ABSTRACT

Title of dissertation: TOWARD LAYERLESS COOPERATION AND RATE CONTROL IN WIRELESS MULTI-ACCESS CHANNELS

Beiyu Rong, Doctor of Philosophy, 2010

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

In wireless networks, a transmitted message may successfully reach multiple nodes simultaneously, which is referred to as the Wireless Multicast Advantage. As such, intermediate nodes have the ability to capture the message and then contribute to the communication toward the ultimate destination by cooperatively relaying the received message. This enables cooperative communication, which has been shown to counteract the effects of fading and attenuation in wireless networks. There has been a great deal of work addressing cooperative methods and their resulting benefits, but most of the work to date has focused on physical-layer techniques and on information-theoretic considerations. While compatible with these, the main thrust of this dissertation is to explore a new approach by implementing cooperation at the network layer.

First, we illustrate the idea in a multi-hop multi-access wireless network, in which a set of source users generate packets to deliver to a common destination. An opportunistic and dynamic cooperation protocol is proposed at the network

level, where users with a better channel to the destination have the capability and option to relay packets from users that are farther afield. The proposed mode of cooperation protocol is new and relies on MAC/Network-level of relaying, but also takes into account physical-layer parameters that determine successful reception at the destination and/or the relay. We explicitly characterize the stable throughput and average delay performance. Our analysis reveals that cooperation at the network layer leads to substantial performance gains for both performance metrics.

Next, on top of the network-layer cooperation, we investigate enhanced cooperative techniques that exploit more sophisticated physical-layer properties. Specifically, we consider dynamic decode-and-forward, superposition coding, and multi-packet reception capability, and we quantify the extent to which the enhancement techniques can further improve the stable throughput region. Then we revert back to the two-user multi-access channel with single-packet reception, which has been extensively studied in the case of no cooperation. After cooperation is permitted between the two users, we revisit the relationship between the *stability region* and the *throughput region* under both scheduled access and random access schemes.

Finally, we shift our focus from the packet-level to bit-level multi-access channels. By exploiting the bit-nature of a packet, we create a bridge between traditional physical-layer-based transmission rates and classical MAC/Network-layer-based throughput rates. We first obtain the closed form of the stability region in bits/slot. Then, as a separate, but related issue, we look at the minimum delivery time policy; for any initial queue size vector, the optimal policy that empties all bits in the system within the shortest time is characterized.

TOWARD LAYERLESS COOPERATION AND RATE CONTROL IN WIRELESS MULTI-ACCESS CHANNELS

by

Beiyu Rong

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
2010

Advisory Committee:
Professor Anthony Ephremides, Chair/Advisor
Professor Michael Fu
Professor Richard J. La
Professor Prakash Narayan
Professor Lawrence C. Washington, Dean's Representative

Acknowledgments

My deepest gratitude goes to my advisor, Professor Anthony Ephremides, for his enormous help and support during my Ph.D. study in Maryland. He is the person who led me to the road of research. His way of doing research and his enthusiasm in both research and life will continue to influence me. I am so fortunate to be advised by an extraordinary person like him.

I would like to thank Professor Ioannis Krikidis for giving me the opportunity to collaborate with him. I am much grateful to Professor Jie (Rockey) Luo, for his useful advice and continuous encouragement.

I owe my gratitude to all people who have helped me and made this dissertation possible, including my friends, my officemates, and so many others that I am unable to list their names here.

Table of Contents

1 Introduction 1
- 1.1 Cooperative Communication . . . 2
  - 1.1.1 Network-Layer Cooperation . . . 2
- 1.2 Rate Control . . . 5
- 1.3 Performance Measures . . . 7
- 1.4 Outline of the Dissertation . . . 9

2 Network-Level Cooperative Access: Stable Throughput and Delay Analysis 10
- 2.1 Introduction . . . 10
- 2.2 Model . . . 13
  - 2.2.1 Cooperation Strategy . . . 14
  - 2.2.2 Queue Stability . . . 17
- 2.3 Cooperative Work-Conserving Policy . . . 18
  - 2.3.1 Stable Throughput Region . . . 20
  - 2.3.2 Delay Minimizing Policy . . . 21
  - 2.3.3 Average Delay Characterization . . . 25
- 2.4 Cooperative TDMA policy . . . 29
  - 2.4.1 Stable Throughput Region . . . 30
  - 2.4.2 Average Delay Characterization . . . 31
- 2.5 Numerical Results . . . 33
- 2.6 Discussion . . . 35
- 2.7 Appendix . . . 37
  - 2.7.1 Proof of Theorem 2.1 . . . 37
  - 2.7.2 Proof of Theorem 2.2 . . . 43
  - 2.7.3 Proof of Proposition 2.1 . . . 45
  - 2.7.4 Proof of Theorem 2.4 . . . 50
  - 2.7.5 Proof of Proposition 2.2 . . . 53

3 Enhanced Cooperation Based on Physical-Layer Techniques 56
- 3.1 Introduction . . . 56
- 3.2 DDF and Superposition Coding . . . 58
  - 3.2.1 Model . . . 58
  - 3.2.2 Dynamic Decode-and-Forward . . . 61
    - 3.2.2.1 Non-Cognitive DDF (NC-DDF) . . . 64
    - 3.2.2.2 Cognitive DDF (C-DDF) . . . 67
  - 3.2.3 Superposition Coding . . . 69
    - 3.2.3.1 Conventional Cooperation with Superposition (S-CC) 71
    - 3.2.3.2 Cognitive DDF with Superposition (SC-DDF) . . . 72
  - 3.2.4 Numerical Results . . . 73
- 3.3 Multipacket Reception Capability . . . 75
  - 3.3.1 Model . . . 75
  - 3.3.2 Opportunistic Cooperation Scheme . . . 78

3.3.3 Numerical Results . . . 89
3.4 Discussion . . . 92

4 Stability and Throughput Regions for Cooperative Multi-Access 94
4.1 Introduction . . . 94
4.2 Model . . . 96
4.3 Scheduled Access . . . 99
4.3.1 Stability Analysis . . . 101
4.3.2 Throughput Analysis . . . 101
4.3.3 Optimal Policy for the Throughput Region . . . 107
4.4 Random Access . . . 107
4.4.1 Stability Analysis . . . 108
4.4.2 Throughput Analysis . . . 111
4.4.3 Optimal Policy for the Throughput Region . . . 116
4.5 Network Coding at the Relay Node . . . 118
4.6 Numerical Results . . . 120
4.7 Discussion . . . 122
4.8 Appendix . . . 123
4.8.1 Proof of Theorem 4.4 . . . 123

5 Rate Control 131
5.1 Introduction . . . 131
5.2 Model . . . 134
5.3 Stability Region . . . 137
5.4 Minimum Delivery Time Policy . . . 140
5.5 Discussion . . . 145
5.6 Appendix . . . 146
5.6.1 Proof of Theorem 5.1 . . . 146
5.6.2 Proof of Lemma 5.1 . . . 149

6 Conclusion 155
6.1 Summary of Contributions . . . 155
6.2 Future Directions . . . 158

Bibliography 160

Chapter 1

Introduction

Traditional analysis and implementation of wireless networks still view the network as a collection of point-to-point links. However, this approach is not appropriate, and in most cases, not accurate. Unlike the wired counterpart where links are well established, the connection between any pair of nodes in wireless networks depends on a lot of factors such as transmission power, distance, fading, target bit-error-rate, and interference from concurrent transmissions. In addition, if we assume isotropic medium and omnidirectional antennas, a transmitted message is capable of being received by multiple nodes simultaneously, which is referred to as the "Wireless Multicast Advantage" [1]. In light of these facts, extraordinary efforts have been made on "cross-layer" design, which links different layers but still treats them separately. In this dissertation, we aim at a "layerless" approach that "fuses" the physical, MAC and network layers.

In a wireless multi-access system, user nodes share the common channel. Their transmissions are not only interfering, but also, users are capable of overhearing each other's transmission. Thus, the multi-access channel exhibits almost every facet for the basic understanding of wireless communications and the ultimate performance limits.

1.1 Cooperative Communication

As implied by the wireless multicast advantage, intermediate nodes have the ability to capture the transmission and then contribute to the communication toward the destination by cooperatively relaying the data packets. Cooperative communication exploits spatial diversity inherent to wireless channels, which is an important tool to counteract the effects of fading, shadowing and attenuation. Initially, spatial diversity was achieved in single communication links via the use of multiple antennas [2, 3]. However, in wireless networks, the effect of multiple antennas can be realized through the combined use of the antenna resources that actually reside at different nodes.

The notion of cooperative relaying has received a great deal of attention. Most of it, so far, has focused on the physical layer and on information-theoretic considerations [4,5,6,7,8,9,10,11]. While compatible with these, our work in this dissertation focuses on simply and intelligently implementing cooperation at the network layer, while taking into account the physical layer properties and realizations as well as the MAC layer protocols.

1.1.1 Network-Layer Cooperation

A number of cooperative signaling methods have been addressed in physical-layer studies, including decode-and-forward (DF), amplify-and-forward (AF), coded cooperation, compress-and-forward (CF). In contrast to these sophisticated signaling methods, our work considers only "reliable forwarding", in the sense that a node

will only forward the packet from another node if the packet is successfully decoded, the process of which is independent from the source's direct link to the destination. We will show that a simple and practical forwarding protocol can actually lead to significant performance improvement if the cooperation strategies are properly designed.

In this part of work for network-layer cooperation, the communication unit is considered to consist simply of a packet, without regard to its bit-content. If a transmission is successful, the entire packet is considered to be decoded without error; otherwise, the packet is not successfully decoded and is discarded. We assume a channel with fading and attenuation which also includes the effects of additive white Gaussian noise, and is shared by all transmitting nodes. When node i transmits a packet with power P, the probability that node j can successfully decode the packet is given by

$$p_{i,j} = \mathbf{P}\left[\frac{|h_{i,j}|^2 P}{P_{int} + N_0} > \gamma\right] \tag{1.1}$$

where $h_{i,j}$ is the channel gain from node i to node j, which can either be distance-dependent only as in free-space communication or can include the effects of different types of fading represented by a random variable. As usual, the fading process is assumed to remain constant during one slot, but may change independently from one slot to another while retaining the same distribution. The quantity P_{int} is the received power at receiver j resulting from transmissions other than from node i, and N_0 is the noise power level at j. The parameter γ is the signal-to-interference-

plus-noise ratio (SINR) threshold required for correct decoding[1].

The adopted SINR criterion is a useful tool for incorporating the physical-layer properties into the network-layer analysis. This probabilistic reception channel modeling in principle allows the communication between any pair of nodes in the network, but with obviously different levels of likelihood based on the attenuation of the signal.

At the network layer, cognitive cooperation has attracted increasing attention for its efficient utilization of wireless network resources [12, 13, 14]. Basically, cognitive relaying is enabled only when the relay node senses an idle time slot (or other channel resource) of the primary user and hence transmits a relayed packet in that slot. However, our work in this dissertation reaches far beyond the idea of cognitive cooperation. First, without introducing extra "pure" relay to assist the transmission, all users in the network can help relay packets for others. We consider the general case where all users may have their own information data to be delivered to the destination while also relay information of other users. Second, instead of allowing cooperation only during the idle slots, we consider a variety of medium access schemes amongst users. Both conflict-free scheduled access and random access schemes are investigated. Third, the cooperation in our study involves a multi-hop relaying concept. In addition, the relaying pattern is not fixed but dynamic, depending on instantaneous channel outcomes.

Although full-duplex mode is possible (but at an increased implementation

[1]This SINR criterion is fully accurate for Gaussian signals and noise, and increasingly accurate as the number of interfering signals increases (no matter what their statistics are) by the cental limit theorem.

cost), in our work, we assume that the node cannot transmit and receive at the same time. As such, the intermediate node will have to remain silent if it wants to overhear a transmission. A question that naturally arises is, whether the intermediate user nodes will suffer performance loss as they waste time slots for capturing and relaying packets from other users. A remarkable result from our work is that, by opportunistically choosing the relaying node, all user nodes in the network can simultaneously achieve performance gains. The advantages and gains of cooperation are mainly evaluated in terms of two performance metrics: (i) stable throughput region and (ii) average delay. We observe that the improvement is not solely due to cooperative diversity which is the case with information-theoretic studies, but rather it is also due to the concentration of packets from disparately distributed queues into fewer virtual queues. This is to be revealed and discussed in the next chapter.

1.2 Rate Control

As implied by Eq. (1.1), the channel reception quality depends on the value of γ, which is an increasing function of the transmission bit-rate. If we encode more bits into a packet, the transmission is less sustainable under fixed channel condition and interference; on the other hand, higher packet throughput can be achieved by lowering the number of bits in a packet. Therefore, it is more appropriate and accurate to formulate the problem at the *bit* level, and evaluate the performance metrics such as bit throughput and bit delay.

In a multi-access channel, given that there are multiple users, the feasible

transmission bit-rate depends on whether other users transmit simultaneously or not. In this part of dissertation, we extend our study in multi-access channels from the packet reception to more realistic channel model, by exploring the bit-nature of a packet. As a first step, we restrict our attention to the two-user case. Then, a trade-off arises as to whether to activate one transmission at a time with higher instantaneous transmission rate (measured in bits/slot), or to let both users transmit simultaneously but at lower individual instantaneous rates, to combat interference and ensure successful reception at the destination.

The balance of such trade-off is illustrated through two performance metrics: stable throughput region and total delivery time. The two performance metrics are obtained from different assumptions and have different meanings, but are related to each other. The optimal rate control policies with regard to both performance metrics are explicitly characterized in Chapter 5. Our results indicate that the optimal policies for achieving different objectives vary and depend on the values of some crucial system parameters.

There are two forms of "rate" measures: (i) the traditional throughput region, which is based on a networking perspective and is measured in packets per time-unit (slot), and (ii) the Shannon capacity region, which is based on an information-theoretic perspective and is expressed in bits per time-unit. Our bit-level multi-access channel study is, in a sense, "layerless", and combines the two perspectives (to some degree) and is, thus, a step towards the ultimate goal of unifying information theory and networking theory [15].

1.3 Performance Measures

A common assumption with information-theoretic studies and many other networking approaches is that source nodes are always backlogged with information bits. The corresponding capacity limits mostly studied are the Shannon capacity region and the maximum throughput region respectively. The former one, which does not fall into the scope of this dissertation, is the usual Shannon-theoretic region; it can be sometimes characterized when symbol length and delay are allowed to approach infinity. The latter one measures the maximum rates that can be communicated by assuming saturated queues. Such saturated queue assumption neglects the bursty traffic nature of a real network. In this dissertation, the emphasis will be given on the *stable throughput region*. The stable throughput region, also called the stability region, is a "rate" measure based on the networking perspective under the assumption of bursty arrivals. It quantifies the maximum rates sustainable by the network while ensuring that all queues remain stable – i.e., the queues are "infinitely often" empty[2]. We will show in the following chapters that the traffic burstiness is crucial for the performance improvement of our network-layer cooperation amongst different users. With the proposed cooperation protocol, theoretical analysis proves that the stable throughput region is substantially improved; in other words, all users in the network can simultaneously increase their stable throughput rates.

Despite the fact that the stability region and the throughput region are defined with different assumptions and have different impacts, there is strong evidence that

[2] A precise technical definition of queue stability is provided in [16].

the two regions are identical in multi-access channels, under a variety of channel and traffic models [17, 18, 19, 20, 21]. If it is true, it supports the conjecture that the empty state of the queue is immaterial in determining stability. The relationship of these two regions has been extensively studied in the case of no cooperation. In this dissertation, we would like to revisit it after cooperation is permitted. As there are multiple data streams at the intermediate node, it is not immediately clear whether the priority policies assigned over these streams will affect the result.

Queue stability in the cases we study implies finite queueing delay. With infinite reservoir of bits or packets, only the transmission delay is accounted in the delay performance. However, the queueing delay can be a significant component of overall delay if the information data is assumed to arrive randomly at the source user. The total delay of a packet is counted from the epoch it enters the system to the epoch it is delivered to the destination, which includes the queueing delay at multiple intermediate nodes the packet has stayed with.

In addition, we address another notion of "delay" in the rate control problem: the total delivery time. Given that an initial amount of traffic volume is located at the sources at a given time, and if there are no more arrivals, the minimum delivery time policy is defined as the rate allocation policy that drains all bits in the network within the shortest time.

1.4 Outline of the Dissertation

The first part of this dissertation is devoted to cooperative communication at the network layer. In Chapter 2, we present a new cooperation concept in a general wireless multi-access system, where there are a finite number of users and a common destination. We provide the option to users that have a better user-destination channel to relay packets from those that are farther afield; as a result, almost all users in the network behave as both user and relay. We consider conflict-free, work-conserving transmission policies as well as plain Time Division Multiple Access (TDMA) policy. For the case of an erasure channel, the stable throughput and delay performance under both medium access control policies are investigated. Chapter 3 deals with the enhancement of the network-layer cooperation through the exploitation of sophisticated physical-layer techniques. Then, in Chapter 4, we revisit the relationship between the stability region and the throughput region in the context of a network-layer cooperative multi-access channel, under a centralized scheduled access and a random access scheme respectively. The second part of this dissertation aims to address the subtle interactions between bits and packets in multi-access channels. By associating the transmission rates with the channel access policies, the achievable stability region (measured in bits/slot) in a random access system is derived; in addition, the optimal policy to empty all bits in the minimum time is explicitly characterized. This is presented in Chapter 5. Finally, Chapter 6 summarizes the contributions of this dissertation and points out possible future research directions.

Chapter 2

Network-Level Cooperative Access: Stable Throughput and Delay Analysis

2.1 Introduction

Previous work in cooperative communication has mainly focused on the physical layer and on information-theoretic considerations [4, 5, 6, 7, 8, 9, 10, 11]. In these settings, information bit streams have been modeled as continuous data flows, while the rate regions have been defined as Shannon rate regions which can be sometimes characterized when symbol length and packet delay are allowed to approach infinity. Some recent work, however, has implemented cooperation at the network protocol level, and performance gains in terms of (stable) throughput, delay and energy efficiency were demonstrated. In [12, 13, 14], cognitive cooperation was enabled by considering the bursty nature of real source users, in which the cognitive relay transmitted the packets relayed from the source during the idle time slots.

In this chapter, we study a more general relaying tactic at the network layer by exploiting cooperation amongst source users. We consider a wireless multi-access system which consists of N source users and a single common destination. Instead of allowing for only the N point-to-point links between the sources and the destination, here, any node pair in the network may form a direct communication link as

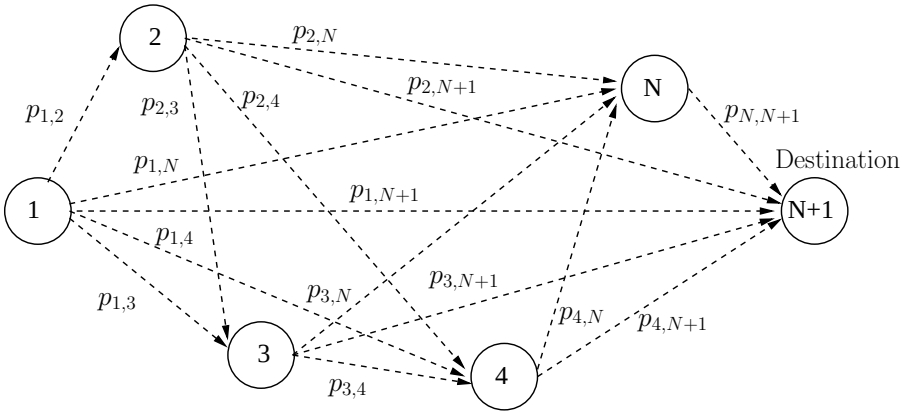

Figure 2.1: Wireless multi-access system with packet-erasure channels. A total of N users, indexed by $\{1, 2, \ldots, N\}$, transmit unicast traffic to the common destination node $(N+1)$. The channel reception probabilities are denoted above the channels.

depicted in Fig. 2.1. The N source nodes are ordered in the following way: node 1 is the "farthest" node from the destination node $(N+1)$, and every subsequent node i is progressively "closer" to $(N+1)$, and node N is the "closest" to $(N+1)$. This distance notion can either be represented as real distance in free-space communication, or "virtual" distance that also includes all types of fading. Therefore node i is considered to have a better channel to $(N+1)$ than $(i-1)$ does. Thus, we provide the option to node i to relay packets of its predecessor nodes $1, 2, \ldots, (i-1)$ in lieu of its own packets if the transmission by these predecessor nodes is unsuccessful but correctly received by node i. Specifically, the cooperation idea is as follows: when node i transmits a packet, if the packet is not successfully decoded by the destination, it may still be decoded by some subsequent source nodes from $(i+1)$ to N; then a selection process is activated and the node among the decoding set with the best source-destination channel will relay that packet, while all other nodes drop that packet; if none of i's subsequent nodes or the destination node decodes the packet, node i will retransmit that packet later on.

Due to the subtle complexity incurred by cooperation, we restrict our attention to conflict-free scheduling only; that is, we do not allow contention. Basically, we consider two types of MAC policy (i.e. transmission scheduling at the source nodes): conflict-free work-conserving policies and Time Division Multiple Access (TDMA) policy. Under the work-conserving policies, we characterize the stable throughput region and establish that all work-conserving policies following the proposed cooperation strategy have the same stable throughput region. This region strictly contains the stable throughput region when cooperation is not permitted. The backpressure algorithm introduced in [22] was proved to achieve the maximum stable throughput region, and it must be conflict-free and work-conserving; but in general, not all conflict-free work-conserving policies are optimal. Our work exhibits a special case where all conflict-free work-conserving policies result in the same region. Subsequently, the optimal work-conserving policy that minimizes the average delay over all packets in the system is determined among all cooperative work-conserving policies; and in the special case of two source users, where the user with a relatively weak channel to the destination has higher priority to transmit, we derive the closed-form expressions for the average delay experienced by each user's packets. Our results indicate that cooperation reduces the delay substantially for all users in the network. Under the TDMA policy, both the stable throughput region and average delay are characterized; likewise, we show that cooperation leads to significant improvement for both performance indices. Finally, we characterize the effect of inter-user channel quality on performance, and show that the gain in performance through cooperation increases as the inter-user channel quality improves.

2.2 Model

We consider the slotted wireless multi-access system shown in Fig. 2.1. N source nodes, indexed by $1, 2, \ldots, N$, transmit unicast traffic to a common destination node $(N + 1)$. The sources are equipped with a buffer of infinite capacity to store the packets. Traffic burstiness is considered by modeling the arrival process for each source node i as a Bernoulli process with rate λ_i ($i \in \{1, 2, \ldots, N\}$). These processes are independent from node to node and i.i.d (independently and identically distributed) over time slots. In the following, we denote i-th source's packets as class-i packets, so there is a total of N classes of packets flowing in the network. The channel is slotted, and the transmission of a packet takes the duration of exactly one time slot.

We adopt the packet-erasure channel model as described in Section 1.1.1. Since we assume conflict-free scheduling only at the MAC layer, corresponding to any node pair (i, j), we denote by $p_{i,j}$ the probability that node j successfully decodes the packet transmitted by node i. Further, the channel reception processes are assumed to be independent across all receivers. As mentioned in Section 2.1, node i has a better channel to the destination than $(i-1)$ does, and hence, has a higher successful delivery probability. Therefore, we have $p_{N,N+1} > p_{N-1,N+1} > \ldots > p_{1,N+1}$. As usual, we assume that acknowledgements (ACKs) are instantaneous and error-free, and all nodes in the system can hear the ACKs. This relatively strong assumption is necessary and common for packet communication and establishes upper bounds to performance for both the cooperative and non-cooperative cases.

2.2.1 Cooperation Strategy

The relaying by nodes with a better channel to the destination is expected to yield performance gains. This intuition leads to the following cooperation strategy idea: When node k transmits a packet,

- If the destination $(N+1)$ successfully decodes the packet, it broadcasts an ACK to the network, so the packet exits the system.

- Otherwise, if the destination doesn't decode the packet, but some nodes from the set $\{k+1, k+2, \ldots, N\}$ decode the packet, the node with the largest index among them (that is the node with the best channel to the destination among those that decode the packet) will keep the packet and take the responsibility to forward the packet, while all other nodes will drop that packet in this case. This can be done if any of k's subsequent node that decodes the packet generates an ACK, then by checking all the ACKs, if a node finds itself to be the one with the largest index, it stores the packet, otherwise it drops the packet.

- Finally, if neither the destination nor any of k's subsequent source nodes decodes the packet, the packet remains at k's queue for retransmission.

With this form of cooperation, except for node 1 which only transmits its own packets, every other source node k, for $2 \leq k \leq N$, may forward some packets from any of its predecessor nodes besides its own packets. When node k transmits a packet, generally k is allowed to randomly select a packet in its queue, either its own

packet or a packet relayed from any of its predecessor nodes. We acknowledge that some performance indices, for example, the delay experienced by each individual packet class, depends on the specific transmission priority policy; we will discuss this issue later when the performance is evaluated.

Before we proceed to the next section, we would like to make a few comments regarding this cooperation strategy:

- This type of cooperation is opportunistic in the sense that it always tries to push the packets to the "best" possible source node. The concept of relaying at the network layer is similar to routing. Some work [23, 24] has focused on multi-hop wireless network routing, using the term "dynamic routing" or "opportunistic routing". However, the objective and the approach in those work are quite different from ours. It focused on finding better routing protocols; the intermediate nodes served only for routing the packets toward the destination. To demonstrate the advantage of the proposed routing protocols, simulation results were provided solely without any analytical solutions. By contrast, our work investigates the impact of cooperation in which source nodes relay packets for each other in case the direct links fail; every node (except for the destination) has its own data to deliver to the destination. A mathematical model is developed for this cooperative multi-access system; by assuming bursty packet arrivals, we are interested in the performance measures of stable throughput and delay. The stable throughput region is defined as the union of all arrival rate vectors (in packets/slot) to the sources such

that all queues in the network remain ergodic. The intermediate nodes have their own data, and it is *not* obvious at first glance that such cooperation can benefit all source nodes including the intermediate nodes, as they take extra workload for relaying packets from predecessor nodes.

- The description of the cooperation strategy appears simple; but we should emphasize that instead of being a fixed relaying strategy, this strategy is dynamic. In each time slot, a packet is selected to be relayed by the "best" possible node, based on the current channel outcome. The total number of all possible routes experienced by a packet originated from source i to the destination $(N+1)$ can be calculated to be 2^{N-i}. We illustrate this fact in a simple network with $N = 3$ source nodes in Fig 2.2. A packet transmitted from node 1 can either be directly delivered to the destination node 4 (Fig. 2.2(a)), or relayed through different sets of intermediate nodes via a varying number of hops (Fig. 2.2(b)(c)(d)).

- This cooperation strategy will require all N source nodes to exchange ACK information, but, considering that only one bit is needed for each ACK to indicate successful reception, the cooperation overhead becomes negligible and we can easily incorporate it into our analysis without affecting the results in any substantial way. Therefore, we provide the results without addressing the cooperation overhead.

- The cooperation strategy has the effect of concentrating packets into fewer "virtual" queues: at each source node k, the packets in node k's queue comprise

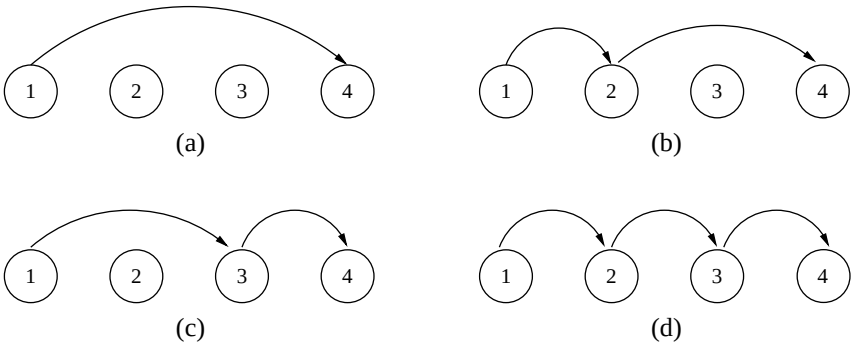

Figure 2.2: An example of the relaying process in a network with $N = 3$ sources and a single destination node indexed by 4. A packet from node 1 reaches destination 4 through (a) direct transmission, (b)(c)(d) relaying by intermediate nodes.

in general a total number of k classes of packets. In [12], packets tended to be pushed to a single relay; here, with a more general relay tactic, packets are gradually concentrated into several intermediate nodes hop-by-hop toward the destination. Given the option to multiplex the traffic streams originated from different sources, the performance limit that can be approached must be an upper bound to that achieved in the non-cooperative case, where packets of different sources are kept in disparate source queues.

2.2.2 Queue Stability

Denote by X_k^t the number of arrivals to the queue at node k during the t-th time slot, and Y_k^t the number of departures from node k. The queue length evolves according to the following form

$$Q_k^{t+1} = [Q_k^t - Y_k^t]^+ + X_k^t \tag{2.1}$$

As defined in [25], the queue is stable if

$$\lim_{t\to\infty} \mathbf{P}\left[Q_k^t < x\right] = F(x) \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1 \tag{2.2}$$

The system is stable if and only if all queues are stable. Under our model, the queue length vector $\mathbf{Q}^t = \{Q_1^t, Q_2^t, \ldots, Q_N^t\}$ forms an irreducible, aperiodic Markov chain, and the queue stability is equivalent to the positive recurrence of the Markov chain. An important tool to determine stability is Loynes' Theorem [16], which states that if the arrival and service processes in a queueing system are jointly stationary, the queue is stable if and only if the average arrival rate is strictly less than the average service rate[3]. We also use the criterion in [26], which establishes that in a work-conserving system, the system is stable if the proportion of time the system is busy is less than 1. With these tools we characterize the stable throughput region as the union of all arrival rate vectors $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ such that all queues in the system remain stable.

2.3 Cooperative Work-Conserving Policy

At the MAC layer we separately consider conflict-free, work-conserving policy on one hand, and a TDMA policy on the other. This section will address the work-conserving policy, while the TDMA policy is described and analyzed in Section 2.4. A policy is *defined* to be conflict-free and work-conserving if two conditions are

[3]Stability at the boundary with equivalence between the arrival rate and service rate is hard to determine, and is out of the scope of this dissertation.

satisfied simultaneously: (i) the system is not idling whenever there are packets in the system; (ii) at most one backlogged source node transmits at each time slot, so there is no contention (i.e., productive work is accomplished when there is demand in the system). Nodes can access the channel via any one of almost countless disciplines as long as they do not violate any of these two conditions. We will show that all conflict-free work-conserving policies have the same stable throughput region, regardless of the transmission priority policy for deciding which class of packets to serve at the transmitter. By incorporating the cooperation strategy stated in Section 2.2.1, the cooperative work-conserving policy is described as follows:

- When a backlogged user node k accesses the channel, it randomly picks up a packet from its queue and transmits. The packet can be from its own arrivals, or a packet that node k relays from any of its predecessor nodes.

- If the destination decodes the packet successfully, it sends back an ACK, and the packet is removed from the system.

- If the destination fails to decode the packet, but there is at least one node in the set $\{k+1, k+2, \ldots, N\}$ that decodes the packet, the node with the largest index among the decoding set is the one that queues the packet; by checking the ACKs, all other nodes that have a copy of the packet can safely drop the packet.

