluate

$$c_{12}^{(new)} \cdot q_1^2 \gtrless \left(c_{12} \cdot q_1^2 + c_{34} \cdot q_3^4\right) \tag{3.16}$$

  and schedule for the option that benefits the network best as a whole.

Interestingly enough, the last point above invokes another natural control rule to the network backlogs. Consider the case where q_1^2 gradually grows without bound, i.e. $\mathbb{E}\left[\lambda_1^2\right]$ has rendered $\Lambda \notin \Pi$. If $\mathbb{E}\left[\lambda_3^4\right]$ is sustainable, then q_3^4 remains finite. We have shown in (2.49) that if node n_3 wishes to assist node n_1 by relaying, any help it provides will result in a non-negative gain. Thus, at some point if the capacity region does not accommodate the network, the left-hand side of (3.16) should be greater than the right-hand side, and node n_3 will support node n_1. When node n_3 has supported node n_1 sufficiently to have comparable queues, it will schedule for itself again. The first and last points lead to the following cooperation decision rule:

$$\mathbb{E}\left[\Delta c_{12}\right] \cdot q_1^2 \underset{DoNotSupport}{\overset{Support}{\gtrless}} c_{34} \cdot q_3^4 \tag{3.17}$$

where $\mathbb{E}\left[\Delta c_{12}\right]$ is inferred from (2.48) and the exchange of handshaking messages (to be described in a later section). To address the second point, we first consider the quality of the 2-hop relay link. From (2.27), we know that the $\gamma_{132} = \mu_H\left(\gamma_{13}, \gamma_{32}\right)$. Because the harmonic mean is dominated by the minimum of its arguments, to prevent further degradation of the signal, n_3 should allocate enough power to match γ_{32} to γ_{13}. To control excessive power expenditure supporting as a relay, we adopt the concept of generalized queueing from [18, Neely] and [19, Supittayapornpong and Neely]. Suppose n_3 is subject to an average power constraint $\bar{P}_3$. This constraint is trivially satisfied if at each time node n_3 schedules with $P_3 \leq \bar{P}_3$. Since we are considering average power expenditure, node n_3 is allowed to violate this constraint occasionally. If we queue up the violations of this power constraint, then we can determine a queue control rule that we can apply to power. Define the *power queue* as

$$W_i\left(t+1\right) = \max\left[W_i\left(t\right) + \left(P_i\left(t\right) - \bar{P}_i\right), 0\right] \tag{3.18}$$

If $W_3 \neq 0$, we simply do not allow the option for node n_3 to support as a relay. This power queue will naturally dissipate if node n_3 stops scheduling, indicating low amounts of incoming traffic.

3.3 Media Access Control for Cooperation

In this section, we briefly discuss the properties of a distributed MAC scheduler to function in a backpressure network with cooperation. We assume that every node has information on its own queue, as well as long term average statistics on the channels to their 1-hop neighbors σ_{ij}^2. This can be determined through collection on the exchange of control messages, as well as known environment information inferred with other sensors ([10, Ding et al.] refers to one such process as "Collaborative Virtual Sensing").

To consider the latency of time sensitive or priority packets in the network, we consider Little's law, which states that the average length of a queue is directly proportional to the average latency on the delivery of packets and the arrival rate. In the networks we are discussing, we can consider an average delay requirement to satisfy some quality of service. In the variables we have defined so far, Little's law states

$$E\left[q_i^j(t)\right] = E\left[\lambda_i^j\right] D_i^j \Longrightarrow D_i^j = \frac{E\left[q_i^j(t)\right]}{E\left[\lambda_i^j\right]} \tag{3.19}$$

where D_i^j denotes the average delay of information from n_i to n_j. Because packets need to be queued before they can be transmitted, D_i^j is lower bounded by 1. Different applications have different delay requirements and because of Little's law, the delay constraints can be met by control of the average queue length. Because queues can only diminish with the transmission of packets, channel access and resources should be allocated to meet delay constraints. Specifically, for n_i, the probability of channel access should approach 1 as q_i^j approaches $q_i^{j*} = E\left[\lambda_i^j\right] d_i^j$, where d_i^j is the latency requirement of information from node n_i to n_j. Conversely, probability of channel access should approach 0 as q_i^j approaches 0.