- If none of the nodes from $(k+1)$ up to $(N+1)$ decode the packet, the packet remains in the queue of node k for retransmission.

2.3.1 Stable Throughput Region

As the total arrival process to each source node depends on the departure processes of all its predecessor nodes, the interaction among the N queues makes the stability analysis extremely difficult. In this part, we present the main results with respect to the stable throughput region.

Theorem 2.1 *$\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ is the external Bernoulli arrival rate vector. The stable throughput region under any cooperative work-conserving policy is the same, and is independent of the transmission priority policy adopted by the source nodes in scheduling among different classes of packets. The stable throughput region is characterized by*

$$\Re = \left\{ \Lambda : \sum_{k=1}^{N} \frac{r_k}{1 - \prod_{i=k+1}^{N+1} (1 - p_{k,i})} < 1 \right\} \tag{2.3}$$

where

$$\begin{aligned} r_1 &= \lambda_1 \\ r_k &= \lambda_k + \sum_{i=1}^{k-1} \frac{\left(p_{i,k} \prod_{m=k+1}^{N+1} (1 - p_{i,m}) \right) r_i}{1 - \prod_{j=i+1}^{N+1} (1 - p_{i,j})} \qquad k \in [2, N] \end{aligned} \tag{2.4}$$

Proof: See Appendix 2.7.1. ■

Theorem 2.2 *Nodes in the network are ordered such that $p_{N,N+1} > p_{N-1,N+1} > \ldots > p_{1,N+1}$. With such ordering, the stable throughput region of the cooperative system strictly contains the stable throughput region when cooperation is not performed.*

The maximum stabilizable arrival rates for source nodes from 1 up to $(N-1)$ *strictly increase, while the maximum stabilizable arrival rate for node* N *stays unaffected.*

Proof: See Appendix 2.7.2. ■

The stabilizable point $(0, 0, \ldots, \lambda_N^{\max})$ at the boundary is unchanged after cooperation is performed; this does not mean that cooperation is not beneficial at this point. Indeed, this point corresponds to the case where there are no arrivals to all the other sources, but only source node N is delivering data to the destination without assistance from any other node (since source N has the best source-destination channel). The multi-access system in this case reduces to a single channel, and the maximum stabilizable rate for source N is always $p_{N,N+1}$. If we draw a line across the region, we can easily see that all source users simultaneously increase their stable throughput rates.

2.3.2 Delay Minimizing Policy

In the case of no cooperation, the problem of optimal server allocation for scheduling among the source-destination links has been extensively studied, including in [27, 28, 29]. The stability analysis in Section 2.3.1 establishes that all cooperative conflict-free work-conserving policies have the same stable throughput region; however, the delay performance depends on the specific work-conserving policy. The class of conflict-free, work-conserving policy includes also dynamic policies that base their decisions on the past observations and past actions of the system. We are interested in characterizing the optimal work-conserving policy that minimizes the

average delay over all packets in the system. In this case, the transmission priority policy at each node in deciding what class of packet to serve is irrelevant, and the optimal policy needs only to schedule what source node to be activated at each time slot. The optimal work-conserving policy is given in the following theorem:

Theorem 2.3 *Among all cooperative, conflict-free, work-conserving policies, the policy that minimizes the average delay over all packets in the system is the policy that at each time slot, activates the backlogged source node with the best channel to the destination.*

Proof: Denote by G the set of all cooperative work-conserving policies, and π_0 the policy which activates the backlogged source node with the best channel to the destination at each time slot. Before we give out the proof, a few preliminaries are needed.

Let $A_i(t)$ denote the Bernoulli arrivals to node i at time slot t. For any work-conserving policy, at each time slot, if there are packets in the system, the system never idles, and one and only one backlogged node is activated. This is represented by the binary random variable $M_i(t)$, which is equal to one if the backlogged node i is activated to transmit at slot t; and equal to zero otherwise. If all N source nodes are empty, $M_i(t)$ is equal to zero for all i ($i \in \{1, 2, \ldots, N\}$). When node i is activated, the transmitted packet will either exit the system with a certain probability, or will remain at i's queue or will be relayed to a subsequent source node j for $j > i$. In the latter case, the total number of packets in the system stays unaffected. Denote by $U_i(t)$ the service process at slot t, which is equal to one if and only if the transmitted

packet from node i is decoded by the destination, and hence, exits the system, and is equal to zero otherwise. At each time slot, there can be at most one departure.

We denote by $X(t)$ the total number of packets in the system by the end of slot t. Under our assumptions, the total number of packets in the system evolves according to

$$X(t+1) = \left[X(t) - \sum_{i=1}^{N} U_i(t) M_i(t) \right]^+ + \sum_{i=1}^{N} A_i(t) \tag{2.5}$$

And, as stated above, at each slot t, there can be at most one value of i, for which $U_i(t)M_i(t)$ is equal to 1.

Let $X_0 = \{X(t)\}_{t=1}^{\infty}$ be the process of the total number of packets in the system under policy π_0, and X the corresponding process when some policy $\pi \in G$ acts on the system. To prove the optimality of π_0, first we prove that when the initial state is the same, X_0 is stochastically smaller than X, that is,

$$X_0 \leq_{st} X \tag{2.6}$$

We first show that at $t = 0$, inequality (2.6) is satisfied; then we show that if inequality (2.6) is satisfied at some time t, then it is satisfied at $(t+1)$ as well. To prove this, recall the following property of stochastic ordering:

Property 2.1 *If $u : \Re^n \to \Re$ is an increasing function and A_i and B_i are independent sets of random variables with $A_i \leq_{st} B_i$ for each i, then $u(A_1, \ldots, A_n) \leq_{st} u(B_1, \ldots, B_n)$.*

The initial state is the same under either policy; so we have

$$X_0(0) \leq_{st} X(0) \tag{2.7}$$

If at time slot t, we have

$$X_0(t) \leq_{st} X(t) \tag{2.8}$$

we want to show

$$X_0(t+1) \leq_{st} X(t+1) \tag{2.9}$$

When node i transmits a packet, the packet exits the system if and only if the destination decodes the packet; this occurs with probability $p_{i,N+1}$. Else the packet stays in the system (either in the queue of node i, or in the queue of one of i's subsequent nodes), which occurs with probability $1 - p_{i,N+1}$. So $U_i(t)$ is a Bernoulli random variable which takes the value 1 with probability $p_{i,N+1}$, and 0 with probability $1 - p_{i,N+1}$. We have $p_{1,N+1} < p_{2,N+1} < \ldots < p_{N,N+1}$, so it is true that for every t,

$$U_1(t) <_{st} U_2(t) <_{st} \ldots <_{st} U_{N-1}(t) <_{st} U_N(t) \tag{2.10}$$

and $U_i(t)$ are independent from node to node, and i.i.d over all slots for each i.

Policy π_0 activates the source node i with the largest $p_{i,N+1}$ among all backlogged source nodes at each time slot, so we have

$$\begin{aligned} & \textstyle\sum_{i=1}^{N} U_i(t)M_i(t)\Big|_{\pi_0} \geq_{st} \sum_{i=1}^{N} U_i(t)M_i(t)\Big|_{\pi} \\ \iff \quad & -\textstyle\sum_{i=1}^{N} U_i(t)M_i(t)\Big|_{\pi_0} \leq_{st} -\sum_{i=1}^{N} U_i(t)M_i(t)\Big|_{\pi} \end{aligned} \tag{2.11}$$

Since $X_0(t) \leq_{st} X(t)$, according to the queue length evolution form in Eq. (2.5), and by applying Property 2.1, we show that inequality (2.9) holds.

Therefore, we arrive at the conclusion that the process of the total number of packets in the system under policy π_0 is stochastically smaller than the corresponding process of any other policy $\pi \in G$. This implies that the average number of packets in the system under π_0 is smaller than that of any $\pi \in G$. Finally, by Little's law, we conclude that the average delay over all packets under policy π_0 is the minimum among all cooperative conflict-free work-conserving policies. ■

2.3.3 Average Delay Characterization

Given that the delay minimizing policy has been found, we would also like to characterize the average delay, which is a function of the number of packets. It is known that delay analysis of more than two interacting queues is next to impossible [17, 25]; thus we focus our attention on the case of two users. The work in [30] analyzed delay of interfering queues in packet-radio networks under the collision channel model; the technique relies on solving the moment-generating function of the joint stationary queue lengths. We notice that the delay performance with re-

spect to each individual packet class depends on not only the MAC policy, which schedules the activation of the source users, but also on the transmission priority policy for deciding what class of packet to serve at the transmitter. Therefore, in this part, we investigate the delay performance of a special work-conserving policy that is priority-based. With the objective to somehow balance the network resources in a fair way between users, let us assume that the priority rule assigns higher priority to packets of source user 1, which has a relatively weak user-destination channel. First we describe this cooperative priority-based policy in detail:

- User 1 has higher priority than user 2 to access the channel. When user 1 is backlogged, it transmits a packet. If the destination decodes the packet, the packet exits the system; otherwise, if user 2 decodes the packet, user 2 queues the packet for retransmission, and user 1 drops the packet. If neither user 2 nor the destination decodes the packet, the packet remains at user 1.

- When user 1 is empty, user 2 accesses the channel. If there are both class-1 and class-2 packets in its queue, user 2 picks a packet belonging to class 1 and transmits that packet first; if there are only class-2 packets in its queue, user 2 transmits a packet of class 2. If the transmitted packet is successfully decoded by the destination, the packet exits the system; otherwise, the packet stays at user 2's queue.

We assume that on a control sub-channel, nodes share their queue status information in order for this policy to be implementable. According to the stated cooperative priority-based policy, there are three queues involved in the analysis, as

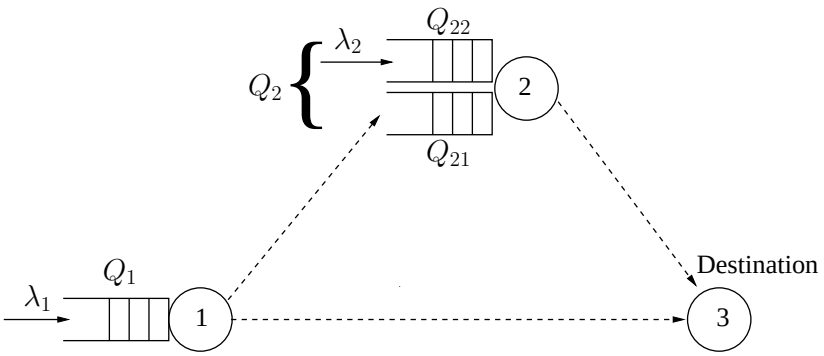

Figure 2.3: Two-user case: under the cooperative priority-based policy, there are three queues involved in the analysis.

illustrated in Fig. 2.3:

1. Q_1: The packets that arrive to user 1 according to the external Bernoulli process with rate λ_1.

2. Q_{21}: The packets at user 2 that are relayed from user 1.

3. Q_{22}: The packets that arrive to user 2 according to the external Bernoulli process with rate λ_2.

Besides, we can merge Q_{21} and Q_{22} at user 2, and denote the merged queue by Q_2. We have the following proposition:

Proposition 2.1 *Under the cooperative priority-based policy, if the system is stable, D_1 and D_2, which represent the average delay of class-1 packets and class-2 packets respectively, are given by*

$$D_1 = \frac{N_1 + N_{21}}{\lambda_1} \tag{2.12}$$

$$D_2 = \frac{N_2 - N_{21}}{\lambda_2} \tag{2.13}$$

where,

1. N_1, *the average queue length of* Q_1, *is given by*

$$N_1 = \frac{-\lambda_1^2 + \lambda_1}{p_{1,2} + p_{1,3} - p_{1,2}p_{1,3} - \lambda_1} \tag{2.14}$$

2. N_{21}, *the average queue length of* Q_{21}, *is given by*

$$N_{21} = \frac{f\lambda_1^2 + g\lambda_1}{a\lambda_1^2 + b\lambda_1 + c} \tag{2.15}$$

with

$$\begin{aligned}
f &= (p_{1,2} - p_{1,2}p_{1,3})\Big(\frac{p_{2,3} - p_{1,3}}{p_{1,2} + p_{1,3} - p_{1,2}p_{1,3}} - p_{1,2} - p_{2,3} + p_{1,2}p_{1,3}\Big) \\
g &= (p_{1,2} - p_{1,2}p_{1,3})\,(p_{1,2} + p_{1,3} - p_{1,2}p_{1,3}) \\
a &= p_{1,2} + p_{2,3} - p_{1,2}p_{1,3} \\
b &= (p_{1,2} + p_{1,3} - p_{1,2}p_{1,3})\,(-p_{1,2} - 2p_{2,3} + p_{1,2}p_{1,3}) \\
c &= p_{2,3}\,(p_{1,2} + p_{1,3} - p_{1,2}p_{1,3})^2
\end{aligned} \tag{2.16}$$

3. N_2, *the average queue length of* Q_2, *is given by*

$$N_2 = \frac{AD - BC}{D - B} \tag{2.17}$$

with

$$
\begin{aligned}
A &= \frac{-\lambda_1^2 - \lambda_2^2 + \lambda_1 + \lambda_2 - \lambda_1\lambda_2}{p_{1,3} - \lambda_1 - \lambda_2} + \frac{\lambda_1^2 - \lambda_1}{p_{1,2} + p_{1,3} - p_{1,2}p_{1,3} - \lambda_1} \\
B &= \frac{p_{1,3} - p_{2,3}}{p_{1,3} - \lambda_1 - \lambda_2} \\
C &= \frac{-\lambda_2^2 + \lambda_2 + \frac{p_{1,2} - p_{1,2}p_{1,3}}{p_{1,2} + p_{1,3} - p_{1,2}p_{1,3}}\lambda_1}{-\lambda_2 - p_{1,2} + p_{1,2}p_{1,3}} \\
D &= \frac{-p_{1,2} - p_{2,3} + p_{1,2}p_{1,3}}{-\lambda_2 - p_{1,2} + p_{1,2}p_{1,3}}
\end{aligned}
\tag{2.18}
$$

Proof: See Appendix 2.7.3. ■

2.4 Cooperative TDMA policy

In this section, we consider the TDMA policy at the MAC layer. User nodes access the channel via a TDMA-based schedule; each source node is allocated a fraction of time in each frame to access the channel. Let $\Omega = (\omega_1, \omega_2, \ldots, \omega_N)$ denote the time allocation vector for the N nodes; all feasible time allocation vectors should satisfy $\Sigma_{i=1}^{N}\omega_i \leq 1$. Together with the proposed cooperation strategy, the cooperative TDMA policy is described as follows:

- At the beginning of node k's assigned time slot, k transmits a packet from its queue if it is backlogged; if k is empty, the slot is not utilized. Just as before, node k randomly picks a packet to transmit, either a packet from its own arrivals, or a packet it relays from any predecessor node.

- If the destination decodes the packet successfully, the packet is removed from the system.

- If the destination fails to decode the packet, then, if there are some nodes among k's subsequent source nodes that decode the packet, only the node with the largest index among them queues the packet for retransmission, while all others drop that packet.

- If none of the nodes from $(k+1)$ up to $(N+1)$ decode the packet, the packet remains in the queue of node k for retransmission.

2.4.1 Stable Throughput Region

Under the TDMA policy, for each allocation vector $(\omega_1, \omega_2, \ldots, \omega_N)$, we can find a corresponding region containing all arrival rates that are stabilizable; then, we will take the union of such regions over all possible allocation vectors that satisfy $\Sigma_{i=1}^{N}\omega_i \leq 1$. Therefore, for any point $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ inside the thus obtained stable throughput region, there exist time allocations $(\omega_1, \omega_2, \ldots, \omega_N)$ such that the queues in the system remain stable.

Theorem 2.4 *The stable throughput region under the cooperative TDMA policy is independent of the transmission policy adopted by the source nodes in scheduling among different classes of packets, and is equal to the stable throughput region of the cooperative work-conserving policy, with the region given by Eq. (2.3) of Theorem 2.1.*

Proof: See Appendix 2.7.4. ■

Under the non-cooperative TDMA policy, by using the same technique, the non-cooperative stable throughput region can be easily obtained, and is the same as

the stable throughput region of the non-cooperative work-conserving policy. Such region is given by Eq. (2.34) in Appendix 2.7.2. Therefore, as stated by Theorem 2.2, the cooperative TDMA policy provides a larger stable throughput region than the non-cooperative TDMA policy.

At light traffic, the TDMA policy idles the time slot even when there are packets in the system. As the arrival rates increase until they reach the boundary of the stable throughput region, some or all of the queues grow without bound. Hence, at the boundary, the TDMA policy becomes conflict-free and work-conserving. This explains the phenomenon that the cooperative work-conserving policy and the cooperative TDMA policy have the same stable throughput region.

2.4.2 Average Delay Characterization

As explained in Section 2.3.3, the delay analysis will be conducted for the case of two users only. Under the cooperative TDMA policy, when user 2 accesses the channel, it randomly selects a packet from its queue to transmit, either its own packet, or a packet relayed from user 1. Therefore, at user 2, the relayed packets and source packets are merged into a single queue to be served. In this case, there are only two queues involved in the analysis: the queue at user node 1, and the queue at user node 2, denoted by Q_1 and Q_2 respectively.

When the system is stable, we derive the analytical delay expressions as functions of the reception probabilities, as well as of the allocation vector (ω_1, ω_2). Among all allocation vectors that stabilize the system, we solve for the optimal

allocation (ω_1^*, ω_2^*) that minimizes the average delay over all packets in the system. The results lead to:

Proposition 2.2 *Under the cooperative TDMA policy, if the system is stable, the minimum average delay over all packets that can be achieved by the optimal* (ω_1^*, ω_2^*) *is*

$$D^*_{Avg}(TDMA) = \frac{\left(\sqrt{b_1\left(-r_2^2 + r_2 + \alpha\lambda_1\lambda_2\right)} + \sqrt{b_2(-r_1^2 + r_1)}\right)^2}{(\lambda_1 + \lambda_2)(b_1b_2 - b_1r_2 - b_2r_1)} \tag{2.19}$$

With (ω_1^*, ω_2^*)*, the average delay of class-1 packets is*

$$\begin{aligned} D_1(TDMA)|_{(\omega_1^*,\omega_2^*)} = & \frac{(-r_1+1)\left(b_2 + \sqrt{\frac{b_1b_2(-r_2^2+r_2+\alpha\lambda_1\lambda_2)}{-r_1^2+r_1}}\right)}{b_1b_2 - b_1r_2 - b_2r_1} \\ & + \frac{\alpha\left(-r_2^2 + r_2 + \alpha\lambda_1\lambda_2\right)\left(b_1 + \sqrt{\frac{b_1b_2(-r_1^2+r_1)}{-r_2^2+r_2+\alpha\lambda_1\lambda_2}}\right)}{r_2\left(b_1b_2 - b_1r_2 - b_2r_1\right)} \end{aligned} \tag{2.20}$$

the average delay of class-2 packets is

$$D_2(TDMA)|_{(\omega_1^*,\omega_2^*)} = \frac{\left(-r_2^2 + r_2 + \alpha\lambda_1\lambda_2\right)\left(b_1 + \sqrt{\frac{b_1b_2(-r_1^2+r_1)}{-r_2^2+r_2+\alpha\lambda_1\lambda_2}}\right)}{r_2\left(b_1b_2 - b_1r_2 - b_2r_1\right)} \tag{2.21}$$

with

$$
\begin{aligned}
r_1 &= \lambda_1 \\
r_2 &= \frac{p_{1,2}(1-p_{1,3})}{p_{1,2}+p_{1,3}-p_{1,2}p_{1,3}}\lambda_1 + \lambda_2 \\
b_1 &= 1-(1-p_{1,2})(1-p_{1,3}) \\
b_2 &= p_{2,3} \\
\alpha &= \frac{p_{1,2}(1-p_{1,3})}{p_{1,2}+p_{1,3}-p_{1,2}p_{1,3}}
\end{aligned}
\tag{2.22}
$$

Proof: See Appendix 2.7.5. ■

2.5 Numerical Results

In Fig. 2.4, we compare the stable throughput regions of the cooperative policies and the non-cooperative policies in the two-user case. The stable throughput region of the non-cooperative policies is characterized by Eq. (2.34), with $N = 2$, in Appendix 2.7.2. In this plot, the reception probabilities of the user-destination channels are chosen to be $p_{1,3} = 0.3$, $p_{2,3} = 0.8$; while for the inter-user channel we study three channel conditions with reception probabilities: $p_{1,2} = 0.4, 0.6, 0.85$ (note that $p_{1,2}$ affects the performance of the cooperative case only). The stable throughput region of the cooperative policies is found to strictly contain that of the non-cooperative policies, and the region increases as the inter-user channel condition improves.

In Fig. 2.5 and Fig. 2.6, we demonstrate the benefit of cooperation in the delay

performance for class-1 packets and class-2 packets respectively, under the priority-based policy. The channel reception probabilities are chosen as those for plotting Fig. 2.4. Using the same technique, we obtain the delay performance under the non-cooperative priority-based policy as:

(a) The average delay of class-1 packets is

$$D_1|\text{non-coop} = \frac{-\lambda_1 + 1}{p_{1,3} - \lambda_1} \tag{2.23}$$

(b) The average delay of class-2 packets is

$$D_2|\text{non-coop} = \frac{\left(-p_{1,3} + p_{2,3} - p_{1,3}p_{2,3}\right)\lambda_1 - p_{1,3}^2\lambda_2 + p_{1,3}\lambda_1\lambda_2 + p_{1,3}^2}{\left(p_{2,3}\lambda_1 + p_{1,3}\lambda_2 - p_{1,3}p_{2,3}\right)\left(\lambda_1 - p_{1,3}\right)} \tag{2.24}$$

We let $\lambda_1 = \lambda_2 = \lambda$ and vary λ to obtain the shown plots. It is seen that, when cooperation is permitted, both class-1 and class-2 packets experience significantly reduced average delay. And when the inter-user channel quality improves, cooperation leads to higher performance gains for both users.

In Fig. 2.7-2.9, we illustrate the benefit of cooperation in terms of the delay performance under the TDMA policy. Using the same technique, the minimum overall average delay achievable under the non-cooperative TDMA policy is obtained as

$$D^*_{\text{Avg}}(\text{TDMA})\big|_{\text{non-coop}} = \frac{\left(\sqrt{p_{1,3}\lambda_2(1-\lambda_2)} + \sqrt{p_{2,3}\lambda_1(1-\lambda_1)}\right)^2}{\left(\lambda_1 + \lambda_2\right)\left(p_{1,3}p_{2,3} - p_{1,3}\lambda_2 - p_{2,3}\lambda_1\right)} \tag{2.25}$$

With the optimal allocation vector that achieves the minimum delay in Eq. (2.25),

the average delay of class-1 packets and class-2 packets in the non-cooperative case is given by

$$D_1(\text{TDMA})|_{\text{non-coop}} = \frac{(1-\lambda_1)\left(p_{2,3} + p_{1,3}\sqrt{\frac{p_{2,3}\lambda_2(1-\lambda_2)}{p_{1,3}\lambda_1(1-\lambda_1)}}\right)}{p_{1,3}p_{2,3} - p_{1,3}\lambda_2 - p_{2,3}\lambda_1} \tag{2.26}$$

$$D_2(\text{TDMA})|_{\text{non-coop}} = \frac{(1-\lambda_2)\left(p_{1,3} + p_{2,3}\sqrt{\frac{p_{1,3}\lambda_1(1-\lambda_1)}{p_{2,3}\lambda_2(1-\lambda_2)}}\right)}{p_{1,3}p_{2,3} - p_{1,3}\lambda_2 - p_{2,3}\lambda_1} \tag{2.27}$$

Fig. 2.7 shows that the minimum overall average delay is substantially decreased after cooperation is performed under the TDMA policy. It is also seen that the delay performance for both the class-1 packets and class-2 packets is greatly improved over that of the non-cooperative case, as illustrated by Fig. 2.8 and Fig. 2.9 respectively. Moreover, a better inter-user channel reduces delay when cooperation is permitted.

2.6 Discussion

In this chapter, we studied a new concept of cooperation at the network level. In a wireless multi-access system consisting of a number of sources and a single destination, the cooperation among the source users was defined so that a transmitted packet, if not decoded by the destination, was relayed by the source user with the best possible source-destination channel among those that decoded it. With an erasure channel model which captures the effects of fading and attenuation, as well as the multicast advantage of wireless channels, the impact of our cooperation strategy was evaluated under two classes of cooperative MAC policies separately.

In Section 2.3, we investigated the class of cooperative work-conserving policy. First, the stable throughput region was explicitly characterized, and was proved to be the same for all cooperative, work-conserving policies. This region strictly outer-bounds the corresponding region of the non-cooperative policy. Then, among the class of all cooperative work-conserving policies, which also includes dynamic policies designed according to the whole system history, we explicitly characterized the optimal policy that minimizes the average delay. And for the two-user cooperative priority-based policy, we derived the analytical delay expressions for each packet class. It was shown that both users' packets experience reduced average delay after cooperation is performed.

In Section 2.4, we investigated the cooperative TDMA policy. Similar observations were obtained as for the cooperative work-conserving policy. By taking the union of all time allocation vectors, the achievable stable throughput region was explicitly characterized, and was shown to be the same as the stable throughput region we obtained for the cooperative work-conserving policy, which strictly outer-bounds the non-cooperative stable throughput region. The average delay performance for the two-user case was quantified subsequently; our results showed that cooperation substantially reduces the delay for both users.

Although the motivating reason for cooperation is not discussed, the fact that all users can simultaneously increase stable throughput rates and experience less average delay can serve as a motivation for cooperation. This is somehow "anti-intuitive", since it seems that the intermediate user nodes will suffer performance loss by wasting time slots for relaying others' packets that could be rather used for

their own packets. For example, in the simple case of two users shown in Fig. 2.3, it seems that user 2 "wastes" a fraction of time for transmitting user 1's packets. However, it turns out that user 2 is indeed gaining from the relaying. This is because, by opportunistically relaying user 1's packets through user 2 which has a better user-destination channel, the cooperation strategy results in emptying user 1's queue faster; in return, more network resources can be utilized for delivering user 2's packets. As a result, all users simultaneously achieve performance gains.

So far, we considered only the single-packet reception channel model, and the transmission of different users was scheduled in different time slots. In Chapter 3, we will continue our line of research by implementing cooperation at the network layer, while incorporating several enhancement techniques that are based on the physical layer. The work in this chapter touches on two elements that are ignored by classical information-theoretic studies of cooperation: source burstiness and delay. In Chapter 4, we will consider both cases of bursty source and saturated source; corresponding to these two cases, we will assess two "rate" measures and examine their relations.

2.7 Appendix

2.7.1 Proof of Theorem 2.1

In a conflict-free, work-conserving system, the system is stable if the proportion of time the system is busy is less than 1. Denote by ρ_k the proportion of time source

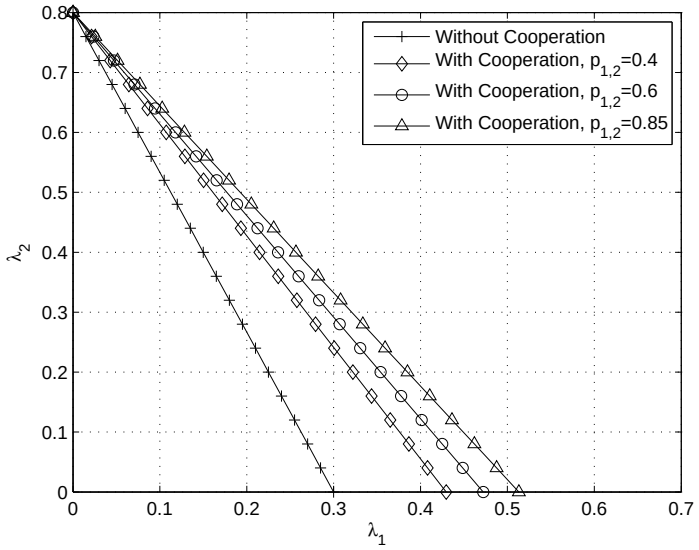

Figure 2.4: Comparison of the stable throughput regions under cooperative and non-cooperative policies.

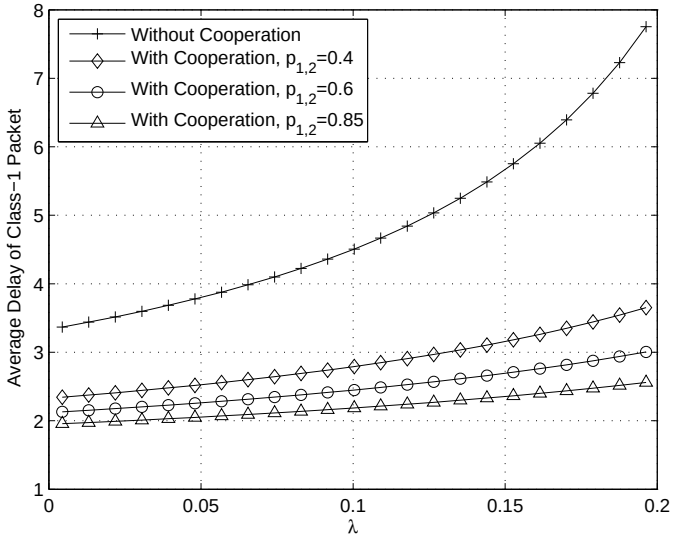

Figure 2.5: Comparison of the average delay for class-1 packets under cooperative and non-cooperative priority-based policies.

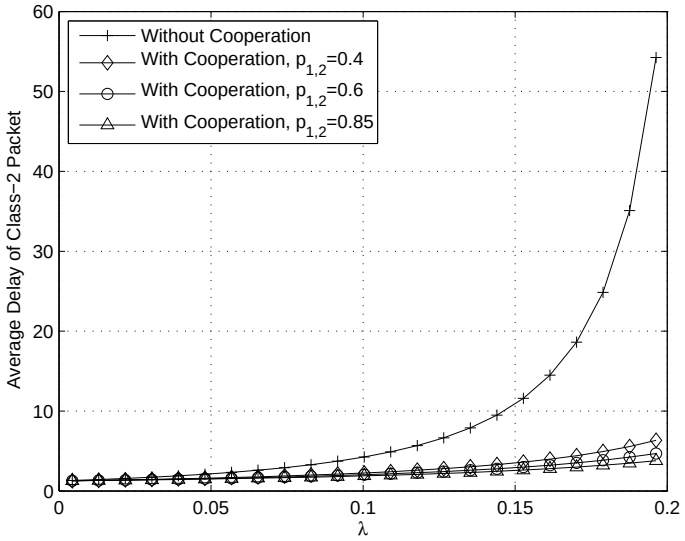

Figure 2.6: Comparison of the average delay for class-2 packets under cooperative and non-cooperative priority-based policies.

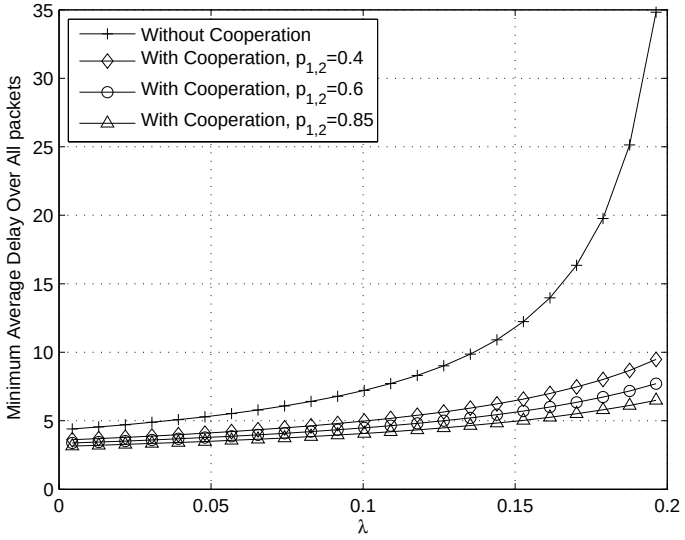

Figure 2.7: Comparison of the minimum average delay over all packets under cooperative and non-cooperative TDMA policies.

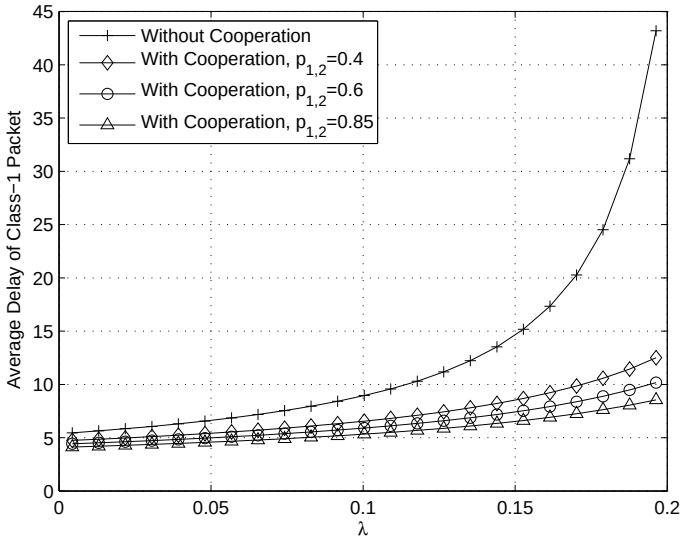

Figure 2.8: Comparison of the average delay for class-1 packets under cooperative and non-cooperative TDMA policies.

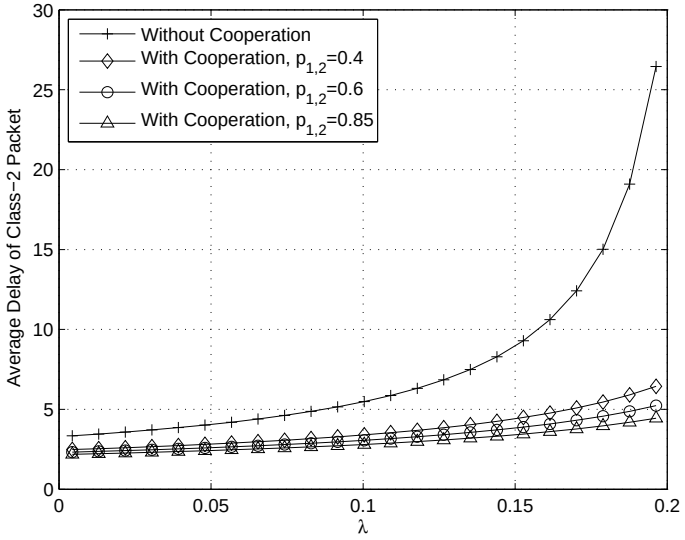

Figure 2.9: Comparison of the average delay for class-2 packets under cooperative and non-cooperative TDMA policies.

node k is transmitting ($k \in \{1, 2, \ldots, N\}$); the system is stable if

$$\sum_{k=1}^{N} \rho_k < 1 \tag{2.28}$$

where ρ_k is the utilization factor of node k, given by $\rho_k = r_k/\mu_k$, where r_k is the average total arrival rate to node k, and μ_k is the average service rate seen by k. Then we need to calculate r_k and μ_k for every node k, $1 \leq k \leq N$.

At node 1's queue, we have $r_1 = \lambda_1$. A packet will depart node 1 if the packet is decoded by at least one node from node 2 up to node $(N+1)$, so the service rate received by node 1 is

$$\mu_1 = 1 - \prod_{j=2}^{N+1} (1 - p_{1,j}) \tag{2.29}$$

Then we analyze node k's queue, for $k \geq 2$. The total arrival processes to node k consist of two parts: first, the external Bernoulli arrival process to node k with rate λ_k; and second, the arrivals from the predecessor node 1 up to node $(k-1)$. It remains to calculate the arrival rate to node k from the predecessor nodes. When node i (for $1 \leq i \leq k-1$) transmits a packet, the packet will be dropped by i if it is decoded by at least one node in the set $\{i+1, i+2, \ldots, N+1\}$, which has probability $1 - \prod_{j=i+1}^{N+1} (1 - p_{i,j})$. When i transmits, the packet will be relayed to node k's queue, if and only if k decodes the packet and none of the nodes from $(k+1)$ to $(N+1)$ decode the packet, which happens with probability $p_{i,k} \prod_{m=k+1}^{N+1} (1 - p_{i,m})$. Hence, we obtain the conditional probability that a transmitted packet by node i is relayed

to the queue at node k given that the transmitted packet departs node i's queue, which is

$$\frac{p_{i,k}\prod_{m=k+1}^{N+1}(1-p_{i,m})}{1-\prod_{j=i+1}^{N+1}(1-p_{i,j})} \tag{2.30}$$

We know that when the queue at node i is stable, the departure rate is the same as the total arrival rate to the queue, which is r_i. Therefore, the total arrival rate to node k from node i is

$$\frac{\left(p_{i,k}\prod_{m=k+1}^{N+1}(1-p_{i,m})\right)r_i}{1-\prod_{j=i+1}^{N+1}(1-p_{i,j})} \tag{2.31}$$

If we take the summation of all arrival rates from node 1 up to $(k-1)$, and add the external Bernoulli arrival rate λ_k, we obtain that the total arrival rate r_k to node k is

$$r_k = \lambda_k + \sum_{i=1}^{k-1}\frac{\left(p_{i,k}\prod_{m=k+1}^{N+1}(1-p_{i,m})\right)r_i}{1-\prod_{j=i+1}^{N+1}(1-p_{i,j})} \tag{2.32}$$

Then we evaluate the average service rate node k receives. A packet will be dropped from k's queue if and only if the packet is decoded by at least one node from the set $\{k+1, k+2, \ldots, N+1\}$, so the average service rate node k receives is

$$\mu_k = 1 - \prod_{m=k+1}^{N+1}(1-p_{k,m}) \tag{2.33}$$

Finally, by substituting r_k and μ_k into the stability condition expressed by

Eq. (2.28), the stable throughput region of the cooperative work-conserving policy is shown to be given by Eq. (2.3) of Theorem 2.1. Besides, the above analysis and derivation is independent of the transmission policy by each node k ($k \in \{1, 2, \ldots, N\}$) in deciding what class of packet to transmit. We should also remark here that the above analysis is the same for any conflict-free work-conserving policy, so the results of Theorem 2.1 hold for any cooperative policy as long as it is conflict-free and work-conserving. ■

2.7.2 Proof of Theorem 2.2

It is easy to show by repeating the same argument as before that, the stability condition for the non-cooperative, work-conserving policy is given by

$$\sum_{k=1}^{N} \frac{\lambda_k}{p_{k,N+1}} < 1 \tag{2.34}$$

while the stability condition for the cooperative work-conserving policy is given in Eq. (2.3) of Theorem 2.1. By checking Eq. (2.3) and Eq. (2.34), we see that they both can be expressed by the following form

$$\sum_{k=1}^{N} \alpha_k \lambda_k < 1 \tag{2.35}$$

where the coefficients α_k are functions of the reception probabilities, and are strictly greater than zero. The region characterized by this form is the region of all points

under a linear N-dimension hyperplane. Therefore to compare the two regions, it is necessary and sufficient to compare the maximum rate $\lambda_k^{\max}$ that can be achieved for every k, since all points under the hyperplane connected by the N maximum points can be achieved through time-sharing.

Under the non-cooperative work-conserving policy, the maximum stabilizable arrival rate for node k ($1 \leq k \leq N$) is

$$\lambda_k^{\max}|_{\text{non-cooperative}} = p_{k,N+1} \tag{2.36}$$

Under the cooperative work-conserving policy, when $k = N$, the maximum stabilizable rate for node N is

$$\lambda_N^{\max}|_{\text{coopeartive}} = p_{N,N+1} = \lambda_N^{\max}|_{\text{non-coopeartive}} \tag{2.37}$$

For $1 \leq k \leq N-1$, after we rearrange the terms in Eq. (2.3), the maximum stabilizable rate for node k can be expressed as

$$\frac{1}{\left(\frac{\beta_{k,k}}{p_{k,N+1}} + \frac{\beta_{k,k+1}}{p_{k+1,N+1}} + \ldots + \frac{\beta_{k,N}}{p_{N,N+1}}\right)} \tag{2.38}$$

where $\beta_{k,i}$ are functions of the reception probabilities, strictly greater than zero, and satisfy $\sum_{i=k}^{N} \beta_{k,i} = 1$. Since $p_{N,N+1} > p_{N-1,N+1} > \ldots > p_{k,N+1}$, the following

inequality holds for every $1 \leq k \leq N-1$,

$$\frac{1}{\left(\frac{\beta_{k,k}}{p_{k,N+1}} + \frac{\beta_{k,k+1}}{p_{k+1,N+1}} + \ldots + \frac{\beta_{k,N}}{p_{N,N+1}}\right)} > p_{k,N+1} \tag{2.39}$$

Hence, for $1 \leq k \leq N-1$, we have

$$\lambda_k^{\max}|_{\text{cooperative}} > \lambda_k^{\max}|_{\text{non-cooperative}} \tag{2.40}$$

Therefore we prove Theorem 2.2. The stable throughput region of the cooperative work-conserving policy strictly contains the stable throughput region of the non-cooperative work-conserving policy. ■

2.7.3 Proof of Proposition 2.1

According to the priority rule, the three queues in Fig. 2.3 are assigned priorities such that Q_1 has the highest priority, Q_{21} has the second highest priority and Q_{22} has the lowest priority. First we analyze class-1 packets. We notice that a class-1 packet may experience two parts of queueing delay: 1) the queueing delay at user 1, that is, the delay in Q_1; and 2) the queueing delay at user 2, that is, the delay in Q_{21}. If the packet is directly delivered to the destination by user 1, the packet experiences the queueing delay in Q_1 only. This happens with probability $1-\alpha = \frac{p_{1,3}}{1-(1-p_{1,2})(1-p_{1,3})}$; otherwise, if the packet is relayed to user 2 first before it finally reaches the destination, the packet experiences the total delay which is the

queueing delay in Q_1 plus the queueing delay in Q_{21}. This happens with probability $\alpha = \frac{p_{1,2}(1-p_{1,3})}{1-(1-p_{1,2})(1-p_{1,3})}$. Therefore, the average queueing delay experienced by class-1 packets should be computed as follows

$$D_1 = (1-\alpha)T_1 + \alpha(T_1 + T_{21}) = T_1 + \alpha T_{21} \tag{2.41}$$

where T_1 is the average queueing delay in Q_1, and T_{21} is the average queueing delay in Q_{21}. The arrival rates to Q_1 and Q_{21} are λ_1 and $\alpha\lambda_1$ respectively; by Little's Law, we have

$$T_1 = N_1/\lambda_1, \qquad\qquad T_{21} = N_{21}/\alpha\lambda_1 \tag{2.42}$$

where N_1 and N_{21} denote the average queue sizes of Q_1 and Q_{21} respectively. By substituting Eq. (2.42) into Eq. (2.41), the average delay of class-1 packets is computed by

$$D_1 = \frac{N_1 + N_{21}}{\lambda_1} \tag{2.43}$$

Now it remains to calculate N_1 and N_{21}. We notice that Q_1 is a discrete-time $M/M/1$ queue[4] with arrival rate λ_1 and service rate $\mu_1 = p_{1,2} + p_{1,3} - p_{1,2}p_{1,3}$; by

[4]We use the notion of discrete-time M/M/1 queue to describe a queueing system with Bernoulli arrival process and geometrically distributed service times.

applying the Pollaczek-Khinchin formula [31], we obtain N_1 as

$$N_1 = \frac{-\lambda_1^2 + \lambda_1}{p_{1,2} + p_{1,3} - p_{1,2}p_{1,3} - \lambda_1} \tag{2.44}$$

To calculate N_{21}, we solve the moment-generating function of the joint queue lengths of Q_1 and Q_{21}, denoted by $G(x,y) = \lim_{t\to\infty} \mathbf{E}\left[x^{Q_1^t} y^{Q_{21}^t}\right]$. The queue evolution has the form shown in Eq. (2.1). Then from the description of the cooperative priority-based policy stated in Section 2.3.3, it can be seen that:

$$\begin{aligned}
\mathbf{E}\left[x^{Q_1^{t+1}} y^{Q_{21}^{t+1}}\right] = (\lambda_1 x + 1 - \lambda_1) \cdot \Bigg\{ & \mathbf{E}\left[\mathbf{1}\left[Q_1^t = 0, Q_{21}^t = 0\right]\right] \\
& + \left(\frac{p_{1,2}(1-p_{1,3})y}{x} + \frac{p_{1,3}}{x} + (1-p_{1,2})(1-p_{1,3})\right) \mathbf{E}\left[x^{Q_1^t}\mathbf{1}\left[Q_1^t > 0, Q_{21}^t = 0\right]\right] \\
& + \left(\frac{p_{2,3}}{y} + 1 - p_{2,3}\right) \mathbf{E}\left[y^{Q_{21}^t}\mathbf{1}\left[Q_1^t = 0, Q_{21}^t > 0\right]\right] \\
& + \left(\frac{p_{1,2}(1-p_{1,3})y}{x} + \frac{p_{1,3}}{x} + (1-p_{1,2})(1-p_{1,3})\right) \mathbf{E}\left[x^{Q_1^t} y^{Q_{21}^t}\mathbf{1}\left[Q_1^t > 0, Q_{21}^t > 0\right]\right] \Bigg\}
\end{aligned} \tag{2.45}$$

Let $t \to \infty$ at both sides of Eq. (2.45), we obtain

$$G(x,y) = (x\lambda_1 + 1 - \lambda_1)\,\frac{b(x,y)G(0,0) + c(x,y)G(0,y)}{yd(x,y)} \tag{2.46}$$

where,

$$
\begin{aligned}
b(x,y) &= xyp_{2,3} - xp_{2,3} \\
c(x,y) &= xp_{2,3} - yp_{1,3} - y^2p_{1,2}(1-p_{1,3}) + xy\left(p_{1,2}+p_{1,3}-p_{1,2}p_{1,3}-p_{2,3}\right) \\
d(x,y) &= x - (x\lambda_1 + 1 - \lambda_1)\Big(p_{1,3} + yp_{1,2}(1-p_{1,3}) + x(1-p_{1,2})(1-p_{1,3})\Big)
\end{aligned}
\tag{2.47}
$$

Using the same approach as in [30], $G(0,0)$ is solved to be

$$
G(0,0) = 1 - \frac{p_{1,2}(1-p_{1,3}) + p_{2,3}}{p_{2,3}\left(1-(1-p_{1,2})(1-p_{1,3})\right)}\lambda_1 \tag{2.48}
$$

and note that

$$
G(0,1) = \lim_{t\to\infty} \mathbf{P}\left[Q_1^t = 0\right] = 1 - \frac{\lambda_1}{1-(1-p_{1,2})(1-p_{1,3})} \tag{2.49}
$$

Define $G_2(x,y) = \frac{dG(x,y)}{dy}$, so $N_{21} = G_2(1,1)$. By taking the derivative of Eq. (2.46) with respect to y, applying L'Hopital's rule twice, and substituting Eq. (2.48) and Eq. (2.49), we obtain one equation relating $G_2(1,1)$ and $G_2(0,1)$ as

$$
G_2(1,1) = -\frac{\lambda_1}{p_{1,2}+p_{1,3}-p_{1,2}p_{1,3}} + \frac{p_{1,2}+p_{2,3}-p_{1,2}p_{1,3}}{p_{1,2}-p_{1,2}p_{1,3}}G_2(0,1) \tag{2.50}
$$

Then we compute $\frac{dG(y,y)}{dy}\Big|_{y=1}$, and note that $\frac{dG(y,y)}{dy}\Big|_{y=1} = G_1(1,1) + G_2(1,1)$; after some algebra, we obtain

$$
G_2(1,1) = \frac{\left(-\lambda_1^2+\lambda_1\right)\left(p_{1,2}-p_{1,2}p_{1,3}\right)}{\left(p_{1,3}-\lambda_1\right)\left(p_{1,2}+p_{1,3}-p_{1,2}p_{1,3}-\lambda_1\right)} + \frac{p_{1,3}-p_{2,3}}{p_{1,3}-\lambda_1}G_2(0,1) \tag{2.51}
$$

With Eq. (2.50) and Eq. (2.51), eliminating $G_2(0,1)$ yields N_{21}, which is as shown in Eq. (2.15) in Proposition 2.1. Now that we have calculated N_1 and N_{21}, from Eq. (2.43), we obtain the average delay of class-1 packets exactly as given by Proposition 2.1.

Next we analyze class-2 packets. The average queueing delay experienced by class-2 packets is equal to the average queueing delay in Q_{22}. We merge Q_{21} and Q_{22} to be Q_2, and denote by N_2 the average queue length of Q_2, so $N_{22} = N_2 - N_{21}$. N_{21}, the average queue length of Q_{21}, has been derived above; to obtain N_{22}, it is equivalent to obtain N_2. Define $H(x,y) = \lim_{t\to\infty} \mathbf{E}\left[x^{Q_1^t} y^{Q_2^t}\right]$ as the moment-generating function of the joint queue lengths of Q_1 and Q_2. Under the cooperative priority-based policy, $H(x,y)$ can be written as follows:

$$H(x,y) = F'(x,y)\frac{b'(x,y)H(0,0) + c'(x,y)H(0,y)}{yd'(x,y)} \tag{2.52}$$

where,

$$\begin{aligned}
F'(x,y) &= (x\lambda_1 + 1 - \lambda_1)\,(y\lambda_2 + 1 - \lambda_2) \\
b'(x,y) &= xyp_{2,3} - xp_{2,3} \\
c'(x,y) &= xp_{2,3} - yp_{1,3} - y^2 p_{1,2}(1 - p_{1,3}) + xy\,(p_{1,2} + p_{1,3} - p_{1,2}p_{1,3} - p_{2,3}) \\
d'(x,y) &= x - (x\lambda_1 + 1 - \lambda_1)(y\lambda_2 + 1 - \lambda_2)\,(p_{1,3} + yp_{1,2}(1 - p_{1,3}) + x(1 - p_{1,2})(1 - p_{1,3}))
\end{aligned} \tag{2.53}$$

Using the same technique as in deriving N_{21}, we derive N_2 as shown in Eq. (2.17)

of Proposition 2.1. Finally applying Little's Law gives the average queueing delay of class-2 packets as given by Proposition 2.1. ■

2.7.4 Proof of Theorem 2.4

Denote by A_k^t the event that there is an arrival to node k at slot t from the external Bernoulli process, W_k^t the event that slot t is assigned to node k, and $\Theta_{i,k}^t$ the event that node k decodes the packet transmitted by node i but none of the nodes from $(k+1)$ to $(N+1)$ decodes the packet.

At node 1, the total arrival rate is λ_1. A packet leaves node 1 if the slot is assigned to node 1, and at least one node from node 2 up to node $(N+1)$ decodes the transmitted packet. So the average service rate seen by node 1 is

$$\mu_1 = \omega_1 \left(1 - \prod_{j=2}^{N+1} (1 - p_{1,j})\right) \tag{2.54}$$

For node k, where $2 \leq k \leq N$, the total number of arrivals X_k^t in slot t is given by

$$X_k^t = \mathbf{1}\left[A_k^t\right] + \sum_{i=1}^{k-1} \mathbf{1}\left[W_i^t \bigcap \{Q_i^t \neq 0\} \bigcap \Theta_{i,k}^t\right] \tag{2.55}$$

where $\mathbf{1}\,[]$ is the indicator function. The first term on the right-hand side of Eq. (2.55) accounts for the external Bernoulli arrival process to node k; the second term accounts for the arrival processes from the predecessor node 1 up to node $(k-1)$. For

$1 \leq i \leq (k-1)$, the packet transmitted by node i will be relayed to node k if the following three events happen together: first, slot t is assigned to node i; second, node i is nonempty so i transmits, which has probability r_i/μ_i; and third, k decodes the packet while none of the nodes from $(k+1)$ to $(N+1)$ decodes the packet. These three events are independent; adding the Bernoulli arrival rate λ_k, the total arrival rate to node k is

$$r_k = \lambda_k + \sum_{i=1}^{k-1} \left(\omega_i \frac{r_i}{\mu_i} p_{i,k} \prod_{m=k+1}^{N+1} (1 - p_{i,m}) \right) \tag{2.56}$$

To evaluate such r_k, we need to calculate μ_i on the right-hand side of Eq. (2.56), which is the average service rate seen by node i. A packet will leave node i's queue if it is node i's time slot, and at least one node from the set $\{i+1, i+2, \ldots, N+1\}$ decodes the packet transmitted by node i. So we have

$$\mu_i = \omega_i \left(1 - \prod_{j=i+1}^{N+1} (1 - p_{i,j}) \right) \tag{2.57}$$

By substituting Eq. (2.57) into Eq. (2.56), we obtain r_k exactly as shown in Eq. (2.4) of Theorem 2.1.

Given a fixed allocation vector $\Omega = (\omega_1, \omega_2, \ldots, \omega_N)$, by using Loynes' Theorem, the queue at node k is stable if $r_k < \mu_k$; and the system is stable if all queues

are stable, that is

$$\begin{aligned} & r_k < \mu_k \\ \Longleftrightarrow & r_k < \omega_k \left(1 - \prod_{j=k+1}^{N+1} (1 - p_{k,j})\right) \end{aligned} \tag{2.58}$$

for all $1 \leq k \leq N$, with r_k given in Eq. (2.4). The stable throughput region of the cooperative TDMA policy is then given by the union of the stabilizable arrival rates $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ over all feasible allocation vectors that satisfy

$$\begin{aligned} & \sum_{k=1}^{N} \omega_k \leq 1 \\ & \omega_k \geq 0, \qquad 1 \leq k \leq N \end{aligned} \tag{2.59}$$

This problem can be easily solved using vector optimization techniques; after simple algebra, the stability condition of the cooperative TDMA policy is given by

$$\sum_{k=1}^{N} \frac{r_k}{1 - \prod_{i=k+1}^{N+1} (1 - p_{k,i})} < 1 \tag{2.60}$$

Compared with Eq. (2.3) of Theorem 2.1, we arrive at the conclusion that the stable throughput region of the cooperative TDMA policy is identical to the stable throughput region of the cooperative work-conserving policy. It is important to note here that when we construct the proof, any source node k ($k \in \{1, 2, \ldots, N\}$) is allowed to randomly pick a packet from its queue to transmit (that is, either a packet from its own arrivals, or a packet relayed from any of its predecessor nodes);

the above analysis is independent of the transmission priority policy adopted. This concludes the proof of Theorem 2.4. ■

2.7.5 Proof of Proposition 2.2

There are two queues to be analyzed, the queue at user node 1 and the queue at user node 2, denoted by Q_1 and Q_2 respectively. For a given allocation vector (ω_1, ω_2) that stabilizes the system, by using the same technique as in the delay analysis of the priority-based policy, we obtain the average queue length of Q_1 and Q_2 as

$$N_1(\text{TDMA}) = \frac{-r_1^2 + r_1}{\omega_1 b_1 - r_1} \tag{2.61}$$

$$N_2(\text{TDMA}) = \frac{-r_2^2 + r_2 + \alpha\lambda_1\lambda_2}{\omega_2 b_2 - r_2} \tag{2.62}$$

with r_1, r_2, b_1, b_2 and α given in Eq. (2.22) of Proposition 2.2. The arrival rates to Q_1 and Q_2 are r_1 and r_2 respectively; by Little's Law, the average queueing delay in these two queues is

$$T_1(\text{TDMA}) = \frac{N_1(\text{TDMA})}{r_1}, \qquad T_2(\text{TDMA}) = \frac{N_2(\text{TDMA})}{r_2} \tag{2.63}$$

Note that at user 2, the relayed packets and source packets are merged into a single queue to be served, so the average queueing delay experienced by both class-1 packets and class-2 packets in Q_2 is the same, and is equal to $T_2(\text{TDMA})$. By

following the same argument as in the proof of Proposition 2.1, the average delay of class-1 packets and the average delay of class-2 packets in the system are calculated as

$$D_1(\text{TDMA}) = T_1(\text{TDMA}) + \alpha T_2(\text{TDMA}) \tag{2.64}$$

$$D_2(\text{TDMA}) = T_2(\text{TDMA}) \tag{2.65}$$

And the average delay over all packets in the system is given by

$$\begin{aligned} D_{\text{Avg}}(\text{TDMA}) =& \frac{\lambda_1 D_1(\text{TDMA}) + \lambda_2 D_2(\text{TDMA})}{\lambda_1 + \lambda_2} \\ =& \frac{N_1(\text{TDMA}) + N_2(\text{TDMA})}{\lambda_1 + \lambda_2} \end{aligned} \tag{2.66}$$

Then we want to minimize the overall average delay given in Eq. (2.66) by optimizing over all possible allocation vectors under the following constraints:

$$\frac{\lambda_1}{1-(1-p_{1,2})(1-p_{1,3})} < \omega_1 \leq 1 \tag{2.67}$$

$$\frac{\frac{p_{1,2}(1-p_{1,3})}{1-(1-p_{1,2})(1-p_{1,3})}\lambda_1 + \lambda_2}{p_{2,3}} < \omega_2 \leq 1 \tag{2.68}$$

$$\omega_1 + \omega_2 \leq 1 \tag{2.69}$$

The objective function (2.66) is convex in both ω_1 and ω_2. The inequality constraints (2.67) and (2.68) are the conditions for user 1's queue and user 2's queue to be stable, and constraint (2.69) is the condition for a feasible allocation vector. This problem can be solved easily with the Lagrangian multiplier technique, and the

optimal allocation vector (ω_1^*, ω_2^*) that minimizes the overall average delay is found to be

$$\begin{aligned}
\omega_1^* &= \frac{r_1 + (b_2 - r_2)\sqrt{\frac{b_1\left(-r_1^2+r_1\right)}{b_2\left(-r_2^2+r_2+\alpha\lambda_1\lambda_2\right)}}}{b_1 + b_2\sqrt{\frac{b_1\left(-r_1^2+r_1\right)}{b_2\left(-r_2^2+r_2+\alpha\lambda_1\lambda_2\right)}}} \\
\omega_2^* &= 1 - \omega_1^* = \frac{b_1 - r_1 + r_2\sqrt{\frac{b_1\left(-r_1^2+r_1\right)}{b_2\left(-r_2^2+r_2+\alpha\lambda_1\lambda_2\right)}}}{b_1 + b_2\sqrt{\frac{b_1\left(-r_1^2+r_1\right)}{b_2\left(-r_2^2+r_2+\alpha\lambda_1\lambda_2\right)}}}
\end{aligned} \tag{2.70}$$

with r_1, r_2, b_1, b_2 and α given in Eq. (2.22) of Proposition 2.2. By substituting (ω_1^*, ω_2^*) into Eq. (2.66), the minimum average delay over all packets is exactly given by Eq. (2.19) of Proposition 2.2. And by substituting (ω_1^*, ω_2^*) into Eq. (2.64) and Eq. (2.65), we obtain the average delay of class-1 packets and class-2 packets exactly as shown in Eq. (2.20) and Eq. (2.21) of Proposition 2.2 respectively. ■

Chapter 3

Enhanced Cooperation Based on Physical-Layer Techniques

3.1 Introduction

In Chapter 2, we introduced a new cooperation concept at the network layer, and evaluated its impact in a general wireless multi-access system. In that work, we considered only the single-packet reception channel model and allowed for at most one transmission in each time slot. By doing so, too limited and oversimplified assumptions were made on the physical-layer properties. In this chapter, we aim to integrate more sophisticated physical-layer techniques into the network-layer cooperation. Designed properly, the enhanced cooperative techniques are expected to yield higher performance gains.

This chapter is divided into two parts. In Section 3.2, we formulate the problem in a cognitive cooperation system, and employ two enhancement techniques at the cognitive relay: (i) dynamic decode-and-forward, and (ii) superposition coding. In contrast to the previously reported work in cognitive cooperation, where relaying is only enabled during the periods of source silence [12, 13, 14, 32], here, a dynamic decode-and-forward (DDF) scheme is first elaborated as a physical-layer relaying strategy [33, 34, 35]. The proposed method allows the relay node to also provide diversity benefits simultaneously with the source transmission, if the relay node is able to decode the source message based on partial reception. In addition,

the conventional cognitive cooperation is enhanced via an adaptive superposition coding technique. The proposed scheme allows the relay node to switch between a single-queue service and a simultaneous service of several relaying queues by using a 1-bit feedback channel. The proposed schemes are analyzed from a networking perspective; by assuming bursty traffic arrivals, the stable throughput region for a two-user configuration is characterized explicitly. The combination of DDF and superposition coding with cognitive relaying introduces a new concept for protocol-level cooperation in wireless networks.

Then, in Section 3.3, we extend the single-packet reception channel model to a multipacket reception model. Allowing the relay to transmit only when the other user is idle, is motivated by the objective of avoiding interference among multiple simultaneous transmissions which may result in unsuccessful reception of any transmitted packet. However, such destructive interference assumption does not accurately capture the behavior of wireless channels; in wireless environments, a packet might survive the interference caused by concurrent transmissions, if the received signal-to-interference-plus-noise ratio (SINR) exceeds the threshold required for correct decoding. Furthermore, if the receiver is equipped with a multiuser detector, it may decode packets successfully from more than one transmitter at a time. The multipacket reception (MPR) model has been studied extensively without user cooperation, focusing on the stability issue [19,21,36,37]. In the cognitive cooperation system where the relay can possibly transmit together with the other user, there is a trade-off between the scheduling of simultaneous transmissions with reduced success probability against single transmission. Based on these considerations, we investi-

gate in a two-user multi-access system where the user with a better user-destination channel may act as the relay for the other. By observing the other user's queue state, the relay is scheduled to transmit opportunistically with probability p if the other user is transmitting. We are then interested in characterizing the optimal scheduling probability that maximizes the stable throughput region of the network. We show that by optimally choosing the scheduling probability, which in general is a function of the arrival rate, the stable throughput region can be a convex polyhedral region for certain channels; this strictly contains the stable throughput region of the conflict-free scheduling which is bounded by a straight line.