To create a distributed MAC scheduler with the considerations above, each backlogged node determines a contention window timer to begin a handshaking protocol. Nodes with a higher backlog than their application requirements should be given priority. The following contention window is commonly used to determine a backoff counter:

$$CW = 2^{\beta \cdot \max_j \left(\frac{q_i^j(t)}{q_i^{j*}} \right)} + \alpha \tag{3.20}$$

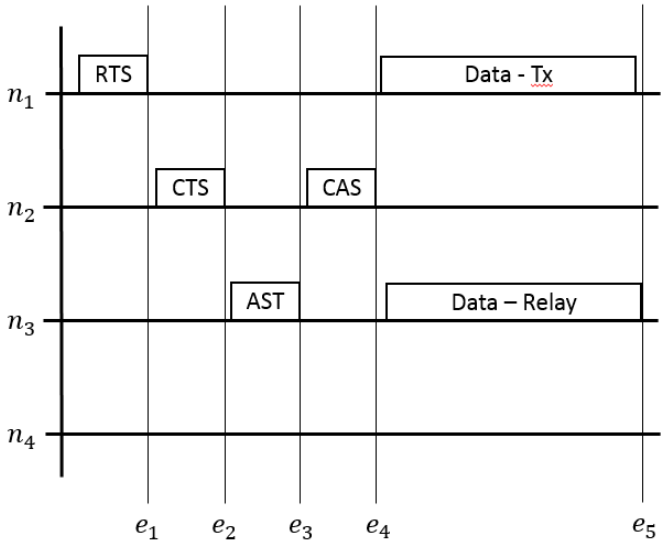

Figure 3.5: Handshaking Between Nodes with Cooperation in $\mathbf{G}_1$

for some $\alpha \geq 0$ and $\beta < 0$. A backoff counter can be generated by multiplying the contention window by a uniform random variable distributed between 0 and 1. Nodes will defer channel access for the duration of their backoff counter prior to attempting to schedule transmission. Using this method, nodes with greater backlog are likely granted channel access earlier than less backlogged nodes. When a node's backoff counter has expired, it will schedule a transmission with handshake messages as shown in Figure 3.5.

In this instance, node n_1 has an expired backoff timer and it would like to transmit to node n_2. It begins the handshaking coordination with a request-to-send (RTS) which has information on node n_1's current backlog and a reference transmission power. At event e_1, all nodes in the network have received node n_1's RTS and determined the channel gain h_{1j} from the quality of the signal. If node n_2 is available for reception, it will respond with a clear-to-send (CTS) message which includes node n_2's backlog and inferred value of h_{12}. Consequently, by event e_2, node n_3 has enough information to compute (2.48). Using that result, node n_3 evaluates (3.17) and determines whether it should cooperate with node n_1's transmission. If node n_3 chooses to cooperate, it will send out an assist (AST) message with information on h_{13} and relaying parameters (bandwidth, frequency, power, etc.). Node n_2 responds with a confirm-assist (CAS) message which includes information on

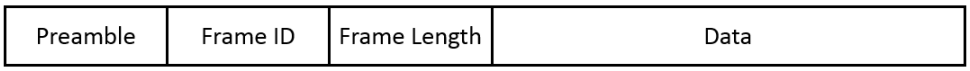

Figure 3.6: Frame Structure for Cooperative Transmission

h_{32}. By event e_4, node n_1 has collected information on h_{12}, h_{13}, and h_{32} and can determine the channel conditions and select the appropriate coded rate of transmission.

3.3.1 Packet Designs

Figure 3.6 shows a proposed frame structure for the sending and recombination of messages from different diversity branches

The preamble at the beginning of the frame is used for detection, as well as frame synchronization. If the message is being relayed through a supporting node, there is an inherent latency due to a longer path traversal and the diversity branches need to be synchronized in order to avoid self-interference through combination. The frame id field is used to assist in latency estimation between different diversity branches and the combination of correct frames. The length field is used to determine where the frame ends to terminate reception and combining of messages. We assume the length of the frame is selected appropriately to stay within the coherence time of the channel.