3.2 DDF and Superposition Coding

3.2.1 Model

We assume a simple multiple-access relay channel (MARC) configuration consisting of two primary users A, B, one common cognitive relay S and one common destination D as shown in Fig. 3.1. Both primary users (A, B) have a buffer of infinite capacity to store incoming packets and Q_i denotes the queue of the i-th user ($i \in \{A, B\}$). Time is considered to be slotted with a normalized slot duration ($T = 1$) and both users share the channel through Time-Division Multiple-Access (TDMA) scheduling which allows the users A and B to access the channel over disjoint fractions of time ω_A and ω_B, respectively, where $0 \leqslant \omega_A \leqslant 1$, $0 \leqslant \omega_B \leqslant 1$ and $\omega_A + \omega_B = 1$. The packet length for the i-th user is fixed, and contains R_i bits which are transmitted during one slot, thus resulting in a spectral efficiency of R_i

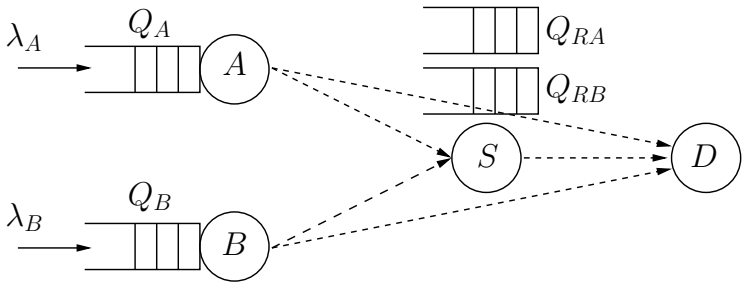

Figure 3.1: The system model; Users A and B communicate with a common destination D via a TDMA policy.

bits/slot. The packet arrivals at the users are independent and stationary Bernoulli processes with mean λ_i (packets per slot) for the i-th user.

The retransmission process is based on an Acknowledgement/Negative Acknowledgement (ACK/NACK) mechanism, in which short-length error-free packets are broadcast over a separate narrow-band channel in order to inform the network of that packet's reception status. When supporting the cognitive cooperation, the relay node is equipped with two relaying queues Q_{RA}, Q_{RB}, to relay some packets from the source users A and B respectively. The relaying queues Q_{Ri} ($i \in \{A, B\}$) are formed in the following way: when the i-th user transmits a packet, if the destination D decodes the packet successfully, it sends back an ACK and the packet exits the system; otherwise, if D cannot decode the packet but S decodes the packet, S sends back an ACK and keeps the packet in its queue Q_{Ri} for retransmission, while the i-th source drops the packet; if neither S nor D decodes the packet, the packet remains at Q_i for retransmission in the next TDMA frame.

All wireless links exhibit fading and additive white Gaussian noise (AWGN). The fading is assumed to be stationary, with frequency non-selective Rayleigh block fading. This means that the fading coefficients $h_{i,j}$ (for the $i \to j$ link) remain con-

stant during one slot, but change independently from one slot to another according to a circularly symmetric complex Gaussian distribution with zero mean and variance $\sigma_{i,j}$ for the link $i \rightarrow j$. Furthermore, the variance of the AWGN is assumed to be normalized with zero mean and unit variance which corresponds to an average signal-to-noise ratio (SNR) equal to $\rho_{i,j} = P_0\sigma_{i,j}^2$, where P_0 denotes the transmission power common to all nodes. Each link $i \rightarrow j$ is characterized by the success probability $f_{i,j}(R_i) \triangleq \mathbf{P}\left[\log(1+\rho_{i,j}|h_{i,j}|^2) > R_i\right] = \exp(-\frac{2^{R_i}-1}{\rho_{i,j}})$ which denotes the probability that the link $i \rightarrow j$ is not in outage ($\overline{f_{i,j}}(R_i) = 1 - f_{i,j}(R_i)$ denotes the outage probability). An outage occurs when the instantaneous capacity of the link $i \rightarrow j$ is lower than the transmitted spectral efficiency rate R_i. The channel is assumed to be known only at the receivers (not at the transmitters) and perfect radio sensing is assumed for the cognitive relay which allows the node S to access the channel only when the latter is unutilized.

Under our model, the arrival processes and the departure processes of each queue are jointly stationary, we can apply Loynes' Theorem [16] to determine queue stability, which requires the average arrival rate to be less than the average service rate. The stable throughput region is defined as the union of all arrival rate vectors (in packets/slot) such that all queues in the network remain stable. The assumed TDMA structure decouples the queues of the terminals, and, hence, bypasses the thorny aspects of stability that arise when the queues interact.

3.2.2 Dynamic Decode-and-Forward

The DDF scheme which enables a partial relaying during the source transmission has been well studied from an information-theoretic standpoint [33, 34], but here, we use it to assess its impact on improving the stable throughput region. According to the DDF protocol, the source codeword is divided into M blocks which are delivered to the destination in two dependent phases during the time of one slot. More specifically, in the first phase of the protocol which is called listening phase, the source broadcasts the source message towards the relay and the destination. During this phase, the destination receives the source message via the direct link without diversity. At a certain instant, referred to as the decision time m, if the relay is able to decode the source information message (based on the received m blocks) before receiving the whole codeword, it starts to assist the source transmission by providing cooperative diversity. More specifically, after successful decoding, the relay can correctly anticipate the future transmissions from the source since it knows the source codebook by using an Alamouti constellation.

The duration of the second phase is from the decision time to the end of the codeword and is a random variable which depends on the instantaneous quality of the $i \rightarrow S$ link. If we restrict the decision time to coincide with the end of a block, the decision time can be represented by a random variable $m \in [1, 2 \ldots, M]$ ($m = M$ corresponds to the case where the relay does not help the destination).

Before we utilize the DFF technique, we first characterize the stable throughout regions of the non-cooperation (NC) scheme and the conventional cooperation

(CC) scheme, which can serve as a baseline for evaluating the impact of DDF on affecting the stable throughput performance. Without cooperation, the users A, B deliver their data via their direct links to the destination without any assistance from the relay S. Since both the arrival processes and the service processes are stationary and since the queues operate without interdependence (because of the time-division between the two users), the stability analysis of the two users' queues can be carried out separately for each queue; and by applying the Loynes' Theorem, for a fixed (ω_A, ω_B), the stability condition is defined by

$$\lambda_i < \mu_i^{(max)} = \omega_i f_{i,D}(R_i) \qquad \text{with } i \in \{A, B\} \tag{3.1}$$

where $\mu_i^{(max)}$ denotes the maximum service rate for the i-th user's queue. If we take the union over all (ω_A, ω_B) such that $\omega_A + \omega_B \leq 1$, we obtain the non-cooperative stable throughput region which is

$$L_{\text{NC}} = \left\{ (\lambda_A, \lambda_B) : \frac{\lambda_A}{f_{A,D}(R_A)} + \frac{\lambda_B}{f_{B,D}(R_B)} < 1 \right\} \tag{3.2}$$

Under the conventional cooperation scheme, the relaying queue Q_{Ri} is served only when the time slot is assigned to the i-th user, and the i-th user's queue Q_i is empty. At each user's time slot, the user transmits a packet if it is backlogged; and the packet is removed from its queue if it is decoded by either the destination or the relay, or both of them. By Loynes' Theorem, the i-th user's queue is stable

if and only if

$$\lambda_i < \mu_i^{(\max)} = \omega_i \Bigg[f_{i,D}(R_i) + \underbrace{[1 - f_{i,D}(R_i)] f_{i,S}(R_i)}_{\triangleq g_{i,S}(R_i)} \Bigg] \tag{3.3}$$

Then we analyze the stability of the cognitive relay S. A packet transmitted by the i-th user, if not successfully decoded by the destination but decoded by relay S, will be removed from the user's queue Q_i and stored at the relaying queue Q_{Ri} for cognitive retransmission, thus forming the arrival process at Q_{Ri}. Relay S serves its relaying queues when the corresponding time slot becomes idle, and the transmitted packet will leave the relaying queue if it is decoded by the destination. The average arrival rate (λ_{Ri}) and average service rate ($\mu_{Ri}^{(\max)}$) of the relaying queues Q_{Ri} for $i \in \{A, B\}$ can be easily obtained as

$$\lambda_{Ri} = \omega_i \mathbf{P}\left[Q_i \neq 0\right] g_{i,S}(R_i) = \omega_i \frac{\lambda_i}{\mu_i^{(\max)}} g_{i,S}(R_i) \tag{3.4}$$

$$\mu_{Ri}^{(\max)} = \omega_i \mathbf{P}\left[Q_i = 0\right] f_{S,D}(R_i) = \omega_i \left[1 - \frac{\lambda_i}{\mu_i^{(\max)}}\right] f_{S,D}(R_i) \tag{3.5}$$

The relaying queue Q_{Ri} is stable if and only if $\lambda_{Ri} < \mu_{Ri}^{(\max)}$. Given the fixed (ω_A, ω_B), the stability condition for the whole network is the intersection of the stability conditions for both the user queues and relaying queues. Finally by taking the union over all (ω_A, ω_B), the stable throughput region for the conventional cognitive

cooperation is given by

$$L_{\mathrm{CC}} = \left\{ (\lambda_A, \lambda_B) : \frac{\lambda_A[g_{A,S}(R_A) + f_{S,D}(R_A)]}{f_{S,D}(R_A)[f_{A,D}(R_A) + g_{A,S}(R_A)]} + \frac{\lambda_B[g_{B,S}(R_B) + f_{S,D}(R_B)]}{f_{S,D}(R_B)[f_{B,D}(R_B) + g_{B,S}(R_B)]} < 1 \right\} \tag{3.6}$$

3.2.2.1 Non-Cognitive DDF (NC-DDF)

In the NC-DDF scheme, the cognitive cooperation is not enabled, and the relay can only possibly provide cooperative diversity during the source transmission. If the destination cannot decode the packet successfully by the end of the second phase, the source packet remains in the source queue and will be retransmitted by the source at the next assigned time slot using the same DDF policy. Furthermore, the relay node will drop the packet, even if it achieves to decode it, and will treat the retransmitted packet as a new packet. Again, the reduction in efficiency from this restriction is accepted for simplifying the analysis. The system model for the DDF case can be written as

$$y_k^{(i)} = \begin{cases} h_{i,D}x_{i,k} + w_k & k = 1 \dots m \\ \sqrt{|h_{i,D}|^2 + |h_{S,D}|^2}x_{i,k} + v_k & k = m+1 \dots M \end{cases} \tag{3.7}$$

where $y_k^{(i)}$ denotes the received signal at the destination for the k-th block and the i-th user ($i \in \{A, B\}$), $x_{i,k}$ denotes the transmitted k-th block for the i-th user, $w_k, v_k \sim \mathcal{CN}(0,1)$ denote the normalized AWGN noise for the two phases of the protocol with a zero mean and a variance equal to one, respectively, and

$m = \min\left\{M, \left\lceil \frac{MR_i}{\log(1+P_0|h_{i,S}|^2)} \right\rceil \right\}$ denotes the decision time when the relay node starts to assist the transmission [33] (practical issues concerning the decision time are beyond the scope of this dissertation [35]). It is worth noting that the orthogonality during the second phase of the DDF protocol is achieved by using an Alamouti scheme [34, 35]. In this case, the achievable rate for $m = k$ and i.i.d. Gaussian inputs is given by [35], which is

$$\begin{aligned} C_k^{(i)} = & \underbrace{\frac{k}{M}\log(1+P_0|h_{i,D}|^2)}_{\text{non-cooperation}} \\ & + \underbrace{\frac{M-k}{M}\log(1+P_0|h_{i,D}|^2+P_0|h_{S,D}|^2)}_{\text{cooperation}} \end{aligned} \tag{3.8}$$

The probability that the DDF link is not in outage can be written as

$$\begin{aligned} f_{i,S,D}(R_i) &= \sum_{k=1}^{M} \mathbf{P}\left[m=k\right] \mathbf{P}\left[C_k^{(i)} > R_i\right] \\ &\simeq \sum_{k=1}^{M} \mathbf{P}\left[m=k\right] f_{i,S,D}^{(k)}(R_i) \end{aligned} \tag{3.9}$$

$$\begin{aligned} \mathbf{P}\left[m=k\right] =& \mathbf{P}\left[k\log(1+P_0|h_{i,S}|^2) > MR_i > (k-1)\log(1+P_0|h_{i,S}|^2)\right] \\ =& \exp\left(-\frac{2^{\frac{MR_i}{k-1}}-1}{\rho_{i,S}}\right) - \exp\left(-\frac{2^{\frac{MR_i}{k}}-1}{\rho_{i,S}}\right) \end{aligned} \tag{3.10}$$

where $\log(\cdot)$ denotes the base 2 logarithm. We note that the probability $f_{i,S,D}^{(k)}(R_i)$ is computable precisely but at great computational cost. In order to simplify the analysis, we approximate the achievable rate of the DDF protocol, that has a decision

time equal to k, as follows [35]:

$$C_k^{(i)} = \frac{k}{M}\log(1+P_0|h_{i,D}|^2) + \frac{M-k}{M}\log(1+P_0|h_{i,D}|^2+P_0|h_{S,D}|^2)$$

(Since $X+Y \leq \max[X,Y]$ where $X,Y \in \mathcal{R}^+$ we have)

$$\geqslant \max\left[\underbrace{\frac{k}{M}\log(1+P_0|h'_{i,D}|^2)}_{\triangleq\zeta}, \underbrace{\frac{M-k}{M}\log(1+P_0|h_{i,D}|^2+P_0|h_{S,D}|^2)}_{\triangleq\psi}\right] \quad (3.11)$$

where $h'_{i,D}$ denotes an independent channel coefficient with variance $\sigma^2_{i,D}$ (it is used in order to relax the dependency between the two terms). The outage probability for the random variables ζ, ψ (for $\sigma^2_{i,D} \neq \sigma^2_{S,D}$) is given as

$$P_\zeta(z) = \mathbf{P}\left[\zeta \leqslant z\right] = 1 - \exp\left(-\frac{2^{\frac{Mz}{k}}-1}{P_0\sigma^2_{i,D}}\right) \quad (3.12)$$

$$P_\psi(z) = \mathbf{P}\left[\psi \leqslant z\right] = 1 - \exp\left(-\frac{\rho}{\sigma^2_{S,D}}\right) - \frac{\sigma^2_{i,D}}{\sigma^2_{i,D}-\sigma^2_{S,D}}\exp\left(-\frac{\rho}{\sigma^2_{i,D}}\right) \times \left[1-\exp\left(-\frac{\sigma^2_{i,D}-\sigma^2_{S,D}}{\sigma^2_{i,D}\sigma^2_{S,D}}\rho\right)\right] \quad (3.13)$$

where $\rho \triangleq \frac{2^{\frac{Mz}{M-k}}-1}{P_0}$. By applying basic order statistics, the probability of outage for the DDF protocol is written as

$$\overline{f_{i,S,D}}(R_i) = \mathbf{P}\left[\max[\zeta,\psi] < R_i\right] = P_\zeta(R_i)P_\psi(R_i) \quad (3.14)$$

In the numerical results section, we plot both the approximated results using the closed-form expression, as well as the exact stable throughput region; the approximated result is shown to be very close to the true result. The stable throughput

region follows the formulation of the NC scheme (Eq. (3.2)) and thus can be written as

$$L_{\text{NC-DDF}} = \left\{ (\lambda_A, \lambda_B) : \frac{\lambda_A}{f_{A,S,D}(R_A)} + \frac{\lambda_B}{f_{B,S,D}(R_B)} < 1 \right\} \tag{3.15}$$

3.2.2.2 Cognitive DDF (C-DDF)

In the NC-DDF scheme just discussed, cooperation is performed for each source transmission only when the source is active. Therefore, a packet is removed from the source user's queue only when it is decoded correctly at the destination. In the scheme proposed here, a packet is dropped from the user queue also when it is decoded correctly at the relay node, following the principles of the CC scheme. The C-DDF relaying strategy is described as follows: during the source transmission, if the relay node can decode the packet of the i-th user with a decision time equal to m (with $1 \leq m \leq M$), it assists the source transmission for the rest of the codeword; by the end of the whole codeword, if the packet is still not successfully decoded at the destination, the relay node transmits an ACK signal and the source packet is dropped from the user, while the relay node will take the responsibility to retransmit this packet later on. The service of Q_{Ri} follows the rules of the CC protocol and therefore it is activated when the i-th user's slot is sensed to be idle. The C-DDF scheme provides two different types of relaying collaboration: (1) a dynamic cooperation when the source is active based on the DDF protocol, and (2) a cognitive cooperation when the source is idle. By inserting the DDF outage probabilities into

Eq. (3.6), the stable throughput region of the C-DDF scheme becomes

$$L_{\text{C-DDF}} = \Bigg\{ (\lambda_A, \lambda_B) : \frac{\lambda_A \big[v_{A,S}(R_A) + f_{S,D}(R_A) \big]}{f_{S,D}(R_A) \big[f_{A,S,D}(R_A) + v_{A,S}(R_A) \big]} + \frac{\lambda_B \big[v_{B,S}(R_B) + f_{S,D}(R_B) \big]}{f_{S,D}(R_B) \big[f_{B,S,D}(R_B) + v_{B,S}(R_B) \big]} < 1 \Bigg\} \tag{3.16}$$

where $v_{i,S}(R_i) \triangleq [1 - f_{i,S,D}(R_i)] f_{i,S}(R_i)$.

Comparison between NC-DDF and C-DDF: The C-DDF scheme utilizes a dynamic and cognitive relaying cooperation. More specifically, in addition to dynamic cooperation, a cognitive cooperation is enabled. In this subsection, we determine when this alternative type of cooperation is favorable to the non-cognitive DDF in terms of stable throughput region.

Due to the adopted TDMA transmission policy, the stable throughput regions of both DDF schemes are bounded by a straight line. Therefore, to compare the two regions, it is enough to compare the intersection of these lines with the axes. The intersection points for NC-DDF and C-DDF are given as

$$\lambda_i^*(\text{NC-DDF}) = f_{i,S,D}(R_i) \tag{3.17}$$

$$\lambda_i^*(\text{C-DDF}) = \frac{f_{S,D}(R_i)[f_{i,S,D}(R_i) + v_{i,S}(R_i)]}{v_{i,S}(R_i) + f_{S,D}(R_i)} \tag{3.18}$$

It is clear that the stable throughput region of NC-DDF is completely contained inside the corresponding region of C-DDF, if $\lambda_i^*(\text{C-DDF}) > \lambda_i^*(\text{NC-DDF})$ for

$i \in \{A, B\}$. After a little algebra, this condition reduces to

$$f_{S,D}(R_i) > f_{i,S,D}(R_i) \tag{3.19}$$

The above condition reveals that the integration of the "cognitive" relaying to the NC-DDF scheme is beneficial only when the relay-destination link is better than the combined DDF links.

3.2.3 Superposition Coding

The protocols discussed so far, assume that the relay node serves only one non-empty queue at a time when the source becomes silent. In order to boost cooperation, we investigate an adaptive superposition coding technique which allows the relay node to serve simultaneously both relaying queues via a power split technique [38, Ch. 6]. More specifically, when the relay node senses an idle time slot and the channel from the relay to the destination is good enough, it is allowed to superimpose packets from both relaying queues to transmit to the destination with a total power equal to P_0. To do this, a feedback channel that informs the relay node about the instantaneous condition of the relay-destination link, is introduced. The feedback message consists of one bit, informing the relay node whether the relay-destination link can support the superimposed spectral efficiency $R_A + R_B$ or not. It is assumed to be error-free and always available at the relay node. In the case that the relay-destination channel cannot support the total spectral efficiency, all the power is allocated to the corresponding relaying data flow (which corresponds

to the spare source time slot), and the scheme reduces to the conventional cognitive relaying. To support both data flows when the instantaneous channel capacity is higher than the total spectral efficiency, multiuser detection schemes like interference cancelation (IC) can be implemented at the destination [38, Ch. 9] so that both data flows can be decoded. The transmission schemes are summarized as follows:

1. If the feedback informs that the relay-destination link can support the total spectral efficiency $R_A + R_B$ (feedback is equal to 1):
   - a) If $Q_{RA} \neq 0, Q_{RB} \neq 0$, the relay serves both relaying queues, one packet from Q_{RA}, and one packet from Q_{RB}.
   - b) If $Q_{RA} \neq 0, Q_{RB} = 0$, the relay serves a packet from the relaying queue Q_{RA}.
   - c) If $Q_{RA} = 0, Q_{RB} \neq 0$, the relay serves a packet from the relaying queue Q_{RB}.
2. Otherwise, if the feedback informs that the relay-destination link cannot support the superimposed spectral efficiency $R_A + R_B$ (feedback is equal to 0), then if the i-th user's slot is sensed to be idle, the relay serves a packet from the Q_{Ri} relaying queue without superposition coding (conventional transmission).

It is worth noting that the feedback is a form of strongly compressed side-information that concerns the relay-destination link only and consists of 1 bit; therefore the related complexity overhead is negligible. Furthermore, a feedback equal to 1, guarantees that a packet can be correctly decoded at the destination at

the sum rate. In the following part of this section, we integrate the abovementioned adaptive superposition mechanism into the cognitive protocols (CC, C-DDF) and carry out the stability analysis.

3.2.3.1 Conventional Cooperation with Superposition (S-CC)

The stability analysis of the user queues under the S-CC scheme is similar to that of the CC scheme, and the stability condition of the user queues is given by

$$\left.\begin{aligned} \lambda_i < \omega_i \big[\underbrace{f_{i,D}(R_i) + g_{i,S}(R_i)}_{\triangleq \Gamma_i} \big] \\ \omega_A + \omega_B \leq 1 \end{aligned}\right\}$$
$$\Longrightarrow L_{\text{S-CC}^1} = \left\{ (\lambda_A, \lambda_B) : \frac{\lambda_A}{\Gamma_A} + \frac{\lambda_B}{\Gamma_B} < 1 \right\} \tag{3.20}$$

Then the stability condition for the relaying queue Q_{Ri} ($i \in \{A, B\}$) can be found to be:

$$\begin{aligned} &\lambda_{Ri} < \mu_{Ri}^{(\max)} \\ &\Longrightarrow \frac{\lambda_i}{\Gamma_i} g_{i,S}(R_i) < \omega_i \underbrace{\left[1 - \frac{\lambda_i}{\omega_i \Gamma_i}\right]}_{\mathbf{P}[Q_i = 0]} \underbrace{\Big[\theta \cdot 1 + \big[1 - \theta\big] f_{S,D}(R_i)\Big]}_{\triangleq \delta_i} \\ &\qquad\qquad + \omega_{\bar{i}} \underbrace{\left[1 - \frac{\lambda_{\bar{i}}}{\omega_{\bar{i}} \Gamma_{\bar{i}}}\right]}_{\mathbf{P}\left[Q_{\bar{i}} = 0\right]} \theta \cdot 1 \end{aligned} \tag{3.21}$$

$$\omega_A + \omega_B \leq 1 \tag{3.22}$$

$$\Longrightarrow L_{\text{S-CC}^2} = \Bigg\{(\lambda_A, \lambda_B) : \frac{\lambda_A}{\Gamma_A}\left[\frac{g_{A,S}(R_A)+\delta_A}{\delta_A-\theta}+\frac{\theta}{\delta_B-\theta}\right] + \frac{\lambda_B}{\Gamma_B}\left[\frac{g_{B,S}(R_B)+\delta_B}{\delta_B-\theta}+\frac{\theta}{\delta_A-\theta}\right] - \theta\left[\frac{1}{\delta_A-\theta}+\frac{1}{\delta_B-\theta}\right] < 1\Bigg\} \tag{3.23}$$

where $\theta \triangleq \mathbf{P}\left[\text{feedback is equal to 1}\right] = f_{S,D}(R_A + R_B)$ denotes the success probability for the sum rate and $\bar{i}$ denotes the complementary of $i \in \{A, B\}$. The arrival rate to the relaying queue Q_{Ri} is the same as that in the CC scheme; regarding the service process of Q_{Ri}, a packet will depart from Q_{Ri} if any of the following events happens: (1) the slot is assigned to the i-th user, $Q_i = 0$, and the feedback is equal to 1, in which case the packet departs with probability 1, (2) the slot is assigned to the i-th user, $Q_i = 0$, and the feedback is equal to 0, in which case the packet departs with probability $f_{S,D}(R_i)$, (3) the slot is assigned to the $\bar{i}$-th user, $Q_{\bar{i}} = 0$, and the feedback is equal to 1, the packet departs with probability 1. These events are disjoint and the service rate of Q_{Ri} can be obtained as in Eq. (3.21). The resulting stable throughput region of S-CC is given by the intersection of the two regions $L_{\text{S-CC}^1} \bigcap L_{\text{S-CC}^2}$, which is easily shown to be equal to $L_{\text{S-CC}^2}$.

3.2.3.2 Cognitive DDF with Superposition (SC-DDF)

The superposition approach can be further built on the C-DDF protocol. In a manner similar to the S-CC scheme, it can be shown that the stable throughput region of the SC-DDF scheme is given by Eq. (3.23) by replacing $g_{i,S}(R_i)$, Γ_i with $v_{i,S}(R_i)$ and $\gamma_i \triangleq f_{i,S,D}(R_i) + v_{i,S}(R_i)$, respectively.

3.2.4 Numerical Results

Numerical plots are shown to support the analytical results of the proposed schemes and compare their performance. For clarity of presentation, a symmetric configuration is considered with $R_A = R_B = 2$ bits per channel per use (BPCU), $\rho_{A,D} = \rho_{B,D} = 5$ dB, $\rho_{A,S} = \rho_{B,S} = 12$ dB, $\rho_{S,D} = 30$ dB, and $M = 3$ for the DDF-based schemes. Fig. 3.2 plots the stable throughput regions for the investigated protocols. As expected, cooperation increases the stable throughput region as it overcomes deep-fading of the direct links, resulting in faster emptying of the user queues. Regarding the enhanced cooperative methods, it can be seen that the NC-DDF scheme provides a superior stable throughput region than the CC scheme ($L_{\text{NC-DDF}} \supset L_{\text{CC}}$). The NC-DDF scheme allows the relay node to assist the source transmission when the source is active but not during the periods of source silence which creates more relaying opportunities than the CC scheme (for most of the time the source slots are not idle unless the traffic is light). Furthermore, the integration of the cognitive relaying (relaying when the source is sensed to be idle) to the NC-DDF scheme, improves further the stable throughput region (C-DDF). The same observation is made for the integration of the superposition coding into the cognitive relaying, that the stable throughput region of the S-CC scheme strictly contains the corresponding region of the CC scheme ($L_{\text{S-CC}} \supset L_{\text{CC}}$). Moreover, the combination of DDF with superposition cognitive relaying (SC-DDF) provides the largest stable throughput region and is introduced as an efficient cross-layer technique for bursty cooperative applications.

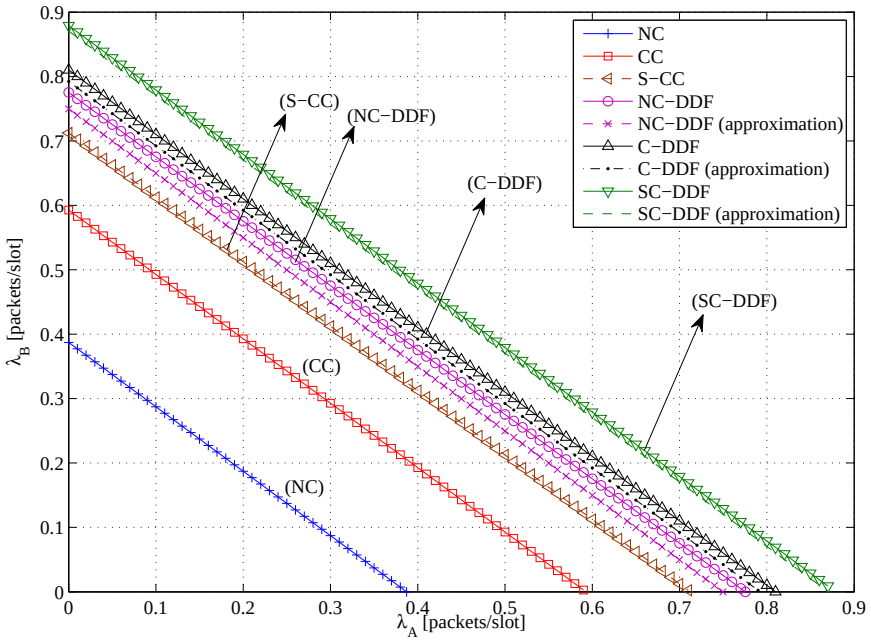

Figure 3.2: Stable throughput regions for the NC, CC, S-CC, NC-DDF, C-DDF and SC-DDF; $R_A = R_B = 2$ BPCU, $\rho_{A,D} = \rho_{B,D} = 5$ dB, $\rho_{A,S} = \rho_{B,S} = 12$ dB, $\rho_{S,D} = 30$ dB, and $M = 3$.

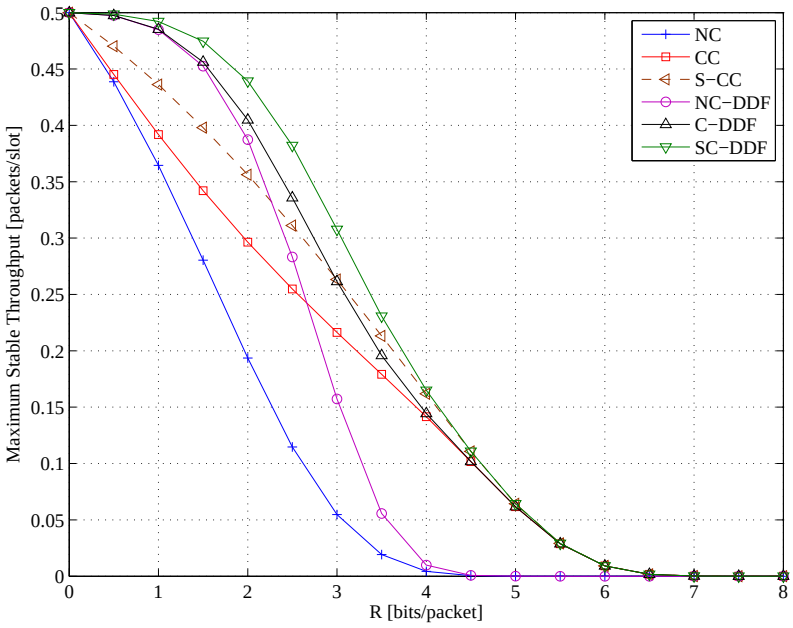

Figure 3.3: Maximum stable throughput versus bits/packet for the NC, CC, S-CC, NC-DDF, C-DDF and SC-DDF; $\lambda_A = \lambda_B$, $\rho_{A,D} = \rho_{B,D} = 5$ dB, $\rho_{A,S} = \rho_{B,S} = 12$ dB, $\rho_{S,D} = 30$ dB and $M = 3$.

Fig. 3.3 shows the maximum stable throughput (MST) for the above symmetric configuration. MST is defined as the maximum common arrival rate (the arrival processes are the same across source users) that stabilizes the system, rather than the full two-dimensional region. The first important observation is that the "cognitive" cooperation becomes more important as the rate (in bits/packet) increases, and hence, the MST for the NC and NC-DDF protocols decays to zero faster than in the cognitive cooperative schemes as the bit rate per packet increases. The fundamental reason for this property is that the cognitive protocols convey a major part of the traffic from the primary users to the relay-destination link, which has a better successful delivery probability than the direct links. Furthermore, it can be seen that the superposition coding approach significantly increases the MST for all protocols, which is expected as the superposition coding technique uses more efficiently the relay-destination link.

It should be noted here that cooperation for cognitive systems is a beneficial solution only when the direct links are in deep-fading and the relay-destination link is strong enough in order to establish communication, which motivated our particular choice of system parameters.

3.3 Multipacket Reception Capability

3.3.1 Model

As another line to address more realistic and advanced physical-layer model, this section exploits the MPR capability at the physical layer. The slotted multi-

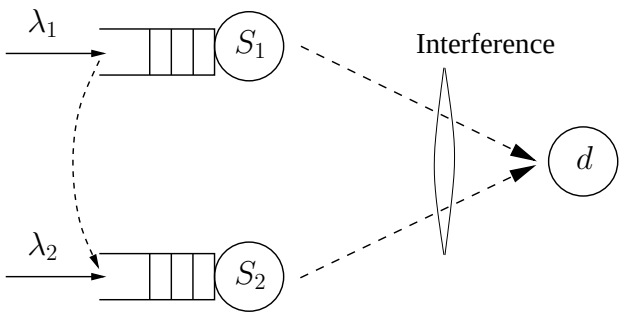

Figure 3.4: The slotted two-user multi-access system. User S_2 acts as the relay for S_1. With an MPR channel model, S_2 is scheduled to transmit opportunistically.

access channel we consider in this part of work is shown in Fig. 3.4. It consists of two source users S_1 and S_2, and a common destination node d. Packets of fixed length arrive to the source node S_i independently according to a Bernoulli process with rate λ_i (packets per slot) in each time slot, $i \in \{1, 2\}$. The transmission of one packet takes the duration of exactly one time slot, and the packets to be transmitted are stored in the buffer of the source nodes, which are assumed to be of infinite size. The MPR channel model is similar to that specified in [19, 21]; given that the set $\mathcal{M}$ of source users transmit, the probability that node n decodes the packet from the source m (for $m \in \mathcal{M}$) is denoted by

$$q^{(n)}_{m|\mathcal{M}} = \mathbf{P}\left[\text{packet from } m \text{ is received at node } n | \text{users in set } \mathcal{M} \text{ transmit}\right] \quad (3.24)$$

This MPR model captures the effects of fading, attenuation and interference at the physical layer. If a packet is not successfully decoded, the receiver simply drops the packet. In the two-user cooperative multiple access channel, the source users are ordered in such a way that S_2 has a better channel to the destination than S_1 does. Thus, S_2 serves to relay S_1's packets if these packets are not decoded by the

destination but decoded by S_2 successfully. In such settings, there are five reception probabilities associated with this network, specifically,

$$\begin{aligned}
q_{1|1}^{(d)} &= \mathbf{P}\left[\text{packet from } S_1 \text{ is received at } d \mid \text{only } S_1 \text{ transmits}\right] \\
q_{2|2}^{(d)} &= \mathbf{P}\left[\text{packet from } S_2 \text{ is received at } d \mid \text{only } S_2 \text{ transmits}\right] \\
q_{1|1}^{(2)} &= \mathbf{P}\left[\text{packet from } S_1 \text{ is received at } S_2 \mid \text{only } S_1 \text{ transmits}\right] \\
q_{1|1,2}^{(d)} &= \mathbf{P}\left[\text{packet from } S_1 \text{ is received at } d \mid \text{both } S_1 \text{ and } S_2 \text{ transmit}\right] \\
q_{2|1,2}^{(d)} &= \mathbf{P}\left[\text{packet from } S_2 \text{ is received at } d \mid \text{both } S_1 \text{ and } S_2 \text{ transmit}\right]
\end{aligned} \tag{3.25}$$

We assume throughout that interference cannot increase the reception probability, hence, the probability that a packet transmitted by S_i is decoded by d given that only S_i transmits, is greater than the corresponding probability given that both S_1 and S_2 transmit, that is,

$$q_{1|1}^{(d)} > q_{1|1,2}^{(d)}, \qquad\qquad q_{2|2}^{(d)} > q_{2|1,2}^{(d)} \tag{3.26}$$

With these reception probabilities, the condition that S_2 has a relatively better user-destination channel than S_1 does can be expressed by

$$q_{2|2}^{(d)} > q_{1|1}^{(d)}, \qquad\qquad q_{2|1,2}^{(d)} > q_{1|1,2}^{(d)} \tag{3.27}$$

Nodes are equipped with single transceivers so that they cannot transmit and receive at the same time. Therefore, S_2 can decode the packet transmitted by S_1 with positive probability if only S_1 transmits, as is illustrated by the reception probability

$q_{1|1}^{(2)}$. We also assume instantaneous and error-free acknowledgements (ACKs), which are broadcast to the network over a separate channel with negligible bandwidth. With all these settings, the relaying process of the network is as follows: when S_1 transmits a packet and S_2 remains silent, if the destination d decodes the packet successfully, it sends back an ACK and the packet exits the network; otherwise, if d doesn't decode the packet but S_2 decodes the packet, S_2 sends back an ACK and queues the packet for retransmission, while S_1 drops that packet upon receiving the ACK from S_2; if neither S_2 nor d decodes the packet, the packet remains at S_1's queue for retransmission later on.

When opportunistic scheduling at S_2 is carried out, the queues at S_1 and S_2 interact in a complicated manner which makes the stability analysis difficult to track. We will adopt the stochastic dominance approach introduced in [17] to decouple the queues. Then for each single queue, a primary tool we use to determine queue stability is Loynes' Theorem [16].

3.3.2 Opportunistic Cooperation Scheme

In the context of this considered multi-access system, we can derive the stable throughput regions of the non-cooperation and conventional cognitive cooperation schemes similarly as in Section 3.2.2. The techniques used are the same, so we omit the detailed analysis and present only the results here.

The stable throughput region of the non-cooperation (NC) scheme is given by

$$\Re_{\text{NC}} = \left\{ (\lambda_1, \lambda_2) : \frac{\lambda_1}{q_{1|1}^{(d)}} + \frac{\lambda_2}{q_{2|2}^{(d)}} < 1 \right\} \tag{3.28}$$

In the conventional cooperation (CC) scheme, S_2 accesses the channel only when it senses an idle time slot of S_1 (we assume perfect sensing). The resulting stable throughput region is given by

$$\Re_{\text{CC}} = \left\{ (\lambda_1, \lambda_2) : \frac{\left(q_{1|1}^{(2)} + q_{2|2}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)}\right) \lambda_1}{q_{2|2}^{(d)} \left(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)}\right)} + \frac{\lambda_2}{q_{2|2}^{(d)}} < 1 \right\} \tag{3.29}$$

Similarly as before, it can be easily verified that the stable throughput region of the conventional cooperation scheme strictly outer-bounds the non-cooperative region, that is, $\Re_{\text{CC}} \supset \Re_{\text{NC}}$.

In contrast to the conventional cooperation where S_2 only transmits during the idle time slot of S_1, under the Opportunistic Cooperation (C-OPP) scheme, S_2 observes the queue length Q_1 at S_1, and is scheduled to transmit opportunistically [39, 40]: if $Q_1 = 0$, S_2 will transmit with probability 1 if it is backlogged; otherwise, if $Q_1 \neq 0$ and hence S_1 transmits, S_2 is scheduled to transmit with probability p if it has packets. In the latter case, if S_2 is scheduled to transmit together with S_1, the two transmissions will cause interference to each other; and since S_2 cannot receive and transmit at the same time, the relay-assistance of S_2 is sacrificed for taking advantage of the multipacket reception capability, under which both transmissions can reach the destination successfully with probability $q_{1|1,2}^{(d)} q_{2|1,2}^{(d)}$,

strictly greater than zero. It should be noted here that when S_2 transmits a packet, it randomly picks up a packet from its queue, either its own packet or a packet relayed from S_1; we will show that the stability analysis is the same for both situations.

With such opportunistic scheduling, the two queues interact in such a way that the service rate of each queue depends on whether the other queue is empty or not; in addition, the arrival process to S_2 from S_1 depends on the departure process from S_1. A useful tool to bypass this thorny problem is the stochastic dominance approach [17]. By utilizing this dominance approach, a complete description of the stability condition of the C-OPP scheme can be characterized for each fixed scheduling probability p; then, by applying a constrained optimization technique, the achievable stable throughput region is obtained by varying over all possible values of p. The main results are provided in the following theorem.

Theorem 3.1 *By optimizing over all possible scheduling probability $p \in [0,1]$ at S_2, the stable throughput region of the C-OPP scheme ($\Re_{C\text{-}OPP}$) for the MPR channel model is defined as follows:*

1. *if*

$$\begin{aligned}\eta =& q_{1|1}^{(2)}(1-q_{1|1}^{(d)})q_{1|1,2}^{(d)} + (q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)}q_{1|1}^{(d)})q_{2|1,2}^{(d)} \\ & - (q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)}q_{1|1}^{(d)} - q_{1|1,2}^{(d)})q_{2|2}^{(d)} > 0 \end{aligned} \tag{3.30}$$

the achievable stable throughput region (by optimizing the scheduling probability

p) is given by $\Re_{C\text{-}OPP} = \Re_1 \bigcup \Re_2$, *where*

$$\Re_1 = \left\{ (\lambda_1, \lambda_2) : \frac{(q_{2|2}^{(d)} - q_{2|1,2}^{(d)})\lambda_1}{q_{2|2}^{(d)} q_{1|1,2}^{(d)}} + \frac{\lambda_2}{q_{2|2}^{(d)}} < 1, \quad for \ 0 \leq \lambda_1 \leq q_{1|1,2}^{(d)} \right\} \tag{3.31}$$

$$\Re_2 = \left\{ (\lambda_1, \lambda_2) : \frac{q_{1|1}^{(2)}(1 - q_{1|1}^{(d)}) + q_{2|1,2}^{(d)}}{q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - q_{1|1,2}^{(d)}} \lambda_1 + \lambda_2 \right.$$
$$\left. < \frac{q_{1|1}^{(2)}(1 - q_{1|1}^{(d)})(q_{1|1,2}^{(d)} + q_{2|1,2}^{(d)}) + q_{1|1}^{(d)} q_{2|1,2}^{(d)}}{q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - q_{1|1,2}^{(d)}}, \quad for \ \lambda_1 > q_{1|1,2}^{(d)} \right\} \tag{3.32}$$

In the rate region defined by $\lambda_1 \leq q_{1|1,2}^{(d)}$, *the boundary of the subregion* $\Re_1$ *is achieved by setting the optimal scheduling probability to be* $p^* = 1$; *and in the rate region defined by* $\lambda_1 > q_{1|1,2}^{(d)}$, *the optimal* p^* *which achieves the boundary of* $\Re_2$ *is given by* $p^* = \frac{q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - \lambda_1}{q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - q_{1|1,2}^{(d)}}$.

2. *otherwise, that is, if*

$$\begin{aligned} \eta =& q_{1|1}^{(2)}(1 - q_{1|1}^{(d)}) q_{1|1,2}^{(d)} + (q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)}) q_{2|1,2}^{(d)} \\ & - (q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - q_{1|1,2}^{(d)}) q_{2|2}^{(d)} \leq 0 \end{aligned} \tag{3.33}$$

the optimal scheduling probability is given by $p^* = 0$, *which means that* S_2 *only transmits during the idle time slot of* S_1, *hence the C-OPP scheme reduces to the CC scheme, and the stable throughput region is known to be given by Eq. (3.29), which is,*

$$\Re_{C\text{-}OPP} = \left\{ (\lambda_1, \lambda_2) : \frac{\left(q_{1|1}^{(2)} + q_{2|2}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} \right) \lambda_1}{q_{2|2}^{(d)} \left(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} \right)} + \frac{\lambda_2}{q_{2|2}^{(d)}} < 1 \right\} \tag{3.34}$$

Proof: The key part of the dominance approach is as follows: first, construct an appropriate dominant system in which one queue can be decoupled from the other, which makes the stability analysis tractable; then, prove that the dominant system and the original system behave in the same way at the boundary of the stable throughput region. We construct the dominant system $\mathcal{S}'$ which dominates the original system $\mathcal{S}$ in the following fashion:

1. If $Q_1 = 0$ and $Q_2 = 0$, S_2 transmits a dummy packet with probability 1;

2. If $Q_1 \neq 0$ and $Q_2 = 0$, S_2 transmits a dummy packet with probability p.

All the other assumptions, channel models, arrival and reception processes remain unaltered in the dominant system. Since the transmission of dummy packets will not contribute to the throughput but cause interference to the concurrent transmission from the other source, it is true that each queue will have a successful departure in $\mathcal{S}$ whenever it has one in $\mathcal{S}'$. Therefore, the queue lengths of the dominant system can no longer be smaller than the corresponding queue lengths of the original system; and the stability of the dominant system implies the stability of the original system.

In the dominant system $\mathcal{S}'$, the service rate of the queue at S_1 depends on whether S_2 transmits or not: (a) if S_2 transmits together with S_1 (which happens with probability p), the service rate seen by S_1 is $q^{(d)}_{1|1,2}$; (b) if S_2 remains silent when S_1 transmits, S_1 receives the service rate of $q^{(2)}_{1|1} + q^{(d)}_{1|1} - q^{(2)}_{1|1}q^{(d)}_{1|1}$, because the packet is dropped from S_1 when either S_2 or the destination d decodes the packet. Hence,

the average service rate of S_1 is

$$\mu_1 = (1-p)(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)}) + p q_{1|1,2}^{(d)} \tag{3.35}$$

And the queue at S_1 is stable if and only if

$$\lambda_1 < \mu_1 = (1-p)(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)}) + p q_{1|1,2}^{(d)} \tag{3.36}$$

The analysis of the queue at S_2 will be a bit more complicated. First we analyze the total arrival process to S_2. The total arrival process consists of two parts: 1) the external Bernoulli arrivals with rate λ_2; and 2) the arrivals from S_1, with the rate $\lambda_{1\to 2}$ to be calculated. There is an arrival to S_2 from S_1 if the following events happen together: (1) S_1 is non-empty and transmits, which has probability λ_1/μ_1 as S_1's queue is a discrete-time M/M/1 queue, (2) S_2 remains silent, which happens with probability $1-p$, (3) the packet transmitted by S_1 is decoded by S_2 but not decoded by d, which occurs with probability $q_{1|1}^{(2)}(1-q_{1|1}^{(d)})$. These three events are independent, and the total arrival rate to S_2 is

$$\begin{aligned}\lambda_{S_2} = \lambda_{1\to 2} + \lambda_2 =& \mathbf{P}\left[Q_1 \neq 0\right](1-p) q_{1|1}^{(2)}(1-q_{1|1}^{(d)}) + \lambda_2 \\ =& \frac{(1-p) q_{1|1}^{(2)}(1-q_{1|1}^{(d)})}{(1-p)(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)}) + p q_{1|1,2}^{(d)}} \lambda_1 + \lambda_2 \end{aligned} \tag{3.37}$$

Then we analyze the service rate of S_2, which also depends on the queue state at S_1. Specifically, if $Q_1 \neq 0$ and S_1 transmits, S_2 transmits with probability p, and

the packet from S_2 is decoded by d with probability $q^{(d)}_{2|1,2}$; otherwise, if $Q_1 = 0$, S_2 transmits with probability 1 and sees a successful delivery probability $q^{(d)}_{2|2}$. So the average service rate of S_2 is

$$\begin{aligned}\mu_{S_2} =& \mathbf{P}\left[Q_1 = 0\right] q^{(d)}_{2|2} + \mathbf{P}\left[Q_1 \neq 0\right] p q^{(d)}_{2|1,2} \\ =& \left(1 - \frac{\lambda_1}{\mu_1}\right) q^{(d)}_{2|2} + \frac{\lambda_1}{\mu_1} p q^{(d)}_{2|1,2}\end{aligned} \tag{3.38}$$

where μ_1 is given in Eq. (3.35). The queue at S_2 is stable if and only if

$$\lambda_{S_2} < \mu_{S_2} \tag{3.39}$$

with λ_{S_2} and μ_{S_2} as written in Eq. (3.37) and Eq. (3.38) respectively.

The network is stable if both queues are stable; after some simple algebra, the stability condition for a fixed scheduling probability p is defined by

$$\lambda_1 < (1-p)(q^{(2)}_{1|1} + q^{(d)}_{1|1} - q^{(2)}_{1|1} q^{(d)}_{1|1}) + p q^{(d)}_{1|1,2} \tag{3.40}$$

$$\frac{\left(q^{(2)}_{1|1}(1 - q^{(d)}_{1|1}) + q^{(d)}_{2|2}\right) - p\left(q^{(2)}_{1|1}(1 - q^{(d)}_{1|1}) + q^{(d)}_{2|1,2}\right)}{(1-p)\left(q^{(2)}_{1|1} + q^{(d)}_{1|1} - q^{(2)}_{1|1} q^{(d)}_{1|1}\right) + p q^{(d)}_{1|1,2}} \lambda_1 + \lambda_2 < q^{(d)}_{2|2} \tag{3.41}$$

We argue that the boundary of the stable throughput region of the dominant system indeed *coincides* with that of the original system: given that $\lambda_1 < (1-p)(q^{(2)}_{1|1} + q^{(d)}_{1|1} - q^{(2)}_{1|1} q^{(d)}_{1|1}) + p q^{(d)}_{1|1,2}$, if for some λ_2 the queue at S_2 is stable in the dominant system, then the corresponding queue in the original system must be stable; conversely, if for some λ_2 the queue at S_2 is unstable in the dominant system,

then this queue never empties, and S_2 always transmits source information when it accesses the channel. In other words, S_2 will not transmit dummy packets, and as long as S_2 never empties, the dominant system and the original system behave exactly in the same way. Thus, we can conclude that the original system and the dominant system are indistinguishable at the boundary points, and $\Re_{\text{C-OPP}}$ is indeed the union of rate pairs (λ_1, λ_2) constrained by Eq. (3.40) and Eq. (3.41) as p varies over $[0, 1]$.

To obtain the closure, we utilize the constrained optimization technique similarly as in [19]. We fix λ_1 and maximize λ_2 as p varies over $[0, 1]$. By replacing λ_1 by x and λ_2 by y, the boundary of the stable throughput region given by Eq. (3.40) and Eq. (3.41) for a fixed p can be written as

$$y = q_{2|2}^{(d)} - \frac{\left(q_{1|1}^{(2)}(1-q_{1|1}^{(d)}) + q_{2|2}^{(d)}\right) - p\left(q_{1|1}^{(2)}(1-q_{1|1}^{(d)}) + q_{2|1,2}^{(d)}\right)}{(1-p)\left(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)}q_{1|1}^{(d)}\right) + pq_{1|1,2}^{(d)}} x$$
$$\text{for} \quad 0 \le x \le (1-p)(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)}q_{1|1}^{(d)}) + pq_{1|1,2}^{(d)} \tag{3.42}$$

Now we consider the following constrained optimization problem

$$\max_{p\in[0,1]} y = \max_{p\in[0,1]} q_{2|2}^{(d)} - \frac{\left(q_{1|1}^{(2)}(1-q_{1|1}^{(d)}) + q_{2|2}^{(d)}\right) - p\left(q_{1|1}^{(2)}(1-q_{1|1}^{(d)}) + q_{2|1,2}^{(d)}\right)}{(1-p)\left(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)}q_{1|1}^{(d)}\right) + pq_{1|1,2}^{(d)}} x \tag{3.43}$$

Differentiate it with respect to p gives

$$\frac{dy}{dp} = \frac{\eta x}{\left((1-p)(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)}q_{1|1}^{(d)}) + pq_{1|1,2}^{(d)}\right)^2} \tag{3.44}$$

with η as given by Eq. (3.30) or Eq. (3.33). The denominator of Eq. (3.44) is strictly greater than zero; and the numerator excluding x, which is η, can be either positive or negative.

- In the case that $\eta > 0$, the first derivative $\frac{dy}{dp}$ is strictly positive and y is an increasing function of p. Thus, it appears that the optimal value of p^* is 1. But caution is needed here. As seen by Eq. (3.42), this constraint is valid only for $x \leq (1-p)(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)}) + p q_{1|1,2}^{(d)}$. Clearly, p^* can take the value 1 if and only if $x \leq q_{1|1,2}^{(d)}$. So in the subregion defined by $0 \leq x \leq q_{1|1,2}^{(d)}$, the optimal probability is $p^* = 1$; substituting $p = 1$ into Eq. (3.43) gives the boundary of the subregion characterized by $\Re_1$ in Eq. (3.31). Now consider x for $x > q_{1|1,2}^{(d)}$. Since y increases with p, and p satisfies $p \leq \frac{q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - x}{q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - q_{1|1,2}^{(d)}}$ according to Eq. (3.42), the optimal p is given by $p^* = \frac{q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - x}{q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)} q_{1|1}^{(d)} - q_{1|1,2}^{(d)}}$. By substituting p^* into Eq. (3.43), we obtain the boundary of the subregion as characterized by $\Re_2$ in Eq. (3.32).

- In the case that $\eta \leq 0$, we have $\frac{dy}{dp} \leq 0$ for all $p \in [0, 1]$, and so y is a decreasing function of p in the range of all possible values of x. Hence, the optimal p^* is equal to zero, and the opportunistic cooperation scheme reduces to the conventional cooperation scheme, with the stable throughput region given by Eq. (3.29). ■

A) Comparison between CC and C-OPP: When the condition expressed in Eq. (3.30) is satisfied, $\Re_{\text{C-OPP}}$, characterized by $\Re_1 \bigcup \Re_2$, becomes a convex polyhedron. This

can be easily verified, and the procedure is as follows: both subregions $\Re_1$ and $\Re_2$ are bounded by a straight line, and the intersection point connecting the two straight lines can be calculated to be $(\lambda_1'(\text{C-OPP}), \lambda_2'(\text{C-OPP}))$, where

$$\lambda_1'(\text{C-OPP}) = q_{1|1,2}^{(d)}, \qquad \lambda_2'(\text{C-OPP}) = q_{2|1,2}^{(d)} \tag{3.45}$$

From Eq. (3.31), the intersection point of the line which bounds $\Re_1$ with the y-axis is

$$\lambda_2^*(\Re_1) = q_{2|2}^{(d)} \quad (\text{for } \lambda_1^*(\Re_1) = 0) \tag{3.46}$$

and from Eq. (3.32), the intersection point of the line which bounds $\Re_2$ with the x-axis is

$$\lambda_1^*(\Re_2) = \frac{q_{1|1}^{(2)}(1 - q_{1|1}^{(d)})(q_{1|1,2}^{(d)} + q_{2|1,2}^{(d)}) + q_{1|1}^{(d)} q_{2|1,2}^{(d)}}{q_{1|1}^{(2)}(1 - q_{1|1}^{(d)}) + q_{2|1,2}^{(d)}} \quad (\text{for } \lambda_2^*(\Re_2) = 0) \tag{3.47}$$

If we form a straight line by connecting these two points, and denote it by L, it turns out that $(\lambda_1'(\text{C-OPP}), \lambda_2'(\text{C-OPP}))$ strictly lies above L. This can be easily checked, as the value of λ_2 at L when $\lambda_1 = \lambda_1'(\text{C-OPP})$ is

$$\lambda_2'(L) = \frac{q_{2|2}^{(d)} q_{2|1,2}^{(d)}(q_{1|1}^{(d)} + q_{1|1}^{(2)}(1 - q_{1|1}^{(d)}) - q_{1|1,2}^{(d)})}{q_{1|1}^{(2)}(1 - q_{1|1}^{(d)})(q_{1|1,2}^{(d)} + q_{2|1,2}^{(d)}) + q_{1|1}^{(d)} q_{2|1,2}^{(d)}} \tag{3.48}$$

This is strictly less than $\lambda_2'(\text{C-OPP})$ if and only if $\eta > 0$. Therefore, when $\eta > 0$, the stable throughput region $\Re_{\text{C-OPP}}$ is strictly convex.

Then we compare $\Re_{\text{C-OPP}}$ and $\Re_{\text{CC}}$. If we can show that L strictly outer-bounds $\Re_{\text{CC}}$, it suffices to say that $\Re_{\text{C-OPP}}$ strictly outer-bounds $\Re_{\text{CC}}$ since the boundary of $\Re_{\text{C-OPP}}$ strictly lies above L as shown above. Because $\Re_{\text{CC}}$ is bounded by a straight line, to compare L and $\Re_{\text{CC}}$, it is enough to compare the intersection points of these lines with the axes. The two intersection points of L are given by Eq. (3.46) and Eq. (3.47); and from Eq. (3.29), the corresponding values for $\Re_{\text{CC}}$ are given by

$$\lambda_1^*(\text{CC}) = \frac{q_{2|2}^{(d)}\left(q_{1|1}^{(2)} + q_{1|1}^{(d)} - q_{1|1}^{(2)}q_{1|1}^{(d)}\right)}{q_{1|1}^{(2)} + q_{2|2}^{(d)} - q_{1|1}^{(2)}q_{1|1}^{(d)}}, \qquad \lambda_2^*(\text{CC}) = q_{2|2}^{(d)} \tag{3.49}$$

Clearly, $\lambda_2^*(\Re_1) = \lambda_2^*(\text{CC})$, and it can be shown that $\lambda_1^*(\Re_2) > \lambda_1^*(\text{CC})$ if and only if $\eta > 0$. Hence, in this case, we conclude that $\Re_{\text{C-OPP}}$ strictly contains $\Re_{\text{CC}}$, that is, $\Re_{\text{C-OPP}} \supset \Re_{\text{CC}}$.

On the other hand, when the condition in Eq. (3.33) is satisfied, the MPR channel is not strong enough to support simultaneous transmissions, and the optimal opportunistic scheduling becomes the conflict-free scheduling, which is the CC scheme. So even in the worst case, the C-OPP scheme can do as well as the CC scheme.

B) Effects that affect η: The value of η determines whether the channel supports simultaneous transmissions to some degree so that the opportunistic cooperation scheme can outperform the conventional scheme: if $\eta > 0$, simultaneous transmissions are supported; if $\eta \leq 0$, the opportunistic scheme reduces to the conventional one by scheduling S_1 and S_2 separately. A simple examination of η reveals how it is affected by those reception probabilities defined in Eq. (3.25):

- The effect of $q_{1|1}^{(d)}$, $q_{2|2}^{(d)}$: we observe that η decreases as $q_{1|1}^{(d)}$ or $q_{2|2}^{(d)}$ increases. A higher value of $q_{1|1}^{(d)}$ means that S_1 is more likely to deliver its packets to the destination without interference from S_2, so it is more preferable to schedule S_2 not to interfere with S_1; likewise, a higher value of $q_{2|2}^{(d)}$ implies a good channel from S_2 to d, and intuitively, we would like to exploit more relaying opportunity by scheduling S_2 to be silent when S_1 transmits, so that S_2 can possibly decode the transmitted packet of S_1 and then relay the packet.

- The effect of $q_{1|1,2}^{(d)}$ and $q_{2|1,2}^{(d)}$: η is an increasing function of both $q_{1|1,2}^{(d)}$ and $q_{2|1,2}^{(d)}$, this is not surprising because the action of simultaneous transmissions is more preferable by a strong MPR channel.

- The effect of $q_{1|1}^{(2)}$: η can be rearranged to be written as $q_{1|1}^{(2)}(1-q_{1|1}^{(d)})(q_{1|1,2}^{(d)}+q_{2|1,2}^{(d)}-q_{2|2}^{(d)})+q_{1|1}^{(d)}q_{2|1,2}^{(d)}-q_{1|1}^{(d)}q_{2|2}^{(d)}+q_{1|1,2}^{(d)}q_{2|2}^{(d)}$. Setting $q_{1|1}^{(2)}$ to be zero gives the value of the non-cooperative case with MPR channel. Hence, if $q_{1|1,2}^{(d)}+q_{2|1,2}^{(d)}-q_{2|2}^{(d)}$ is strictly greater than zero (which implies a strong MPR channel), cooperation adds a positive contribution to η which makes η more likely to be positive, and, hence, the opportunistic scheduling can take effect to outperform the conventional scheme.

3.3.3 Numerical Results

In this section, we compare the stable throughput regions of the two cooperation schemes (CC and C-OPP) and the non-cooperation scheme (NC). Two plots are given to illustrate how different categories of MPR channel affect the C-OPP

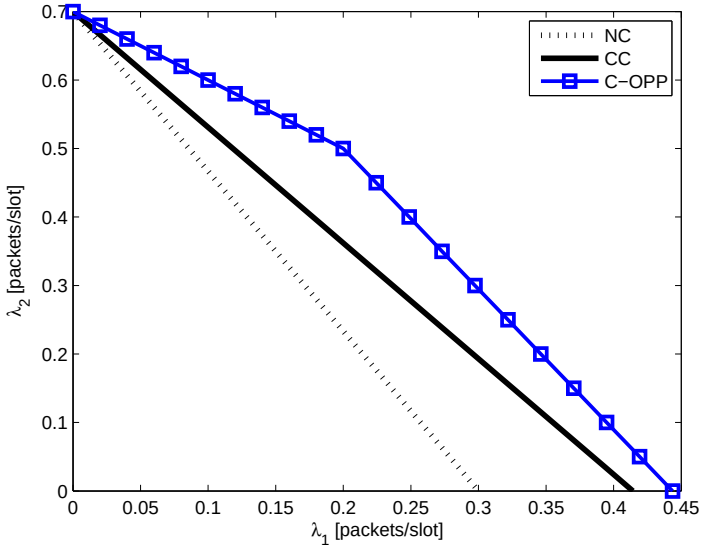

Figure 3.5: Comparison of the stable throughput regions under NC, CC, and C-OPP schemes when $\eta > 0$. $\Re_{\text{C-OPP}} \supset \Re_{\text{CC}} \supset \Re_{\text{NC}}$.

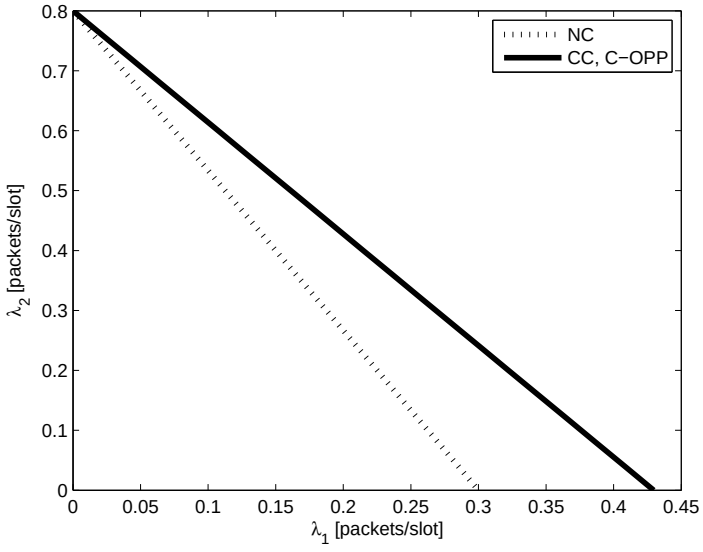

Figure 3.6: Comparison of the stable throughput regions under NC, CC, and C-OPP schemes when $\eta \leq 0$. $\Re_{\text{C-OPP}} = \Re_{\text{CC}} \supset \Re_{\text{NC}}$.