3.4 Simulation of a Backpressure Network with Cooperation

Figures 3.7-3.10 shows the results of two simulations carried out in GNU Octave. In both simulations $\omega = 1$, $\sigma^2 = 1$, and the arrival rates are modeled as Poisson processes with $\Lambda = [4 \quad 1]$; however, γ_0 was selected to be .015. The corresponding achievable capacity for each node is just under 3 bits/Hz, so in general, node n_1 cannot stabilize its queue. The simulation in Figure 3.7 shows the queue growth of network $\mathbf{G}_1$ when nodes are not allowed to cooperate with each other as relays. The simulation in Figure 3.8 shows the same exact settings as in Figure 3.7; however, in this simulation nodes

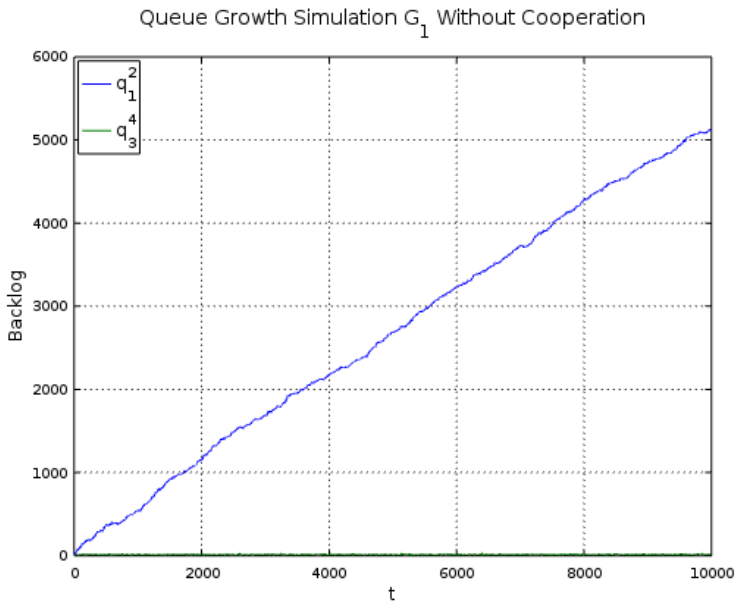

Figure 3.7: Queue Growth of $\mathbf{G}_1$ when Cooperation is Not Allowed

are allowed to cooperate as cooperative amplify-and-forward relays. The simulation was run for a duration of 10000 time steps. In the scenario where nodes were not allowed to cooperate, q_1^2 seems to grow without bound, while q_3^4 was comparatively low because n_3 did not generate as much traffic as n_1. This is indicative that n_3 may have excess resources to support n_1. In the scenario where nodes were allowed to cooperate, n_3 periodically supported n_1 in reducing its backlog, maintaining a stable backlog for q_1^2 of about 100. Figure 3.9 shows the power expenditure of nodes n_1 and n_3, and the high spikes from n_1 are an indication that it is having a difficult time getting rid of packets in its queue. Figure 3.10 shows a steady curve that tabulates the instances that one node supports another (in the cooperative scenario shown in Figure 3.8). The slope in Figure 3.10 shows that n_3 supports n_1 roughly 20% of the time, allowing the remaining time for its own transmissions while providing the necessary help to keep q_1^2 stable. It is visibly clear here that a backpressure network which allows nodes to cooperate as amplify-and-forward relays is more flexible than the same network without cooperation.