Table 3.1: The effects of $q_{1|1}^{(d)}$, $q_{2|2}^{(d)}$, $q_{1|1}^{(2)}$, $q_{1|1,2}^{(d)}$ and $q_{2|1,2}^{(d)}$ on the value of η.

$q_{1\|1}^{(d)}$	$q_{2\|2}^{(d)}$	$q_{1\|1}^{(2)}$	$q_{1\|1,2}^{(d)}$	$q_{2\|1,2}^{(d)}$	η
0.4	0.8	0.4	0.1	0.4	-0.1520
			0.2	0.4	-0.0480
			0.2	0.5	0.0160
			0.3	0.6	0.1840
0.3	0.8	0.4	0.2	0.5	0.0420
0.4					0.0160
0.5					-0.0100
0.4	0.7	0.4	0.2	0.5	0.0600
	0.8				0.0160
	0.9				-0.0280

scheme. In Fig. 3.5, the reception probabilities are chosen to be $q_{1|1}^{(d)} = 0.3$, $q_{2|2}^{(d)} = 0.7$, $q_{1|1}^{(2)} = 0.4$, $q_{1|1,2}^{(d)} = 0.2$ and $q_{2|1,2}^{(d)} = 0.5$ such that $\eta > 0$. As we have proved, the C-OPP scheme supports simultaneous transmissions to some degree, and hence, supports a convex stable throughput region which strictly contains the corresponding region of the CC scheme, which is bounded by a straight line. In the subregion defined by $\lambda_1 \leq q_{1|1,2}^{(d)}$, the optimal scheduling probability p^* is 1; while in the subregion defined by $\lambda_1 > q_{1|1,2}^{(d)}$, the optimal p^* is $\frac{q_{1|1}^{(2)}+q_{1|1}^{(d)}-q_{1|1}^{(2)}q_{1|1}^{(d)}-\lambda_1}{q_{1|1}^{(2)}+q_{1|1}^{(d)}-q_{1|1}^{(2)}q_{1|1}^{(d)}-q_{1|1,2}^{(d)}}$. The convexity of the region also implies that higher sum rates can be achieved.

In Fig. 3.6, the reception probabilities are chosen to be $q_{1|1}^{(d)} = 0.3$, $q_{2|2}^{(d)} = 0.8$, $q_{1|1}^{(2)} = 0.4$, $q_{1|1,2}^{(d)} = 0.2$ and $q_{2|1,2}^{(d)} = 0.4$ such that $\eta \leq 0$. It is seen that the optimal scheduling strategy sets $p^* = 0$, which reduces the C-OPP scheme to the CC scheme by scheduling S_2 to transmit only when S_1 is empty. So even in the worst case, the C-OPP scheme can perform as well as the CC scheme.

Under all channel conditions, both cooperation schemes outperform the non-

cooperation scheme; so cooperation among users improves the stable throughput region, thus leads to higher stable throughput simultaneously for both users.

In addition, a number of channel reception probability sets are tested to demonstrate their effects on affecting the value of η. The results are listed in Table 3.1.

3.4 Discussion

In this chapter, we utilized enhancement techniques that are based on the physical layer to boost network-layer cooperation. Cooperative communication studies at the physical layer usually neglect source burstiness, and on the other hand, traditional work at the network layer makes unrealistic and limited assumptions on the physical layer channel model and the encoding/decoding processes. The proposed cooperation methods in this chapter combine physical and network layer ideas.

In Section 3.2, the DDF cooperative technique which allows relaying assistance also during the source's transmission was first studied. Then an adaptive superposition coding technique was investigated by allowing the relay to simultaneously serve two relaying queues, according to the instantaneous channel feedback. We showed that both advanced cooperative techniques result in more relaying opportunities, which leads to higher performance gains in terms of the stable throughput region.

In Section 3.3, by considering a general asymmetric MPR channel model which better exploits the physical-layer properties, the conventional cognitive cooperation where the relay only transmits during an idle time slot was refined. As such, we

proposed an opportunistic cooperation scheme, in which the relay transmits together with the other user with some probability. The stable throughput region of this opportunistic scheme was characterized explicitly by optimizing the transmission probability at the relay. The conventional cooperation scheme results in higher stable throughput for both users over the non-cooperation scheme, and it was shown that the opportunistic cooperation scheme can yield higher performance gains under certain channels.

Chapter 4

Stability and Throughput Regions for Cooperative Multi-Access

4.1 Introduction

Previous chapters addressed the issue of user cooperation in wireless networks. For the cooperative multi-access system investigated in Chapter 2, if we set the number of source users to be $N = 2$, it becomes a two-user system where the user "closer" to the destination helps to relay packets from the other user. This model resembles the single-relay channel which was first introduced by van der Meulen [41]; later, Cover and El Gamal considered the Shannon information capacity region of the classical relay channel in [42], and were able to determine it for the class of physically degraded channels. Subsequently, numerous papers have contributed to the understanding of the relay channel, including [43, 44, 45], et al. These work was studied from an information-theoretic standpoint; even for the single-relay channel, the Shannon information capacity remains unknown.

In this chapter, we will continue to perform cooperation between the two users at the network level, and assume that both users have their own data packets to be delivered to the destination. The main objective is to revisit the relationship between the *stability region* and the *throughput region* in the context of a packet-based network-level cooperative multi-access system. This would answer the unresolved question of whether the empty state of the queue is immaterial in determining

stability. The stability region (also called the stable throughput region) and the throughput region (also called the saturated throughput region), are two rate measures based on the networking perspective. With different assumptions and different meanings, these two regions need not coincide. In the multi-access system where cooperation is not performed, it was established that these two regions are identical under a variety of channel and traffic models [17, 18, 19, 20, 21]. In the case of our cooperative multi-access system, there are different packet streams at the relay, thus, we apply a class of scheduling policies for the transmission of the source packets and relayed packets at the relay node. We show that, the stability region is independent of the scheduling policies; however, the throughput region does depend on the scheduling policies, and is not necessarily identical to the stability region. This result rekindles interest in the relation between these two regions. In addition, we determine the optimal policy for maximizing the throughput region.

The abovementioned two regions are characterized and compared under both a centralized scheduled access and a random access scheme separately. Another outcome of our work is that, the cooperative stability region under random access does not necessarily outer-bound the non-cooperative stability region. In the last part of this chapter, network coding is performed at the relay node as an alternative to plain store-and-forward routing, and its impact on both the stability and throughput regions is evaluated. We conclude that network coding between the two traffic streams at the relay node leads to the same performance as plain store-and-forward routing.

A simpler version of our model was studied in [46], where the authors con-

sidered a linear tandem network in which each node was only connected to its two neighboring nodes via error-free links. However, this simple connectivity model does not accurately capture the physical-layer property of wireless channels. In [46], the authors did not consider the stability issue under random access; when they analyzed the throughput region, they assumed that all queues in the network are saturated which guarantees permanent availability of packets for transmission. However, even under heavy traffic, the queues of packets to be relayed are built by arrivals from the neighboring nodes and, hence, can be empty with positive probability. In this chapter we characterize the throughput region by assuming that *only* the source queues at the user nodes are saturated, while the relay queue at the relay accepts bursty arrivals departing from the other user and, hence, need not be saturated. We show that by assuming that all queues are saturated, the actual throughput region is over-estimated.

4.2 Model

We consider the slotted two-user multi-access system shown in Fig. 4.1. Two source users, S and R, transmit unicast traffic to the common destination node D. The source nodes are equipped with buffers of infinite size to store incoming packets, and the transmission of each packet takes exactly one time slot. Nodes are ordered according to Fig. 4.1 so that R has a better source-destination channel than S. We adopt the single-packet reception channel model as described in Chapter 1, such that a transmitted packet will be successfully decoded with a certain probability.

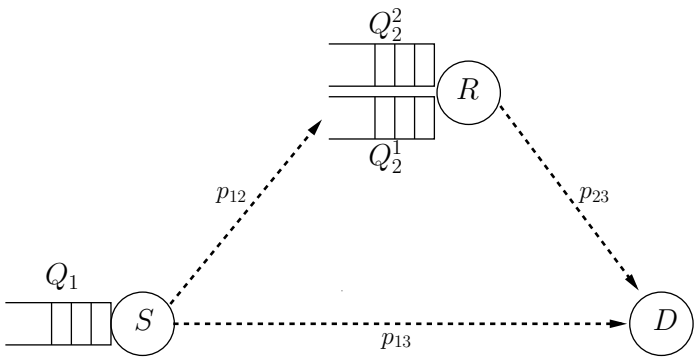

Figure 4.1: The two-user multi-access system with cooperation between the two source users (S and R). D is the common destination. Channel reception probabilities are denoted above the channels.

Further, if both users transmit at the same time, a collision will occur which leads to the failure of both transmissions. For simplicity, we do not consider advanced physical-layer techniques as in Chapter 3; however, we will show that the considered simple channel model can still lend itself to a thorough examination and achieve our objective.

There are three channel reception probabilities associated with this multi-access system. Specifically, p_{13} and p_{23} denote the source-destination channel reception probabilities from the sources S and R to the destination respectively; and p_{12} denotes the inter-user channel reception probability between S and R. By using these notations, the fact that source user R has a better channel to the destination (than S) can be mathematically expressed as $p_{23} > p_{13}$. The reception probabilities p_{13}, p_{23} and p_{12} are determined by parameters such as the network topology, transmission power, targeted bit-error rate, etc., and cannot be altered once these parameters are fixed.

The cooperation strategy used in this chapter is stated as follows: when node

S transmits a packet, if the destination D decodes the packet successfully, it sends back an acknowledgement (ACK) and the packet exits the system; otherwise, if D doesn't decode the packet but R decodes the packet, R sends back an ACK and keeps the packet in its queue for retransmission; upon receiving the ACK from R, S will drop the packet in this case; if neither R nor D decodes the packet, the packet will remain in S's queue for retransmission. For R's own packets, R is responsible for delivering them to the destination without help from S. As usual, ACKs are assumed to be error-free and instantaneous, and broadcast to the whole network over a separate control channel with negligible bandwidth.

With this form of cooperation, there are two queues at node R, as depicted in Fig. 4.1: the queue Q_2^1 stores packets relayed from S, and Q_2^2 stores R's own packets. We refer to Q_1 and Q_2^2 as the source queues, and Q_2^1 as the relay queue. We will focus our attention on investigating the performance metrics of the stability region and throughput region. Accordingly, we model bursty packet arrivals and saturated source queues separately. In the non-saturated system, packets arrive at nodes S and R independently according to Bernoulli processes, with rates λ_1 and λ_2 respectively. Then, the stability region is defined as the union of all (λ_1, λ_2) such that all queues in the network remain stable. In the saturated system, we assume permanent availability of source packets at the two source queues Q_1 and Q_2^2; accordingly, we study the throughput region which is the union of all achievable throughput rate pairs (λ_1, λ_2), where λ_1 and λ_2 stand for the throughput rates for S and R respectively. We should note here that the relay queue Q_2^1 at node R is formed by arrivals from S according to a stationary process, and hence, cannot be

assumed to be saturated.

At the MAC layer, we separately consider a centralized scheme and a fully distributed scheme, namely, scheduled access and random access. The two medium access schemes will be described and analyzed in Section 4.3 and Section 4.4 respectively.

4.3 Scheduled Access

Under scheduled access, the source nodes S and R access the channel at disjoint fractions of time ω_1 and ω_2 respectively, where $0 \leq \omega_1 \leq 1$, $0 \leq \omega_2 \leq 1$. The set of all feasible time allocation satisfies: $\sum_{i=1}^{2} \omega_i \leq 1$. The two source nodes are activated to transmit at different time slots, so there is no contention. For each allocation (ω_1, ω_2), we can find a corresponding stability region and throughput region; then, we will take the union of such regions over all possible (ω_1, ω_2) such that $\sum_{i=1}^{2} \omega_i \leq 1$. Therefore, for any point (λ_1, λ_2) inside the thus obtained stability region, there exist time allocations (ω_1, ω_2) such that the queues in the network remain stable; similarly, for any point (λ_1, λ_2) in the throughput region, there exist time allocations (ω_1, ω_2) such that the throughput rates (λ_1, λ_2) can be achieved.

In this cooperative multi-access system, there are two queues Q_2^1 and Q_2^2 at node R, one stores packets relayed from S, and the other stores R's own packets. Therefore, when it is R's turn to transmit a packet, R can choose to transmit a packet from one of the two queues; the scheduling policy, that is, the policy for deciding which of the two queues to serve, may affect the (stable) throughput achieved by the

two source nodes individually. We consider a class of scheduling policies for node R in deciding between Q_2^1 and Q_2^2: when both Q_2^1 and Q_2^2 are non-empty, R selects a packet from Q_2^1 with probability β, and a packet from Q_2^2 with probability $1-\beta$ (for $0 \leq \beta \leq 1$); when only one of the two queues is non-empty, R selects a packet from that queue. If we set $\beta = 0$, it corresponds to the case which assigns higher priority to Q_2^2; if we set $\beta = 1$, it corresponds to the case which assigns higher priority to Q_2^1. Putting together these two special cases with the general case, the three scheduling policies we consider at the relay node are described as follows:

1. The source queue Q_2^2 has higher priority; that is, only when it is empty, will R transmit a packet from the relay queue Q_2^1 ($\beta = 0$).

2. The relay queue Q_2^1 has higher priority; that is, only when it is empty, will R serve a packet from the source queue Q_2^2 ($\beta = 1$).

3. When both queues Q_2^1 and Q_2^2 are non-empty, R selects a packet from Q_2^1 with probability β, and a packet from Q_2^2 with probability $1 - \beta$, for $0 \leq \beta \leq 1$; when only one of the two queues is non-empty, R selects a packet from that queue.

Note that under policy 3, we will vary over $\beta \in [0, 1]$ to obtain the stability and throughput regions; the thus obtained regions are compared with the corresponding regions under policies 1 and 2.

4.3.1 Stability Analysis

The stability region under scheduled access for a tandem of N users was obtained in Chapter 2; an important observation from the proof in that analysis is that the resulting region is independent of the scheduling policies at the relay node. In addition, the stability region of the cooperative system strictly contains the stability region of the non-cooperative system. By specializing the results from Chapter 2, for $N = 2$, we obtain:

Theorem 4.1 *For any scheduling policy adopted by node R, the stability region under scheduled access in the two-user multi-access system with cooperation is given by*

$$L_{SR} = \left\{ (\lambda_1, \lambda_2) : \frac{(p_{12} + p_{23} - p_{12}p_{13})\lambda_1}{p_{23}\left(p_{12} + p_{13} - p_{12}p_{13}\right)} + \frac{\lambda_2}{p_{23}} < 1 \right\} \tag{4.1}$$

4.3.2 Throughput Analysis

Next, we consider the throughput region, that is, the region of throughput rates that are achievable when we assume that the two source queues Q_1 and Q_2^2 are saturated. We denote by L_i the throughput region when the scheduling policy i ($i \in \{1, 2, 3\}$) is adopted. Note that even in the saturated case, the relay queue Q_2^1 at node R is built by the arrivals from Q_1, so it is non-saturated and can be empty at times.

Under scheduling policy 1 in which Q_2^2 has higher priority, the assumption that Q_2^2 is always saturated implies that the relay queue Q_2^1 will never be served by

node R. Therefore, we have the following theorem:

Theorem 4.2 *Under scheduling policy 1, the throughput region reduces to the throughput region of the non-cooperative system, which is*

$$L_1 = \left\{ (\lambda_1, \lambda_2) : \frac{\lambda_1}{p_{13}} + \frac{\lambda_2}{p_{23}} < 1 \right\} \tag{4.2}$$

This is a strict subset of the stability region given by Eq. (4.1), that is, $L_1 \subset L_{SR}$.

Proof: The source queue Q_2^2 is saturated and never empties, so R always selects a packet from Q_2^2 to transmit when it is R's slot. This reduces to the non-cooperative system as node R never serves the packets relayed from S. The throughput rates are given by

$$\lambda_1 = \omega_1 p_{13}, \qquad \lambda_2 = \omega_2 p_{23} \tag{4.3}$$

If we take the union over all (ω_1, ω_2) such that $\omega_1 + \omega_2 \leq 1$, we obtain the throughput region as given in Eq. (4.2). A simple comparison will find that the throughput region under policy 1 is strictly inferior to the stability region given by Eq. (4.1). ■

Similarly, as Q_2^2 never empties, the scheduling policy 3 becomes: when Q_2^1 is non-empty, R selects a packet from Q_2^1 with probability β, and a packet from Q_2^2 with probability $1 - \beta$, for $0 \leq \beta \leq 1$; when Q_2^1 is empty, R selects a packet from Q_2^2. The throughput analysis under policy 2 and policy 3 is carried out, and the results lead to:

Theorem 4.3 *Under scheduling policies 2 and 3, the throughput regions are the same and are* identical *to the stability region given by Eq. (4.1) of Theorem 4.1. So we have:* $L_2 = L_3 = L_{SR}$.

Proof: Under both policies 2 and 3, the throughput rate for source S consists of two parts: the packets delivered to the destination by S itself, and those relayed by R. The throughput rate contributed by S itself is easily seen to be $\lambda_1^S = \omega_1 p_{13}$, since a packet is delivered from S to the destination if and only if the time slot is assigned to S and the packet is successfully decoded by the destination, which has the probability of $\omega_1 p_{13}$; we need further to calculate the throughput rate contributed by the relaying, that is, the throughput rate contributed by the packet delivery from the queue Q_2^1.

As we have clarified before, Q_2^1 is built by arrivals from source S, and hence, can be empty with positive probability. Therefore, the behavior of Q_2^1 is determined by both its arrival process and its service process. A packet will arrive at Q_2^1 from node S if and only if the following three events happen together: first, the time slot is assigned to S, so S transmits a packet; second, the destination doesn't decode that packet; and third, the relay node R decodes that packet successfully. These three events are independent, so the expected value of the arrival process to the queue Q_2^1 is

$$\lambda_{Q_2^1} = \omega_1 p_{12}(1 - p_{13}) \tag{4.4}$$

Under policy 2, Q_2^1 has higher priority than Q_2^2, and the packet is delivered to

the destination from Q_2^1 if and only if it is node R's time slot, and the destination successfully decodes the transmitted packet. So the average service rate of Q_2^1 is calculated to be

$$\mu_{Q_2^1} = \omega_2 p_{23} \tag{4.5}$$

By Loynes' Theorem, if the arrival and service processes of a queue are jointly stationary, the queue is stable if and only if the average arrival rate is strictly less than the average service rate, and furthermore, the departure rate is equal to the arrival rate. As a result, one and only one, of the following two cases will happen:

1. If $\omega_1 p_{12}(1 - p_{13}) < \omega_2 p_{23}$, the queue Q_2^1 is stable, and the throughput rate for S contributed by the relaying is $\lambda_1^R = \omega_1 p_{12}(1 - p_{13})$. Therefore, the total throughput rate for S is

$$\lambda_1 = \lambda_1^S + \lambda_1^R = \omega_1(p_{13} + p_{12}(1 - p_{13})) \tag{4.6}$$

In this case, Q_2^1 is empty with probability $1-\omega_1 p_{12}(1-p_{13})/\omega_2 p_{23}$ as it is simply a discrete-time M/M/1 queue. According to policy 2, Q_2^2 is successfully served if it is node R's time slot and Q_2^1 is empty, and the packet transmitted from Q_2^2 is decoded by the destination. So the average service rate that Q_2^2 receives is given by

$$\mu_{Q_2^2} = \omega_2 p_{23}\left(1 - \frac{\omega_1 p_{12}(1 - p_{13})}{\omega_2 p_{23}}\right) = \omega_2 p_{23} - \omega_1 p_{12}(1 - p_{13}) \tag{4.7}$$

This is the actual throughput rate for user R. So we obtain the throughput rates as

$$\lambda_1 = \omega_1(p_{13} + p_{12}(1 - p_{13})) \tag{4.8}$$

$$\lambda_2 = \omega_2 p_{23} - \omega_1 p_{12}(1 - p_{13}) \tag{4.9}$$

2. If $\omega_1 p_{12}(1 - p_{13}) \geq \omega_2 p_{23}$, Q_2^1 is unstable and will grow without bound. In this case, the throughput rate for S contributed by the relaying is equal to the service rate of Q_2^1, which is $\omega_2 p_{23}$; besides, Q_2^1 is empty with probability zero, so Q_2^2 is served with probability zero as it is assigned lower priority, and the throughput rate for R is zero. Therefore, the throughput rates are given by

$$\lambda_1 = \omega_1 p_{13} + \omega_2 p_{23} \tag{4.10}$$

$$\lambda_2 = 0 \tag{4.11}$$

By taking the union of these rates over all feasible (ω_1, ω_2) such that $\sum_{i=1}^{2} \omega_i \leq 1$, we obtain that the actual throughput region under policy 2 is characterized by Eq. (4.1).

Under policy 3, given fixed (ω_1, ω_2), a similar analysis will lead to the throughput rates as functions of ω_1, ω_2 and β, given by the following:

1. If $\omega_1 p_{12}(1-p_{13}) < \beta\omega_2 p_{23}$, the throughput rates are given by

$$\lambda_1 = \omega_1(p_{13} + p_{12}(1-p_{13})) \tag{4.12}$$

$$\lambda_2 = \omega_2 p_{23} - \omega_1 p_{12}(1-p_{13}) \tag{4.13}$$

2. If $\omega_1 p_{12}(1-p_{13}) \geq \beta\omega_2 p_{23}$, the throughput rates are given by

$$\lambda_1 = \omega_1 p_{13} + \beta\omega_2 p_{23} \tag{4.14}$$

$$\lambda_2 = (1-\beta)\omega_2 p_{23} \tag{4.15}$$

By taking the union of the throughput rates over all (ω_1, ω_2) such that $\sum_{i=1}^{2} \omega_i \leq 1$ and $\beta \in [0,1]$, the throughput region under policy 3 is also characterized by Eq. (4.1). ■

The above analysis reveals that the arrival and service processes of the non-saturated queue Q_2^1 play a critical role in determining the throughput region under policy 2 and policy 3. If we instead assume that all queues are saturated including the relay queue Q_2^1, as was done in [46], the maximum throughput region is achieved by activating node R all the time, that is, by assigning $\omega_1 = 0$ and $\omega_2 = 1$. Node R switches between Q_2^1 and Q_2^2, and the throughput region is given by $L' = \big\{ (\lambda_1, \lambda_2) : \lambda_1 + \lambda_2 < p_{23} \big\}$. This strictly contains the region expressed in Eq. (4.1). So by assuming all queues saturated, we will over-estimate the throughput region, and the corresponding result provided in [46] turns out to be a strict outer bound.

4.3.3 Optimal Policy for the Throughput Region

As shown above, the throughput region under policy 3 obtained by taking the union of the throughput rates for all possible $\beta \in [0, 1]$, where β is the probability to select a packet from the relay queue Q_2^1, is the same as the throughput region under policy 2 where Q_2^1 has higher priority than Q_2^2. In addition, the throughput regions under policy 2 and policy 3 strictly outer-bound the throughput region under policy 1. These two observations suffice to conclude that, the scheduling policy which assigns higher priority to the relay queue Q_2^1 ($\beta = 1$) is optimal in the sense that it achieves the maximum throughput region among the class of stationary scheduling policies considered at the relay node.

4.4 Random Access

Due to the difficulty of global coordination for scheduled access in ad hoc networks, random access is an attractive and simple alternative that is implemented in a fully distributed fashion. In this setting, when source node S (or R) is backlogged, it transmits a packet with probability q_1 (or q_2) independently of any other event. If both S and R decide to transmit in the same time slot, a collision will occur and both transmissions will fail. We also do not allow simultaneous transmission and reception, and hence, when R is transmitting, it cannot receive a packet transmitted by S.

We are interested in characterizing the stability and throughput regions corresponding to each of the three scheduling policies described in Section 4.3, upon

which we can reveal the relation between the two regions for random access. Likewise, the stability region is obtained by taking the union of all arrival rates (λ_1, λ_2) such that there exist transmission probabilities (q_1, q_2) which stabilize all queues; the throughput region is defined as the union of all throughput rates (λ_1, λ_2) for which there exist (q_1, q_2) that can achieve them.

4.4.1 Stability Analysis

Stability analysis in random access system is known to be notoriously difficult due to the interaction among the queues. So far, the stability region of the slotted random access system is known for no more than three users. Even in the case of two users without cooperation (the two-user ALOHA system), the stability analysis is non-trivial. We again make use of the stochastic dominance approach to decouple the queues and derive the stability region. The main result is summarized in the following theorem:

Theorem 4.4 *The stability region under random access is independent of the scheduling policies at the relay node R, and is characterized by*

$$\Re_{SR} = \left\{ (\lambda_1, \lambda_2) : \sqrt{\frac{\lambda_1}{p_{12} + p_{13} - p_{12}p_{13}}} + \sqrt{\frac{p_{12}(1-p_{13})\lambda_1}{p_{23}(p_{12} + p_{13} - p_{12}p_{13})} + \frac{\lambda_2}{p_{23}}} < 1 \right\} \quad (4.16)$$

Proof: See Appendix 4.8.1. ■

Under scheduled access, we have proved in Chapter 2 that the cooperative stability region strictly contains the stability region when cooperation is not used.

Under random access, the stability region without cooperation is known to be

$$\Re_{\mathrm{SR}}^{\mathrm{nocoop}} = \left\{ (\lambda_1, \lambda_2) : \sqrt{\frac{\lambda_1}{p_{13}}} + \sqrt{\frac{\lambda_2}{p_{23}}} < 1 \right\} \tag{4.17}$$

From Eq. (4.16), the maximum stabilizable rate for source S with cooperation is $\frac{p_{23}(p_{12}+p_{13}-p_{12}p_{13})}{(\sqrt{p_{12}(1-p_{13})}+\sqrt{p_{23}})^2}$, this quantity can be greater than, equal to or less than p_{13}, which is the maximum stabilizable rate for S in the non-cooperative system. By comparing the two regions expressed in Eq. (4.16) and Eq. (4.17), we can easily find that the relation between the two regions depends on the relation between these two quantities that are functions of the system parameters. Specifically,

1. If

$$\frac{p_{23}(p_{12} + p_{13} - p_{12}p_{13})}{(\sqrt{p_{12}(1 - p_{13})} + \sqrt{p_{23}})^2} \geq p_{13} \tag{4.18}$$

   the stability region of the cooperative system under random access strictly contains the stability region of the non-cooperative system. The relation between these two regions in this case is illustrated in Fig. 4.2(a);

2. If

$$\frac{p_{23}(p_{12} + p_{13} - p_{12}p_{13})}{(\sqrt{p_{12}(1 - p_{13})} + \sqrt{p_{23}})^2} < p_{13} \tag{4.19}$$

   the stability region of the cooperative system and the stability region of the non-cooperative system overlap, but neither one properly contains the other.

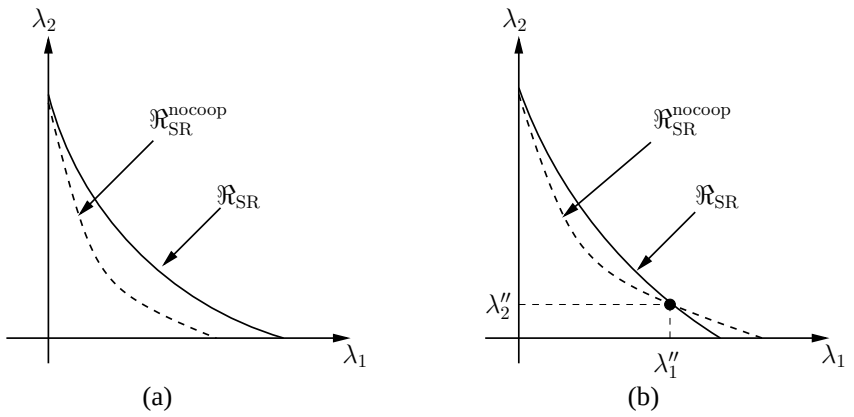

Figure 4.2: Comparison of the stability regions for the cooperative and non-cooperative systems under random access. (a) Condition (4.18) holds, $\Re_{\rm SR} \supset \Re_{\rm SR}^{\rm nocoop}$. (b) Condition (4.19) holds, $\Re_{\rm SR}$ and $\Re_{\rm SR}^{\rm nocoop}$ cross at $(\lambda_1'', \lambda_2'')$.

The boundaries of the two stability regions have two intersection points, which are $(\lambda_1', \lambda_2') = (0, p_{23})$ on the y-axis, and $(\lambda_1'', \lambda_2'') = (\alpha p_{13}, (1-\sqrt{\alpha})^2 p_{23})$, with $\alpha = \frac{4\left(\sqrt{p_{12}-p_{12}p_{13}+p_{13}}-\sqrt{p_{13}}\right)^2 (p_{12}-p_{12}p_{13}+p_{13})p_{23}^2}{(p_{12}-p_{12}p_{13})^2(p_{13}+p_{23})^2}$. In the subset constrained by $\lambda_1 \leq \lambda_1''$ (or equivalently, $\lambda_2 \geq \lambda_2''$), the cooperative stability region strictly contains the non-cooperative stability region; in the subset constrained by $\lambda_1 > \lambda_1''$ (or equivalently, $\lambda_2 < \lambda_2''$), the non-cooperative stability region strictly contains the cooperative stability region. This relationship is illustrated in Fig. 4.2(b).

Condition (4.18) implies that p_{23} is considerably larger than p_{13}, that is, the channel condition from R to D is sufficiently better than the channel condition from S to D. By contrast, condition (4.19) implies that the R to D channel quality is better than the S to D one, but not sufficiently so. When cooperation is used, some of S's packets will enter R's queue for retransmission, and the transmission of these packets from R will collide with the transmission from S, if S decides to transmit at the same time. Therefore, we observe that two forces are at work:

(a) For one thing, some of S's packets will be relayed by R, which has the better user-destination channel. This is a positive effect.

(b) For another, the addition of such an arrival process to R requires R to push more packets out of its queue if it wants to remain stable; and under random access, the simultaneous transmissions of R and S will cause destructive collision and both transmissions will fail. This is a negative effect.

Based on these two effects, under condition (4.19) when the R to D channel is not sufficiently better than the S to D channel, it follows that: when λ_1 is very small, the cooperative system outperforms the non-cooperative system in terms of stability region. In this range, effect (a) plays the major role. As we increase λ_1 until it exceeds a certain threshold value λ_1'', the cooperative stability region falls inside the non-cooperative stability region. This is because, as λ_1 increases, the arrival rate from S to R increases as well, and the occurrence of collision increases. As a result, effect (b) becomes dominant and the cooperation strategy actually degrades the performance.

4.4.2 Throughput Analysis

Under the assumption that the source queues Q_1 and Q_2^2 are saturated, we investigate the throughput regions with respect to each of the three scheduling policies at node R as described in Section 4.3. We denote by $\Re_i$ the throughput region when the scheduling policy i ($i \in \{1, 2, 3\}$) is adopted.

Theorem 4.5 *Under scheduling policy 1, the throughput region is given by*

$$\Re_1 = \left\{ (\lambda_1, \lambda_2) : \sqrt{\frac{\lambda_1}{p_{13}}} + \sqrt{\frac{\lambda_2}{p_{23}}} < 1 \right\} \tag{4.20}$$

Proof: As we have argued in the proof of Theorem 4.2 in Section 4.3.2, the saturated system in which the source traffic at the relay node has higher priority reduces to the non-cooperative system. Under random access, this system behaves the same as the slotted ALOHA system with an erasure channel, and the throughput region is known to be given by Eq. (4.20), which is also the non-cooperative stability region. ■

Theorem 4.6 *Under scheduling policy 2 and policy 3, the throughput regions* $\Re_2$ *and* $\Re_3$ *are given by:*

1. *When the condition in Eq. (4.18) is satisfied, the throughput regions under policy 2 and policy 3 are the same, and are identical to the cooperative stability region as given by Eq. (4.16) in Theorem 4.4. That is,*

$$\Re_2 = \Re_3 = \Re_{SR} \tag{4.21}$$

2. *When the condition in Eq. (4.19) is satisfied, the throughput region under policy 2 is given by the union of two regions, specifically,*

$$\Re_2 = \Re_{SR} \bigcup \Re' \tag{4.22}$$

where $\Re_{SR}$ is the cooperative stability region as shown in Eq. (4.16), and $\Re'$ is defined by

$$\Re' = \left\{ (\lambda_1, \lambda_2) : \begin{array}{l} \frac{p_{23}(p_{12}+p_{13}-p_{12}p_{13})}{(\sqrt{p_{12}(1-p_{13})}+\sqrt{p_{23}})^2} \leq \lambda_1 \leq p_{13} \\ \lambda_2 = 0 \end{array} \right\} \tag{4.23}$$

A diagram of the boundary of $\Re_2$ given by this equation is shown in Fig. 4.3(a). Finally, the throughput region under policy 3 is given by

$$\Re_3 = \Re_{SR} \bigcup \Re_1 \tag{4.24}$$

where $\Re_{SR}$ and $\Re_1$ are given in Eq. (4.16) and Eq. (4.20) respectively. A diagram showing the boundary of $\Re_3$ is depicted in Fig. 4.3(b). In this case, the relationship between the cooperative stability region and the throughput regions under policies 2 and 3 is

$$\Re_3 \supset \Re_2 \supset \Re_{SR} \tag{4.25}$$

Proof: The throughput rates for nodes S and R under policy 2 and policy 3 can be computed in the same way as in the proof of Theorem 4.3 in Section 4.3.2. For fixed (q_1, q_2), under policy 2, we can obtain that

1. If

$$q_1(1-q_2)p_{12}(1-p_{13}) < q_2(1-q_1)p_{23} \tag{4.26}$$

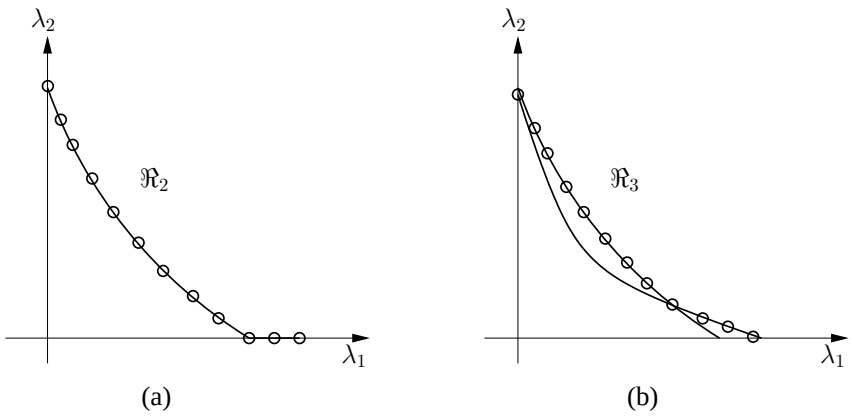

Figure 4.3: Boundaries of the throughput regions under policy 2 and policy 3 for random access when condition (4.19) holds. (a) $\Re_2$, the throughput region under policy 2. (b) $\Re_3$, the throughput region under policy 3.

the throughput rates are given by

$$\lambda_1 = q_1(1-q_2)(p_{12} - p_{12}p_{13} + p_{13}) \tag{4.27}$$

$$\lambda_2 = q_2(1-q_1)p_{23} - q_1(1-q_2)p_{12}(1-p_{13}) \tag{4.28}$$

2. If

$$q_1(1-q_2)p_{12}(1-p_{13}) \geq q_2(1-q_1)p_{23} \tag{4.29}$$

the throughput rates are given by

$$\lambda_1 = q_1(1-q_2)p_{13} + q_2(1-q_1)p_{23} \tag{4.30}$$

$$\lambda_2 = 0 \tag{4.31}$$

The throughput region is the union of all these rates over $(q_1, q_2) \in [0,1]^2$. When the condition written in Eq. (4.26) is satisfied, by following the same procedure as in the proof of Theorem 4.4 which solves a constrained optimization problem, we can obtain that the achievable throughput rates are in the region of

$$\left\{(\lambda_1,\lambda_2) : \sqrt{\frac{\lambda_1}{p_{12}+p_{13}-p_{12}p_{13}}} + \sqrt{\frac{p_{12}(1-p_{13})\lambda_1}{p_{23}(p_{12}+p_{13}-p_{12}p_{13})} + \frac{\lambda_2}{p_{23}}} < 1\right\} \quad (4.32)$$

On the other hand, when the condition in Eq. (4.29) holds, the achievable throughput rates are in the region of

$$\left\{(\lambda_1,\lambda_2) : \begin{array}{l} 0 \le \lambda_1 \le \max\left(p_{13}, \frac{p_{23}(p_{12}+p_{13}-p_{12}p_{13})}{\left(\sqrt{p_{12}(1-p_{13})}+\sqrt{p_{23}}\right)^2}\right) \\ \\ \lambda_2 = 0 \end{array}\right\} \quad (4.33)$$

The throughput region under policy 2 is then the union of these two sub-regions. Depending on the relation between the quantities p_{13} and $\frac{p_{23}(p_{12}+p_{13}-p_{12}p_{13})}{\left(\sqrt{p_{12}(1-p_{13})}+\sqrt{p_{23}}\right)^2}$, the throughput region is as described in Theorem 4.6.

Under policy 3, by following a similar analysis, we can obtain that for fixed (q_1, q_2) and fixed β, we have:

1. If $q_1(1-q_2)p_{12}(1-p_{13}) < \beta q_2(1-q_1)p_{23}$, the throughput rates are given by

$$\lambda_1 = q_1(1-q_2)(p_{12}-p_{12}p_{13}+p_{13}) \quad (4.34)$$

$$\lambda_2 = q_2(1-q_1)p_{23} - q_1(1-q_2)p_{12}(1-p_{13}) \quad (4.35)$$

2. If $q_1(1-q_2)p_{12}(1-p_{13}) \geq \beta q_2(1-q_1)p_{23}$, the throughput rates are given by

$$\lambda_1 = q_1(1-q_2)p_{13} + \beta q_2(1-q_1)p_{23} \tag{4.36}$$

$$\lambda_2 = (1-\beta)q_2(1-q_1)p_{23} \tag{4.37}$$

After taking the union over $(q_1, q_2) \in [0,1]^2$ and $\beta \in [0,1]$, the throughput region under policy 3 can be shown to be given by Theorem 4.6.

Finally, a simple comparison between $\Re_2$, $\Re_3$ and $\Re_{\text{SR}}$ yields the relations between these regions as shown in Eq. (4.21) and Eq. (4.25), under condition (4.18) and condition (4.19) respectively. ■

4.4.3 Optimal Policy for the Throughput Region

From theorems 4.5 and 4.6, we learn that if condition (4.18) is satisfied, the throughput region under policy 2 coincides with the throughput region under policy 3, which strictly outer-bounds the throughput region under policy 1. That is, $\Re_2 = \Re_3 \supset \Re_1$. Hence, in this case, the scheduling policy which always assigns higher priority to the relay queue Q_2^1 (corresponding to $\beta = 1$) is optimal in achieving the maximum throughput region among the stationary scheduling policies considered at the relay node. The optimal scheduling policy under condition (4.18) is shown in Fig. 4.4(a).

Otherwise, that is, if condition (4.19) is satisfied, the optimal scheduling policy is not the same for the entire range. In this case, a simple comparison will find that the maximum throughput region is given by the union of two regions: the throughput

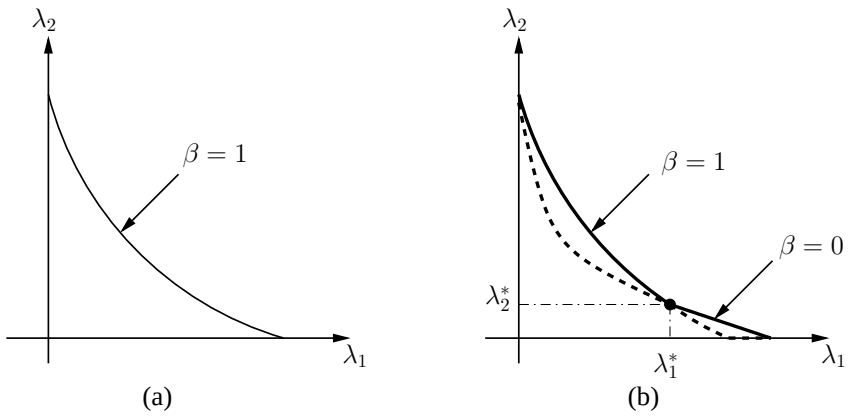

Figure 4.4: The optimal scheduling policies that maximize the throughput region for random access under (a) condition (4.18), and (b) condition (4.19).

region under policy 1 and the throughput region under policy 2. That is, $\Re_{\text{TR}}^{\max} = \Re_1 \bigcup \Re_2$. In addition to the two points on the x-axis and y-axis, the boundaries of the two throughput regions intersect at the point $(\lambda_1^*, \lambda_2^*) = (\alpha p_{13}, (1-\sqrt{\alpha})^2 p_{23})$, with $\alpha = \frac{4\left(\sqrt{p_{12}-p_{12}p_{13}+p_{13}}-\sqrt{p_{13}}\right)^2 (p_{12}-p_{12}p_{13}+p_{13}) p_{23}^2}{(p_{12}-p_{12}p_{13})^2 (p_{13}+p_{23})^2}$. In the subset constrained by $\lambda_1 \leq \lambda_1^*$ (or equivalently, $\lambda_2 \geq \lambda_2^*$), $\Re_2$ strictly contains $\Re_1$; in the subset constrained by $\lambda_1 > \lambda_1^*$ (or equivalently, $\lambda_2 < \lambda_2^*$), $\Re_1$ strictly contains $\Re_2$. Therefore, among the class of scheduling policies considered, the optimal policy at the relay node in achieving the maximum throughput region is described as follows: if we want to achieve the throughput rates for $\lambda_1 \leq \lambda_1^*$ (or equivalently, $\lambda_2 \geq \lambda_2^*$), the optimal scheduling policy is the policy which assigns higher priority to the relay queue Q_2^1 ($\beta = 1$); if we want to achieve the throughput rates for $\lambda_1 > \lambda_1^*$ (or equivalently, $\lambda_2 < \lambda_2^*$), the optimal scheduling policy is the one which assigns higher priority to relay's source queue Q_2^2 ($\beta = 0$). The latter case corresponds to the scenario of the non-cooperative operation, as Q_2^2 never empties and Q_2^1 will never be served. The optimal scheduling policy under condition (4.19) is shown in Fig. 4.4(b).

4.5 Network Coding at the Relay Node

The idea of Network Coding was first introduced in [47]. It allows nodes to perform operations on the bits inside the packets that are received from different sources. In [48] where a configuration similar to the one we are considering here was analyzed. It was shown that certain coding schemes at the relay node achieve the *min-cut capacity*; among the coding schemes at the relay node, greedy random linear coding was proved to be rate-optimal when the channels are noiseless, provided that the relay node can transmit and receive information simultaneously.

With both relayed packets and source packets at the relay node R, we investigate the performance of network coding when it is applied on these two streams of packets, and compare it to the simple retransmission scheme. When node R decides to transmit, network coding is performed in the following fashion: if both queues Q_2^1 and Q_2^2 are non-empty, node R transmits a random linear combination of two packets, one from Q_2^1, the other from Q_2^2; if only one queue is non-empty, node R transmits an uncoded packet from that queue. The random linear combination is assumed to be taken over a sufficiently large finite field, so the coefficient vectors of the same set of packets generated in different time slots are linearly independent with probability approaching one. It is easy to show the following result:

Theorem 4.7 *The superposition of network coding over cooperative relaying does not introduce additional performance gains for either the stability region or the throughput region, under either scheduled access or random access. The performance of network coding at the relay node R in this three-node packet-erasure network is*

the same as that of plain store-and-forward routing.

Proof: Consider the coded packet $\alpha_1 A + \alpha_2 B$, where packets A and B are from the queues Q_2^1 and Q_2^2 respectively, and coefficients α_1, α_2 are randomly generated in each time slot. Since both A and B are new to the destination, the destination needs to successfully receive two such encoded packets to decode both A and B. Therefore, we can view each encoded packet as a new uncoded packet and it follows that network coding yields the same performance as the plain store-and-forward routing in this case. ■

This conclusion can be extended to the case where the relay node is allowed to perform random linear network coding on K packets, for any $K \geq 2$; or even the greedy random coding scheme which allows the relay to transmit a linear combination of all packets in its queue at each time. There will be no performance gain over the plain store-and-forward routing in either the stability or throughput region, by following the same argument. Furthermore, if we take into account the non-zero probability of linear dependence between the coding vectors applied to the same set of combined packets, the destination is required to receive on average more than two encoded packets in order to decode two individual packets, and hence, network coding can possibly decrease the stability or throughput region. This is not surprising since network coding generally yields benefits for multicasting environments with multiple destinations. However, if network coding is performed in different ways, it is possible to yield enhancement to the obtained regions.

4.6 Numerical Results

Numerical results illustrating the relationship between the stability and throughput regions in the cooperative multi-access system are provided in this section. Fig. 4.5 plots the stability and throughput regions under scheduled access, for the channels with reception probabilities $p_{13} = 0.3$, $p_{23} = 0.8$ and $p_{12} = 0.6$. It is seen that the relationship between the regions is consistent with the results in Section 4.3, that is, $L_{\text{SR}} = L_2 = L_3 \supset L_1$.

The situation is more complicated in the random access case. The stability and throughput regions under random access are shown in Fig. 4.6 and Fig. 4.7. In Fig. 4.6, the channel reception probabilities are chosen to be $p_{13} = 0.15$, $p_{23} = 0.8$, $p_{12} = 0.7$, so that Eq. (4.18) is satisfied. In this case, the cooperative stability region and the throughput regions under policies 2 and 3 are observed to be the same, following the solid line; whereas the throughput region under policy 1 depicted by the dotted line is strictly inferior to the other three regions. That is, $\Re_{\text{SR}} = \Re_2 = \Re_3 \supset \Re_1$.

In Fig. 4.7, the channel reception probabilities are chosen to be $p_{13} = 0.4$, $p_{23} = 0.7$, $p_{12} = 0.6$, so that Eq. (4.19) is satisfied. The boundaries of the cooperative stability region and the non-cooperative stability region (which is also the throughput region under policy 1) cross at the point $(\lambda_1, \lambda_2) = (0.2167, 0.0482)$. The throughput regions under policies 2 and 3, and the cooperative stability region are shown to confirm that $\Re_3 \supset \Re_2 \supset \Re_{\text{SR}}$.

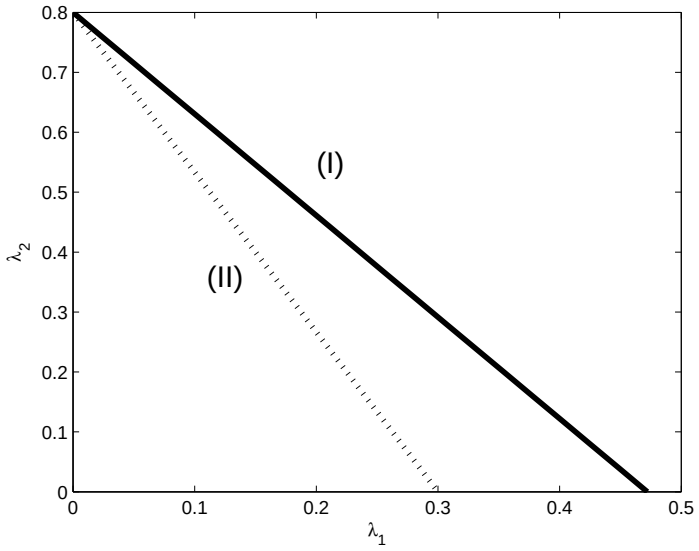

Figure 4.5: Stability region and throughput regions under scheduled access. $p_{13} = 0.3$, $p_{23} = 0.8$, $p_{12} = 0.6$. (I) Stability region; throughput regions under policies 2 and 3. (II) Throughput region under policy 1.

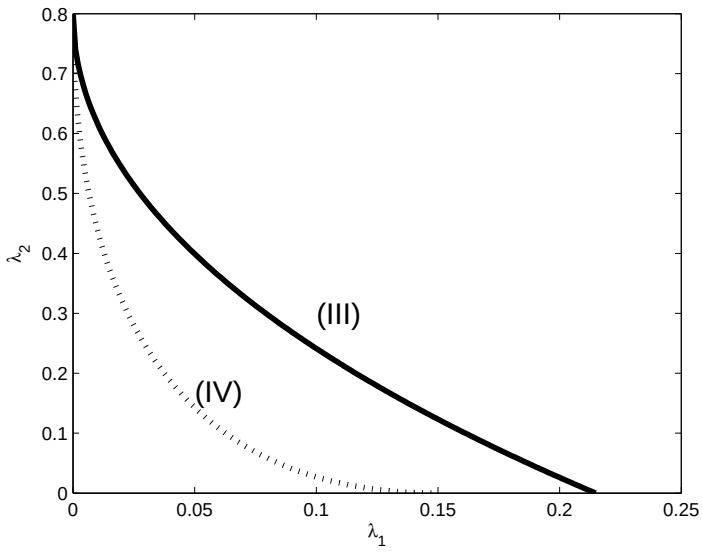

Figure 4.6: Stability region and throughput regions under random access. $p_{13} = 0.15$, $p_{23} = 0.8$, $p_{12} = 0.7$. (III) Stability region; throughput regions under policies 2 and 3. (IV) Throughput region under policy 1.

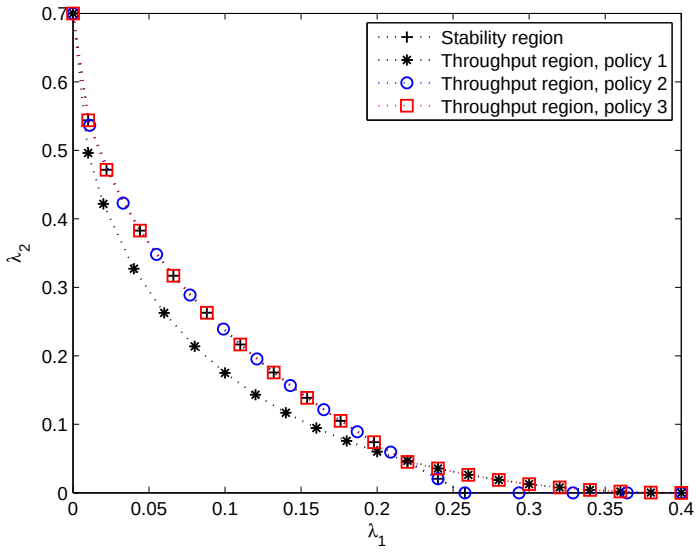

Figure 4.7: Stability region and throughput regions under random access. $p_{13} = 0.4$, $p_{23} = 0.7$, $p_{12} = 0.6$.

4.7 Discussion

In this chapter, we investigated the effects of network-level cooperation in a wireless multi-access system with two sources unicasting traffic to a common destination over erasure channels, under scheduled access and random access schemes respectively. After cooperation was permitted, the stability region and the throughput region were explicitly characterized, and their relationship was revisited, amplified, and explained. We concluded that the stability region is independent of the scheduling policies at the relay node assigned to the source packets and relayed packets; while the throughput region depends on the scheduling policies. Among a class of scheduling policies considered, we identified the two regions and determined when and whether they are identical or not. This observation demonstrates that, the addition of the extra link between the two users brings in a new dimen-

sion of complexity that affects the relationship between the stability region and the throughput region in multi-access systems.

Unlike the case of scheduled access, we observed that the cooperative stability region under random access does not always strictly outer-bound the non-cooperative stability region; this is due to the lack of coordination associated with random access, which leads to collision of simultaneous transmissions.

Finally, integration of network coding at the relay node was considered. It turned out that this does not yield additional performance gains over plain store-and-forward routing in terms of the stability region or throughput region. This is somewhat surprising since the information capacity region of the single-relay channel can be achieved if network-coding-like processing of information at the relay node is permitted. However, our model is different in that it allows the relay node to multiplex its own traffic with the traffic it relays.

4.8 Appendix

4.8.1 Proof of Theorem 4.4

We start by considering a dominant system, denoted by $\mathcal{M}^1$. In system $\mathcal{M}^1$, node S transmits dummy packets with probability q_1 whenever it is empty, while node R behaves in the same way as in the original system. All the other assumptions, channel models, arrival processes and reception processes remain unaltered in the dominant system. Since the dummy packets have no contribution to the throughput but cause collision with the transmission from the other source terminal, it follows

that the queue sizes in the dominant system are not smaller than the corresponding queue sizes in the original system, which implies that the stability of the dominant system is a sufficient condition for the stability of the original system.

We merge the two queues at node R, and denote the merged queue by Q_2. The two queues at node R are stable if and only if the merged queue Q_2 is stable. In the dominant system $\mathcal{M}^1$, node S always transmits with probability q_1, so the probability of success seen by Q_2 is always $q_2(1-q_1)p_{23}$, that is to say, the average service rate of Q_2 is

$$\mu_{Q_2} = q_2(1-q_1)p_{23} \tag{4.38}$$

To derive the stability condition for Q_2, we need to calculate its total arrival rate. There are two independent arrival processes to Q_2: first, the external Bernoulli arrivals to node R that line up in queue Q_2^2, and second, the arrivals from node S, stored in queue Q_2^1. The arrival rate to Q_2^2 is λ_2; it remains to calculate the arrival rate to Q_2^1.

In system $\mathcal{M}^1$, when node R receives a dummy packet from node S, it will simply *discard* the dummy packet. When the dominant system is stable, the queue at node S is stable, so the departure rate of the source packets (dummy packets are excluded) is equal to the arrival rate to node S, which is λ_1. When node S transmits a source packet, the packet will leave node S's queue if the following two events happen together: (i) node R is empty, or node R is backlogged but decides to be silent with probability $1-q_2$, and (ii) either the destination D or node R decodes

the packet, or both of them decode the packet. These two events described under (i) and (ii) are independent, and thus we have

$$\begin{aligned}&\mathbf{P}\left[\text{the transmitted packet of node } S \text{ departs } S\right]\\ =&\left\{(1-q_2)\mathbf{P}\left[Q_2\neq 0\right]+\mathbf{P}\left[Q_2=0\right]\right\}\left(1-(1-p_{12})(1-p_{13})\right)\end{aligned} \tag{4.39}$$

Among the packets that exit node S's queue, some will exit the network if they are decoded by the destination node directly, and some of them will be relayed to node R first. The latter case will happen if the following two events happen together: (i) node R decodes the packet, and (ii) the destination node D doesn't decode the packet. Therefore, we have

$$\begin{aligned}&\mathbf{P}\left[\text{the transmitted packet of node } S \text{ is relayed to node } R\right]\\ =&\left\{(1-q_2)\mathbf{P}\left[Q_2\neq 0\right]+\mathbf{P}\left[Q_2=0\right]\right\}p_{12}(1-p_{13})\end{aligned} \tag{4.40}$$

From Eq. (4.39) and Eq. (4.40), we obtain the conditional probability that a transmitted packet of node S (dummy packets are excluded) arrives at node R given that the transmitted packet exits node S's queue, which is

$$\begin{aligned}&\frac{\left\{(1-q_2)\mathbf{P}\left[Q_2\neq 0\right]+\mathbf{P}\left[Q_2=0\right]\right\}p_{12}(1-p_{13})}{\left\{(1-q_2)\mathbf{P}\left[Q_2\neq 0\right]+\mathbf{P}\left[Q_2=0\right]\right\}\left(1-(1-p_{12})(1-p_{13})\right)}\\ =&\frac{p_{12}(1-p_{13})}{1-(1-p_{12})(1-p_{13})}\end{aligned} \tag{4.41}$$

As we have described above, when the queue at S is stable, the departure rate

of the source packets (dummy packets are excluded) is equal to the arrival source rate λ_1, so the average arrival rate to the queue Q_2^1 at node R is computed to be $\frac{p_{12}(1-p_{13})}{1-(1-p_{12})(1-p_{13})}\lambda_1$. And the total arrival rate to Q_2 is

$$\lambda_{Q_2} = \lambda_2 + \frac{p_{12}(1-p_{13})}{1-(1-p_{12})(1-p_{13})}\lambda_1 \tag{4.42}$$

By Loynes' Theorem, the stability condition for the queue Q_2 at node R is given by $\lambda_{Q_2} < \mu_{Q_2}$, that is

$$\frac{p_{12}(1-p_{13})}{1-(1-p_{12})(1-p_{13})}\lambda_1 + \lambda_2 < q_2(1-q_1)p_{23} \tag{4.43}$$

Then we analyze the queue Q_1 at node S in system $\mathcal{M}^1$. The arrival rate to Q_1 is λ_1; the service process of Q_1 depends on whether Q_2 at node R is empty or not: if Q_2 is empty, Q_1 receives a service rate of $q_1\left(1-(1-p_{12})(1-p_{13})\right)$; otherwise, if Q_2 is not empty, Q_1 receives a service rate of $(1-q_2)q_1\left(1-(1-p_{12})(1-p_{13})\right)$. We now need to calculate the probability that Q_2 is empty. The total arrival process to Q_2 is no longer Bernoulli, since there can be possibly two packets arriving at R in the same time slot, one to Q_2^2, the other to Q_2^1 from Q_1. However, these two arrival processes are independent, and the inter-arrival times regarding the total arrival process remain geometrically distributed. By using the result provided in [49, Ch. 6.2], the stationary probability that Q_2 is empty still has the form of $1-\lambda_{Q_2}/\mu_{Q_2}$.

As a result, the average service rate received by Q_1 is

$$\begin{aligned}\mu_{Q_1} &= \left\{(1-q_2)\mathbf{P}\left[Q_2 \neq 0\right] + \mathbf{P}\left[Q_2 = 0\right]\right\} q_1 \left(1-(1-p_{12})(1-p_{13})\right) \\ &= \left(1 - \frac{\lambda_2 + \frac{p_{12}(1-p_{13})}{1-(1-p_{12})(1-p_{13})}\lambda_1}{(1-q_1)p_{23}}\right) q_1 \left(1-(1-p_{12})(1-p_{13})\right) \end{aligned} \tag{4.44}$$

And the stability condition for Q_1 is

$$\begin{aligned} & \lambda_{Q_1} < \mu_{Q_1} \\ \Longleftrightarrow & \frac{(1-q_1)p_{23} + q_1 p_{12}(1-p_{13})}{q_1(1-q_1)p_{23}\left(1-(1-p_{12})(1-p_{13})\right)}\lambda_1 + \frac{1}{(1-q_1)p_{23}}\lambda_2 < 1 \end{aligned} \tag{4.45}$$

The network is stable if and only if all queues in the network are stable, and from Eq. (4.43) and Eq. (4.45), the stability region of the dominant system $\mathcal{M}^1$ is characterized by

$$\Re_{\mathcal{M}^1} = \left\{ (\lambda_1, \lambda_2) : \begin{array}{l} \frac{(1-q_1)p_{23}+q_1p_{12}(1-p_{13})}{q_1(1-q_1)p_{23}(1-(1-p_{12})(1-p_{13}))}\lambda_1 + \frac{1}{(1-q_1)p_{23}}\lambda_2 < 1 \\ \\ \frac{p_{12}(1-p_{13})}{1-(1-p_{12})(1-p_{13})}\lambda_1 + \lambda_2 < q_2(1-q_1)p_{23} \end{array} \right\} \tag{4.46}$$

We now consider another dominant system $\mathcal{M}^2$. In system $\mathcal{M}^2$, node R transmits dummy packets with probability q_2 whenever it is empty, and node S behaves in the same way as in the original system. By following a parallel argument, we obtain the stability region of the dominant system $\mathcal{M}^2$ as

$$\Re_{\mathcal{M}^2} = \left\{ (\lambda_1, \lambda_2) : \begin{array}{l} \lambda_1 < q_1(1-q_2)\left(1-(1-p_{12})(1-p_{13})\right) \\ \\ \lambda_2 + \frac{(1-q_2)p_{12}(1-p_{13})+q_2p_{23}}{(1-q_2)(1-(1-p_{12})(1-p_{13}))}\lambda_1 < q_2p_{23} \end{array} \right\} \tag{4.47}$$

For a fixed transmission probability pair (q_1, q_2), the union of the two regions $\Re_{\mathcal{M}^1} \bigcup \Re_{\mathcal{M}^2}$ given by Eq. (4.46) and Eq. (4.47) provides the sufficient condition for the stability of the original system. By following the same "indistinguishability" argument as in [17], we observe that at saturation, the dominant systems and the original system become indistinguishable and hence must have the same boundary of the stability regions. As a result, the exact stability region of the original system *coincides* with the union of the stability regions of the two dominant systems, as (q_1, q_2) varies over $[0, 1]^2$.

Finally, to obtain the stability region over all (q_1, q_2), we solve a constrained optimization problem as in [19]. To do this, we fix λ_1 and maximize λ_2 as (q_1, q_2) varies over $[0, 1]^2$, which then generates the boundary of the stability region. The procedure is as follows.

We replace λ_1 by x and λ_2 by y, and the boundary of the stability regions given by Eq. (4.46) and Eq. (4.47) can be written as

$$y = (1-q_1)p_{23} - \frac{\big((1-q_1)p_{23} + q_1 p_{12}(1-p_{13})\big)x}{q_1(p_{12}+p_{13}-p_{12}p_{13})}$$
$$\text{for} \qquad 0 \le x \le \frac{(p_{12}+p_{13}-p_{12}p_{13})(q_2(1-q_1)p_{23}-y)}{p_{12}(1-p_{13})} \tag{4.48}$$

and

$$y = q_2 p_{23} - \frac{\big((1-q_2)p_{12}(1-p_{13}) + q_2 p_{23}\big)x}{(1-q_2)(p_{12}+p_{13}-p_{12}p_{13})}$$
$$\text{for} \qquad 0 \le x \le q_1(1-q_2)(p_{12}+p_{13}-p_{12}p_{13}) \tag{4.49}$$

First we consider the constrained optimization problem suggested by Eq. (4.48) as

$$\max_{q_1 \in [0,1]} y' = \max_{q_1 \in [0,1]} \left((1-q_1)p_{23} - \frac{\big((1-q_1)p_{23} + q_1 p_{12}(1-p_{13})\big)x}{q_1(p_{12}+p_{13}-p_{12}p_{13})} \right) \tag{4.50}$$

Differentiating with respect to q_1 gives

$$\frac{dy'}{dq_1} = -p_{23} + \frac{p_{23}x}{q_1^2(p_{12}+p_{13}-p_{12}p_{13})} \tag{4.51}$$

Setting Eq. (4.51) to zero yields

$$q_1^* = \sqrt{\frac{x}{p_{12}+p_{13}-p_{12}p_{13}}} \tag{4.52}$$

A simple calculation shows that the second derivative at q_1^* is negative, and so q_1^* must be the unique maximizer if it can be reached. Since q_1 is a probability, we have $0 \leq q_1 \leq 1$; according to Eq. (4.52), this is equivalent to require

$$0 \leq x \leq p_{12}+p_{13}-p_{12}p_{13} \tag{4.53}$$

But by Eq. (4.48), y must be non-negative, and so x automatically satisfies Eq. (4.53).