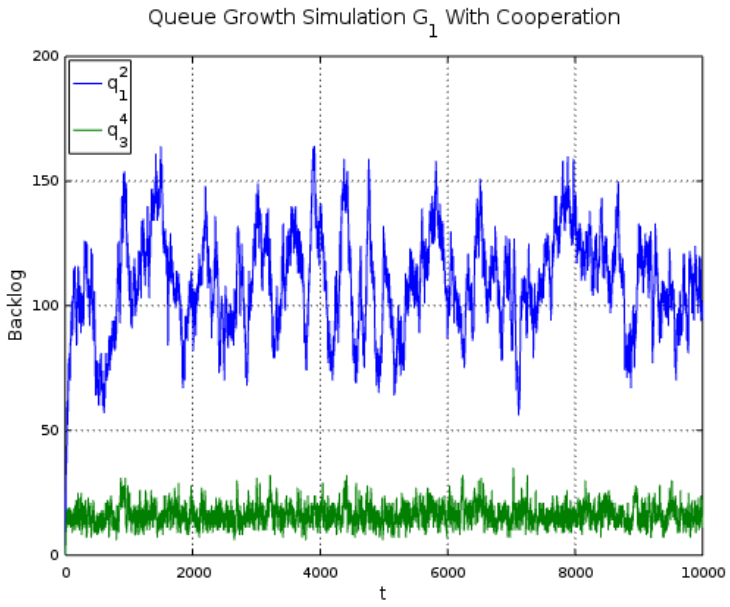

Figure 3.8: Queue Growth of $\mathbf{G}_1$ when Cooperation is Allowed

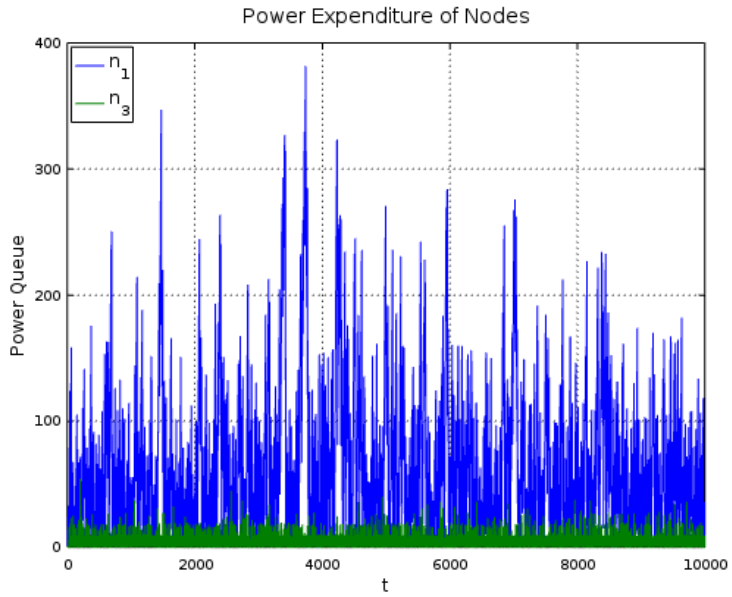

Figure 3.9: Power Queue of Nodes n_1 and n_3 when Cooperation is Allowed

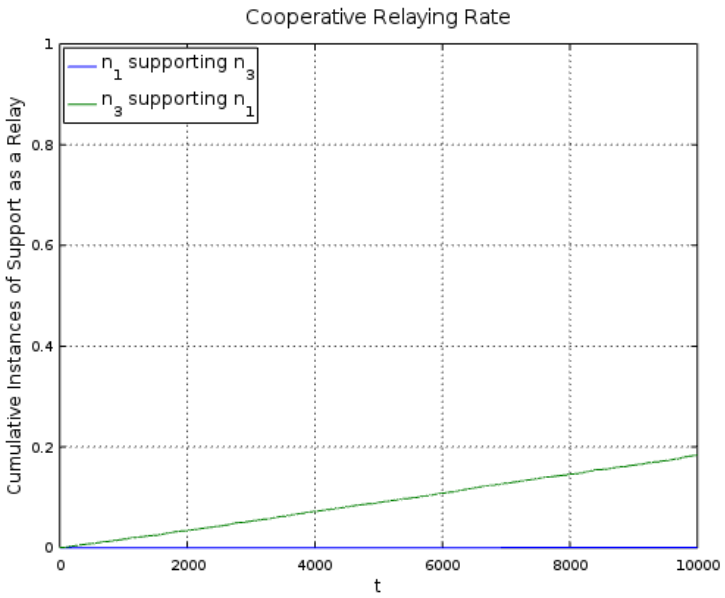

Figure 3.10: Rate of Cooperation Between Nodes n_1 and n_3

CHAPTER 4

CONCLUSIONS

This thesis has shown that the capacity regions of backpressure networks can be further augmented with cooperation models in heterogeneous networks. While cooperative diversity has been studied to reduce outage probability or bit error rate in a physical layer perspective, comparatively little has been investigated on the impact cooperation can have from a network perspective. By judiciously tasking less occupied nodes to support more occupied nodes with an amplify-and-forward relaying technique, we have shown a more flexible and responsive capacity region.

4.1 Capacity of Backpressure Networks with Cooperation

Backpressure Networks have been regarded as throughput optimal networks because of their ability to accommodate more arrival rates than any other scheduling policy while maintaining bounded queue lengths. To consider more general theory in practical networks, backpressure scheduling was then applied to differential backpressure routing for multihop networks, providing queue stability rules for end-to-end delivery of packets.

Differential backpressure routing has been known to have end-to-end delivery rate that is limited by the maximum quantity of information traversing the weakest cut in a multihop route. This shows a natural limitation on the capacity region of backpressure networks. By allowing for an amplify-and-forward relaying protocol, weak cut links can be augmented by resources available from less occupied nodes. This sort of capacity region morphing protocol is sufficiently flexible to include legacy software-based devices, to support newer generation devices, and can naturally react to changing application requirements in a network.

REFERENCES

[1] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, Dec 1992.

[2] D. G. Brennan, "Linear diversity combining techniques," *Proceedings of the IRE*, vol. 47, no. 6, pp. 1075–1102, June 1959.

[3] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, Sep 1979.

[4] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec 2004.

[5] M. O. Hasna and M. S. Alouini, "Optimal power allocation for relayed transmissions over Rayleigh-fading channels," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1999–2004, Nov 2004.

[6] W. Su, A. K. Sadek, and K. J. Ray Liu, "Cooperative communication protocols in wireless networks: Performance analysis and optimum power allocation," *Wireless Personal Communications*, vol. 44, no. 2, pp. 181–217, 2008. [Online]. Available: http://dx.doi.org/10.1007/s11277-007-9359-z

[7] A. S. Ibrahim, A. K. Sadek, W. Su, and K. J. R. Liu, "Cooperative communications with relay-selection: when to cooperate and whom to cooperate with?" *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2814–2827, July 2008.

[8] R. A. M. Khan and H. Karl, "MAC protocols for cooperative diversity in wireless LANs and wireless sensor networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 46–63, March 2014.

[9] Y. Ding and M. Uysal, "Amplify-and-forward cooperative OFDM with multiple-relays: performance analysis and relay selection methods," *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, pp. 4963–4968, Oct 2009.

[10] L. Ding, T. Melodia, S. N. Batalama, J. D. Matyjas, and M. J. Medley, "Cross-layer routing and dynamic spectrum allocation in cognitive radio ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1969–1979, May 2010.

[11] J. Liu, A. Eryilmaz, N. B. Shroff, and E. S. Bentley, "Heavy-ball: A new approach to tame delay and convergence in wireless network optimization," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, Apr 2016, pp. 1–9.

[12] M. J. Neely, "Optimal backpressure routing for wireless networks with multi-receiver diversity," in *2006 40th Annual Conference on Information Sciences and Systems*, March 2006, pp. 18–25.

[13] E. M. Yeh and R. A. Berry, "Throughput optimal control of cooperative relay networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3827–3833, Oct 2007.

[14] J. W. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *Military Communications Conference, 1991. MILCOM '91, Conference Record, Military Communications in a Changing World., IEEE*, Nov 1991, pp. 571–575 vol.2.

[15] M. O. Hasna and M. S. Alouini, "Performance analysis of two-hop relayed transmissions over Rayleigh fading channels," in *Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. 2002 IEEE 56th*, vol. 4, 2002, pp. 1992–1996 vol.4.

[16] M. Abramowitz, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover Publications, Incorporated, 1974.

[17] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, Nov 1997.

[18] M. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan Claypool, 2010. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6813406

[19] S. Supittayapornpong and M. J. Neely, "Time-average stochastic optimization with non-convex decision set and its convergence," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on*, May 2015, pp. 490–497.