Hence, by substituting Eq. (4.52) into Eq. (4.48), we obtain

$$y'_{\max} = -\frac{\big(p_{12}(1-p_{13})-p_{23}\big)x}{p_{12}+p_{13}-p_{12}p_{13}} - \frac{2p_{23}\sqrt{x}}{\sqrt{p_{12}+p_{13}-p_{12}p_{13}}} + p_{23}$$

$$\Longleftrightarrow \sqrt{\frac{x}{p_{12}+p_{13}-p_{12}p_{13}}} + \sqrt{\frac{p_{12}(1-p_{13})x}{p_{23}(p_{12}+p_{13}-p_{12}p_{13})} + \frac{y'_{\max}}{p_{23}}} = 1 \qquad (4.54)$$

This is the boundary of the stability region provided by Eq. (4.48).

By applying the same procedure on Eq. (4.49), we can show that its boundary is also given by Eq. (4.54). Therefore, the stability region of the cooperative multi-access system under random access is described by

$$\Re_{\text{SR}} = \left\{ (\lambda_1, \lambda_2) : \sqrt{\frac{\lambda_1}{p_{12}+p_{13}-p_{12}p_{13}}} + \sqrt{\frac{p_{12}(1-p_{13})\lambda_1}{p_{23}(p_{12}+p_{13}-p_{12}p_{13})} + \frac{\lambda_2}{p_{23}}} < 1 \right\} \qquad (4.55)$$

This concludes the proof of Theorem 4.4. We should remark here that when we construct the proof, node R is allowed to randomly pick a packet from the merged queue Q_2 to transmit (that is, either pick a packet of its own or a relayed packet from S); the analysis and derivation we presented is independent of what scheduling policy node R adopts, and hence, the stability region is independent of the scheduling policies considered and is given by Eq. (4.16). ■

Chapter 5

Rate Control

5.1 Introduction

As another arena for studying wireless multi-access channels, rate control is the main subject of this chapter. The traditional networking-theoretic approach in multi-access channels rests on the packet level, which assumes that data travels in the form of packets, without regard to the bit-content. By doing this, the physical layer properties which include the effects of fading, attenuation and interference are abstracted into a simple packet-erasure channel model: the transmission is considered to be successful with a certain probability, and unsuccessful otherwise. Then the performance measures of interest are the packet throughput, packet delay, etc. However, packets cannot be treated separately from the internal bits. If a packet contains more bits and hence is transmitted at a higher data rate, the successful reception probability of that packet is relatively low; on the other hand, higher packet throughput can be achieved by lowering the data rate. Therefore, it is more appropriate and accurate to consider the problem at the *bit* level, and evaluate the performance metrics such as bit throughput and bit delay.

There have been many papers which investigate the problem of rate allocation in multi-access channels, with the objective to maximize the throughput or minimize the average delay, including [50,51,52,53]. However, those work is subject to a couple

of limitations. First, their channel model is restricted to additive white Gaussian noise multiple access channel only, the capacity region of which has a polymatroidal structure that is essential for their proof of optimality. Second, in those work, the transmission rates assigned to the users are selected from the Shannon capacity region, and the optimal rate control policy is shown to be operated on the boundary of the capacity region. However, the Gaussian capacity region is obtained when we assume the availability of infinite number of source bits at the transmitters, which may not hold under the assumption of bursty arrivals. In addition, the rates on the boundary are asymptotically achievable when the codeword length approaches infinity, thus, such rates cannot be achieved with finite-length codeword transmitted in a finite time slot. And, the infinite-length codeword implies infinite delay, which makes the delay minimization problem questionable.

In this chapter, we first investigate the stability issue of a multi-access channel consisting of two users and a common destination. In each time slot, information bits arrive at the two users independently according to some stationary processes. Due to the difficulty and cost of global coordination in the wireless environment, here, we assume that each user, if it is backlogged, decides to transmit with some probability independently of any other event. Then, the transmission rates are selected from a finite, discrete set, such that the allocated rates are feasible according to the action of both users. The feasible rate set can be derived from any channel model, and is not restricted to the Gaussian channel only. Basically, if only one user decides to transmit, the channel is able to support higher transmission rate; otherwise, if both users decide to transmit together in a time slot, the supportable

data rate from each user must be lowered in order to combat interference and ensure successful reception at the destination. By doing this, we require that the users exchange the information about whether they will transmit or remain idle. Since only one bit is needed for the side information from each user (active or idle), the cost incurred by this information exchange is negligible. Under these assumptions, we explicitly characterize the stability region (in bits/slot) of this multi-access system. The stability region is defined as the set of all arrival rates for which there exist transmission probabilities such that all queues in the network remain stable. Subsequently, the structure property of the stability region is discussed, and we determine the necessary and sufficient condition under which the stability region is convex. In this case, the optimal policy for attaining the maximum achievable stability region is shown to be the rate allocation policy with transmission probability one for both users. According to our model, the stability region is the union of all stabilizable arrival rates by some transmission probabilities. Hence, we *define* a policy to be optimal if it can stabilize all arrival rates that are stabilizable with some transmission policy; and the stability region of the optimal policy, which is indeed the entire stability region, is also referred to as the maximum stability region.

In the second part of this chapter, we focus our attention on the delay performance. We look into the case where users are initially given a certain amount of data to be delivered to the destination, and there are no more arrivals. We are interested in characterizing the optimal policy that drains the queues of two users within the shortest time. Under the same condition for a convex stability region, the optimal policy is shown to have similar structure as the policy investigated in [54,55]

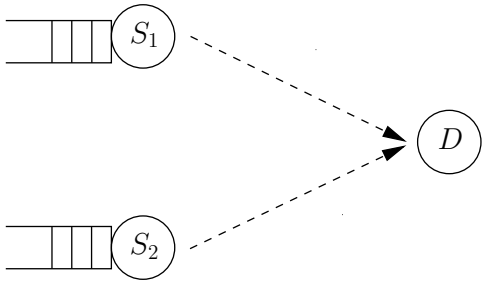

Figure 5.1: The wireless multi-access channel with two users (S_1, S_2), and a common destination (D).

for the Gaussian broadcast channels, and is the one that empties the two queues with the same expected time.

5.2 Model

We consider throughout this chapter the slotted wireless multi-access channel shown in Fig. 5.1. Two users, S_1 and S_2, transmit unicast traffic to the common destination D. In each time slot, packets arrive to source user S_i according to a Bernoulli process with rate η_i packets per slot, and each packet contains an average number of l_i information bits. Hence, the average arrival rate to each user, measured in bits per slot, is $\lambda_i = \eta_i l_i$, for $i = 1, 2$. Further, we assume that the arrival processes are independent across users and independently and identically distributed over time slots.

Different from the packet-based approach in most work at the network layer, which assumes the transmission of one packet in a time slot without regard to the packet-content, in this chapter, we take into account the bit-nature of a packet. As such, we assume that each user, if it transmits, will transmit a finite number of bits. These bits are then encoded into a packet with fixed length that is transmitted for

the duration of one time slot. Generally, the transmission rates can be chosen from a continuous set. But for simplifying the analysis, we assume that user S_i ($i \in \{1, 2\}$) is allowed to transmit with one of the two rates: R_i or r_i. The rates are chosen such that R_i is the supportable rate that S_i can transmit to D without the presence of $S_{\bar{i}}$'s transmission ($\bar{i}$ denotes the complementary of i); and r_i is the rate S_i can transmit if both users transmit simultaneously. Due to the interference caused by concurrent transmissions, the supportable rate under single transmission must be greater than the supportable rate under simultaneous transmissions. That is, $R_i > r_i$, for $i = 1, 2$. These rates depend on various parameters such as the received power, target bit-error-rate, as well as the coding and decoding mechanisms. For example, the rate can be defined as some increasing function of the received signal-to-interference-plus-noise ratio (SINR) as in [56]. In some papers, the rate is represented by the Shannon capacity rate $\log_2(1 + \text{SINR})$. As we have commented in Section 5.1, although this approach based on Shannon capacity is widely used for the rate control problem, its validity is questionable. However, the actual choices for the transmission rates will not affect our analysis.

We would like to remark that our model is general since we do not restrict it to any specific channel model (e.g., additive Gaussian multiple-access channel); all rates chosen following the principle described above are included in our model. In addition, we incorporate the physical-layer properties into the network-layer studies by considering more realistic channel model than the packet reception channel model.

Denote by $q_1[n]$ and $q_2[n]$ the number of bits in the queues at S_1 and S_2

respectively, at the beginning of time slot n. Instead of a centralized scheduling like Time Division Multiple Access (TDMA) or a fully distributed random access scheme, here, we consider a class of transmission policy that is partially distributed with some coordination. The transmission policy is distributed as each user decides to transmit with some probability if it is backlogged, independent of any other event. Then, users communicate over a control sub-channel exchanging the information whether they decide to transmit or not; according to the action of the other user, each user selects the feasible rate to transmit. At each time slot n, the transmission policy is described as follows:

- If $q_i[n] > 0$, user S_i decides to transmit with probability p_i. If $q_i[n] = 0$, S_i has to be idle with probability 1.

- If both users are backlogged and decide to transmit, by exchanging this information they know that they both will transmit, so each user S_i ($i \in \{1, 2\}$) selects the rate r_i to transmit, as r_i is the supportable rate when S_1 and S_2 transmit together.

- If only one of the users, say user S_i, decides to transmit, it transmits with rate R_i.

We note that the control information shared by the two users requires only one bit from each side (transmit or idle), and hence, the overhead incurred by this type of coordination is negligible. Defined in this way, we see that in each time slot, there are three possible allocation rate pairs: $(R_1, 0)$, (r_1, r_2) and $(0, R_2)$, according to the queue state and the action chosen by the two users.

5.3 Stability Region

As defined by the transmission policy, user S_i independently decides to access the channel, and the queues at S_1 and S_2 interact in a complicated manner. As before, by the stochastic dominance approach and Loynes' Theorem, we are able to solve the stability conditions in the context of our bit-level multi-access channel model. The main results are provided in the following theorem.

Theorem 5.1 *The parameters λ_1, λ_2 represent the average arrival rates to the users S_1 and S_2 respectively, measured in bits/slot. The stability region of the investigated multi-access channel is given by*

$$\Re = \bigcup_{(p_1,p_2)\in[0,1]^2} \Big(\Re_1(p_1,p_2) \bigcup \Re_2(p_1,p_2)\Big) \tag{5.1}$$

where

$$\Re_1(p_1,p_2) = \left\{ (\lambda_1,\lambda_2): \begin{array}{l} \lambda_1 < p_1R_1 - \frac{p_1(R_1-r_1)}{R_2-p_1(R_2-r_2)}\lambda_2 \\ \lambda_2 < p_2R_2 - p_1p_2(R_2-r_2) \end{array} \right\} \tag{5.2}$$

$$\Re_2(p_1,p_2) = \left\{ (\lambda_1,\lambda_2): \begin{array}{l} \lambda_1 < p_1R_1 - p_1p_2(R_1-r_1) \\ \lambda_2 < p_2R_2 - \frac{p_2(R_2-r_2)}{R_1-p_2(R_1-r_1)}\lambda_1 \end{array} \right\} \tag{5.3}$$

and $(p_1,p_2) \in [0,1]^2$ is the transmission probability pair.

Proof: See Appendix 5.6.1. ■

Lemma 5.1 *By varying the transmission probabilities (p_1,p_2) over $[0,1]^2$, the union*

of the stability region for the multi-access channel takes one, and only one of the following forms:

(a) If

$$\frac{r_1}{R_1} + \frac{r_2}{R_2} < 1 \tag{5.4}$$

the stability region is given by

$$\Re = \mathcal{L}_1 \bigcup \mathcal{L}_2 \bigcup \mathcal{L}_3 \tag{5.5}$$

where

$$\mathcal{L}_1 = \left\{ (\lambda_1, \lambda_2) : \lambda_2 < R_2 - \frac{(R_2 - r_2)\lambda_1}{r_1}, \quad for \ \ \lambda_1 \in \left[0, \frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)}\right) \right\} \tag{5.6}$$

$$\mathcal{L}_2 = \left\{ (\lambda_1, \lambda_2) : \sqrt{\frac{(R_2 - r_2)\lambda_1}{R_1 R_2}} + \sqrt{\frac{(R_1 - r_1)\lambda_2}{R_1 R_2}} < 1, \right.$$
$$\left. for \ \ \lambda_1 \in \left[\frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)}, \frac{R_1(R_2 - r_2)}{R_2}\right) \right\} \tag{5.7}$$

$$\mathcal{L}_3 = \left\{ (\lambda_1, \lambda_2) : \lambda_2 < \frac{R_1 r_2}{R_1 - r_1} - \frac{r_2 \lambda_1}{R_1 - r_1}, \quad for \ \ \lambda_1 \in \left[\frac{R_1(R_2 - r_2)}{R_2}, R_1\right) \right\} \tag{5.8}$$

(b) Otherwise, that is, if

$$\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1 \tag{5.9}$$

the optimal transmission probability pair for maximizing the stability region is

$(p_1^*, p_2^*) = (1, 1)$. *The resulting stability region is convex, and is given by*

$$\Re = \mathcal{I}_1 \bigcup \mathcal{I}_2 \tag{5.10}$$

where

$$\mathcal{I}_1 = \left\{ (\lambda_1, \lambda_2) : \lambda_2 < R_2 - \frac{(R_2 - r_2)\lambda_1}{r_1}, \quad for \ \ \lambda_1 \in [0, r_1) \right\} \tag{5.11}$$

$$\mathcal{I}_2 = \left\{ (\lambda_1, \lambda_2) : \lambda_2 < \frac{R_1 r_2}{R_1 - r_1} - \frac{r_2 \lambda_1}{R_1 - r_1}, \quad for \ \ \lambda_1 \in [r_1, R_1) \right\} \tag{5.12}$$

Proof: See Appendix 5.6.2. ■

These results imply that, when the condition expressed in Eq. (5.4) is satisfied, the stability region is bounded by straight lines in some part, and bounded by a strictly convex function in the remaining part. Otherwise, when the condition expressed in Eq. (5.9) is satisfied, the stability region becomes convex, and is bounded by two straight lines from the axes.

Further, when Eq. (5.9) is satisfied, each user should choose to transmit whenever it is backlogged. Such policy will stabilize any arrival rate vectors that are stabilizable under some rate control policy we investigate. The resulting stability region is convex, and is bounded by two straight lines that intersect at (r_1, r_2). In Fig. 5.2, we display some diagrams to illustrate the properties of the stability region in different cases. In Fig. 5.2(a), the condition in Eq. (5.4) holds, the stability region is bounded by two straight lines from the axes and a convex function in the interior; whereas, in Fig. 5.2(b), the condition in Eq. (5.9) is satisfied, and the stability region

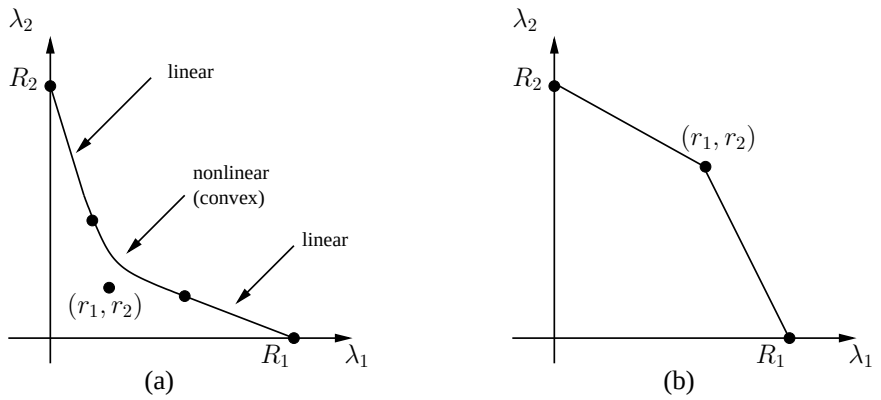

Figure 5.2: The closure of the stability region. (a) Eq. (5.4) is satisfied, $\Re$ is bounded by straight lines close to the axes and a convex curve in the interior. (b) Eq. (5.9) is satisfied, $\Re$ is convex.

becomes convex.

5.4 Minimum Delivery Time Policy

In the previous section, we characterized the optimal policy that yields the maximum stability region for the case $\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1$ and $\frac{r_1}{R_1} + \frac{r_2}{R_2} < 1$ separately. However, a policy that is throughput-optimal does not guarantee optimal delay performance. For example, the backpressure algorithm introduced in [22] was proved to achieve the maximum stability region in a relatively general network model, but it can have poor delay performance.

As a separate, but related issue, we focus our attention on the minimum delivery time policy in this section. We consider the scenario where there is an initial amount of traffic volume at the source users at a given time, and there are no more arrivals; the objective is then to minimize the total time to empty the queues of both users. We first analyze the case of $\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1$, we will show that the minimum delivery time policy in our problem has a similar property as the Queue

Proportional Scheduling (QPS) policy studied in [54,55]. In those work, the authors worked on the Gaussian broadcast channels, and the operation space for all feasible transmission rates is the Shannon capacity region. In [54], the QPS policy was shown to achieve the maximum stability region; in [55], the QPS policy was shown to minimize the total delivery time. In [55], the action space for the users is the Gaussian Shannon capacity region, which is a continuous set. Therefore, as there are no new arrivals, and the system is operating on the rate point that is proportional to the initial queue size vector, it turns out that the queue sizes retain the same proportion. As a result, the system is operating on a fixed transmission rate point until both queues are emptied at the same time. However, in our work, there is only a discrete set for feasible allocation rate pairs: $\{(R_1, 0), (r_1, r_2), (0, R_2)\}$, so the system must operate on these points in a probabilistic fashion such that the average allocated rate is proportional to the initial queue size vector.

Then, we analyze the case of $\frac{r_1}{R_1} + \frac{r_2}{R_2} < 1$; this turns out to be a much simpler case. We will show that the minimum delivery time policy is the policy that separately schedules the transmissions of the two users, and each user transmits R_i bits during the assigned time slots until its queue is empty.

Theorem 5.2 *Denote by* $\mathbf{q_0} = (q_1, q_2)$ *the initial queue size vector at users* S_1 *and* S_2*. When the condition in Eq. (5.9) is satisfied, that is, if it is true that* $\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1$*, the minimum delivery time policy is described as follows:*

1. *If*

$$\frac{q_2}{q_1} \geq \frac{r_2}{r_1} \tag{5.13}$$

the optimal policy is to allocate the rates (r_1, r_2) *to the users with probability* $\tau = \frac{q_1 R_2}{q_1 R_2 + q_2 r_1 - q_1 r_2}$*, and allocate the rates* $(0, R_2)$ *with probability* $1 - \tau = \frac{q_2 r_1 - q_1 r_2}{q_1 R_2 + q_2 r_1 - q_1 r_2}$ *until one queue is emptied, after which the policy serves the remaining backlogged user* S_i *with rate* R_i*;*

2. *Otherwise, that is, if*

$$\frac{q_2}{q_1} < \frac{r_2}{r_1} \tag{5.14}$$

the optimal policy is to allocate the rates (r_1, r_2) *to the users with probability* $\tau' = \frac{q_2 R_1}{q_2 R_1 + q_1 r_2 - q_2 r_1}$*, and allocate the rates* $(R_1, 0)$ *with probability* $1 - \tau' = \frac{q_1 r_2 - q_2 r_1}{q_2 R_1 + q_1 r_2 - q_2 r_1}$ *until one queue is emptied, after which the policy serves the remaining backlogged user* S_i *with rate* R_i*.*

Proof: First, we note that to compare the total delivery time of different policies, it is equivalent to compare the average allocated rates to the two users. Suppose that the total delivery time of a policy is T, then the average allocated rate vector is given by $\mathbf{R}_{\text{Avg}} = \mathbf{q_0}/T$. As we have stated above, the set of all possible allocation rates in each time slot is $\{(R_1, 0), (r_1, r_2), (0, R_2)\}$. When the condition in Eq. (5.9) is satisfied, the union of all achievable average allocated rates coincides with the stability region expressed by Eq. (5.10), which is convex. Hence, the average allocated rate

vector must be inside the stability region, that is, $\mathbf{R}_{\text{Avg}} \in \Re$. Then, we show that the average allocated rate vector of the optimal policy must lie on the boundary of $\Re$. The argument is simple: suppose that the average allocated rate vector (d_1, d_2) of the optimal policy is not on the boundary but in the interior, we can always find a rate vector (d_1', d_2') on the boundary with $d_1' > d_1$ and $d_2' > d_2$, then a different policy with average allocation rates (d_1', d_2') must have better performance. Finally we can show that the optimal policy must empty the two queues using the same expected time. This can be proved as follows. Without loss of generality, we select a different policy that empties S_1's queue first, then the average rate allocated to user S_1 must be increased, at the expense of decreasing the rate allocated to user S_2. As a result, the queue at S_2 needs more time slots to empty, and the total delivery time is determined by the queue that is emptied last. Therefore, under the optimal policy, the two queues are emptied with the same expected time; and a policy with such property yields the average allocation rate vector proportional to the initial queue size vector.

Consequently, the explicit characterization of the optimal policy depends on the initial queue sizes. If $q_2/q_1 \geq r_2/r_1$, the initial queue size vector (q_1, q_2) falls inside the semi-infinite triangle (I) as illustrated by Fig. 5.3(a), and the optimal policy has the average allocation rates (l_1, l_2) as indicated in Fig. 5.3(a). To achieve this rate pair, the policy selects the rates (r_1, r_2) with probability τ, and selects the rates $(0, R_2)$ with probability $1 - \tau$; after simple calculation the probability τ is as given in Theorem 5.2. Otherwise, if $q_2/q_1 < r_2/r_1$ and the initial queue size vector falls inside the region (II) as shown by Fig. 5.3(b), a parallel statement can be made.

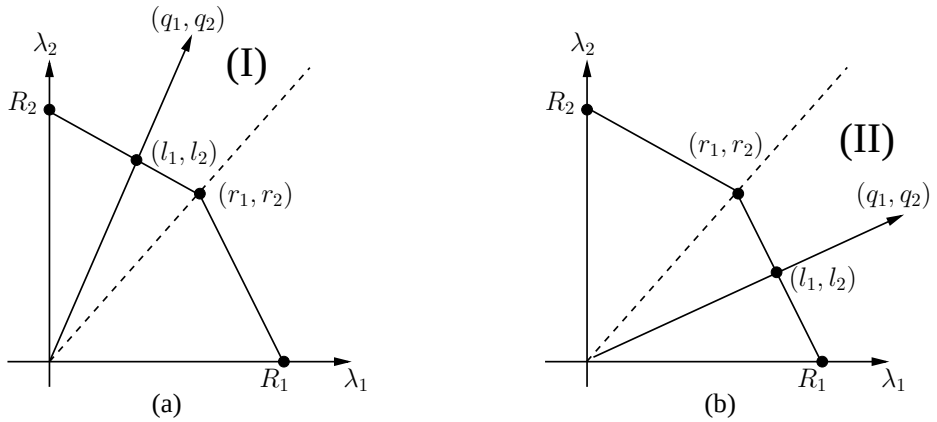

Figure 5.3: The average allocation rates of the optimal policy that minimizes the total delivery time. (a) Condition (5.13) holds, (q_1, q_2) is inside region (I). (b) Condition (5.14) holds, (q_1, q_2) is inside region (II).

In case a queue is emptied first, we allocate the maximum supportable rate to the remaining queue until it is empty. ■

Although it has been proved that, when the condition $\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1$ holds, the transmission policy with transmission probability 1 for both users is optimal for the stable throughput. Under the same condition, such policy does not drain the queues within the shortest time among all policies; and the minimum delivery time policy is the one described in Theorem 5.2 which takes into account the initial traffic demand in both queues. It will be more interesting to address the minimum average delay policy, which minimizes the average delay experienced by all users in the system when there are random arrivals. The minimum average delay problem appears to be challenging even in our two-user multi-access channel model, where both users are allowed to transmit simultaneously, though at a lower rate. This investigation is left to the future work.

Theorem 5.3 *For the case of* $\frac{r_1}{R_1} + \frac{r_2}{R_2} < 1$ *as described by Eq. (5.4), the minimum*

delivery time policy will perform a time sharing between the two rate points: $(R_1, 0)$ *and* $(0, R_2)$*, until both users' queues are empty. That is, the optimal policy is the one that schedules the transmissions of two users separately, until all the backlogged bits are delivered to the destination.*

Proof: Under the condition in Eq. (5.4), the relations of the three rate points are illustrated as in Fig. 5.2(a), where the rate point (r_1, r_2) is below the straight line formed by the points $(R_1, 0)$ and $(0, R_2)$. Suppose that the optimal policy should allocate the rates (r_1, r_2) for some time slots, then by averaging over all allocated rates, the average allocated rate must be at a point that is strictly below the straight line connected by $(R_1, 0)$ and $(0, R_2)$. However, for any such rate point (d_1, d_2), we can always find a rate point (d_1', d_2') that lies on the straight line with $d_1' > d_1$ and $d_2' > d_2$, thus an alternate policy with average allocation rate (d_1', d_2') will result in shorter delivery time. Therefore, we can argue that the optimal policy will perform a time sharing between the rate points $(R_1, 0)$ and $(0, R_2)$, and the transmissions of the users are assigned in disjoint time slots. Each user S_i transmits with rate R_i until its queue is empty. ■

5.5 Discussion

In this chapter, we studied the rate control problem in a wireless multi-access channel. Instead of using the Shannon capacity rate at the network layer, the transmission rates are selected from a finite, discrete set in which the rates are achievable in a finite time slot. In Section 5.3, under a class of transmission policies with some

coordination, we first characterized the stability region in terms of bits/slot. We showed that the optimal policies for maximizing the stability region depend on the values of some crucial network parameters. Under a certain channel condition, the maximum stability region is achieved by activating the users with probability one if they are backlogged, and the stability region is convex.

In Section 5.4, the minimum delivery time problem was investigated in our channel model. For any initial queue size vector, we explicitly characterized the optimal policy that empties the users' queues within the shortest time. Under the same condition for a convex stability region, the minimum delivery time policy was shown to be the one that drains the queues of both users with the same expected time.

5.6 Appendix

5.6.1 Proof of Theorem 5.1

We first find the stability region $\Re(p_1, p_2)$ for a fixed transmission probability pair (p_1, p_2). Tow parallel dominant systems $\mathcal{M}^i$ $(i \in \{1, 2\})$ are constructed in the following way:

- In the dominant system $\mathcal{M}^1$, user S_1 transmits dummy bits with probability p_1 if it is empty, while user S_2 acts the same as in the original system.

- In the dominant system $\mathcal{M}^2$, user S_2 transmits dummy bits with probability p_2 if it is empty, while user S_1 acts the same as in the original system.

The arrival processes, the channel model, the transmission and reception processes and all other assumptions remain unaltered in the dominant systems. Then we analyze the stability conditions with respect to each dominant system. In system $\mathcal{M}^1$, user S_1 always attempts to transmit with probability p_1, regardless of its queue state. Therefore, user S_2 experiences an average service rate of $p_2p_1r_2+p_2(1-p_1)R_2$. This is because: if S_1 decides to transmit, which occurs with probability p_1, S_2 is able to transmit with rate r_2; otherwise, if S_1 decides not to transmit with probability $1-p_1$, S_2 is able to transmit alone with rate R_2 to the destination. And if S_2 is backlogged, it will attempt to transmit with probability p_2. Hence, the service rate for S_2 in system $\mathcal{M}^1$ takes an average of $p_2p_1r_2 + p_2(1-p_1)R_2$. Then, according to Loynes' Theorem, the queue at S_2 is stable if and only if

$$\lambda_2 < p_2p_1r_2 + p_2(1-p_1)R_2 = p_2R_2 - p_1p_2(R_2 - r_2) \tag{5.15}$$

Now we analyze the stability condition of the queue at S_1. The service process of S_1 depends on the state of S_2's queue. When S_2 is not empty, S_2 will decide to transmit with probability p_2. In this case, S_1 is only able to deliver r_1 bits to the destination. Otherwise, if S_2 decides not to transmit, S_1 is able to transmit with rate R_1. On the other hand, when S_2 is empty, S_1 can always transmit with rate R_1. As a result, the average service rate of S_1 should be calculated as

$$\mu_1 = p_1\Big\{\mathbf{P}\left[q_2 \neq 0\right](p_2r_1 + (1-p_2)R_1) + (1-\mathbf{P}\left[q_2 \neq 0\right])R_1\Big\} \tag{5.16}$$

Since the queue at S_2 is a discrete-time M/M/1 queue, the stationary probability that the queue is not empty is $\mathbf{P}[q_2 \neq 0] = \frac{\lambda_2}{\mu_2} = \frac{\lambda_2}{p_2 R_2 - p_1 p_2 (R_2 - r_2)}$; by substituting this into Eq. (5.16), we can obtain the average service rate of S_1, with which we can determine the stability condition of S_1 by Loynes' Theorem. The queue at S_1 is stable if and only if

$$\lambda_1 < p_1 R_1 - \frac{p_1(R_1 - r_1)}{R_2 - p_1(R_2 - r_2)} \lambda_2 \tag{5.17}$$

Hence, the dominant system $\mathcal{M}^1$ is stable if and only if the queues at both S_1 and S_2 are stable; and its stability region, denoted by $\Re_1(p_1, p_2)$, is as shown in Eq. (5.2).

By following a parallel argument for system $\mathcal{M}^2$ in which user S_2 transmits dummy bits, the stability region of $\mathcal{M}^2$, denoted by $\Re_2(p_1, p_2)$, is characterized by Eq. (5.3) in Theorem 5.1.

We know from the stochastic dominance argument that, the stability of the dominant systems implies the stability of the original system. That is, the stability region of the original system $\Re(p_1, p_2)$ is lower-bounded by $\bigcup_{i=1,2} \Re_i(p_1, p_2)$, so we have $\Re(p_1, p_2) \supseteq \bigcup_{i=1,2} \Re_i(p_1, p_2)$. Then we proceed by arguing that the boundary of the original system indeed *coincides* with the boundary of the dominant systems, which is $\Re(p_1, p_2) = \bigcup_{i=1,2} \Re_i(p_1, p_2)$. Consider system $\mathcal{M}^1$ where S_1 continues to transmit dummy bits when it is empty, given that $\lambda_2 < p_2 R_2 - p_1 p_2 (R_2 - r_2)$, for those λ_1 that S_1's queue is stable in the dominant system, the corresponding queue is also stable in the original system; conversely, for those λ_1 which makes the queue

at S_1 unstable in the dominant system, S_1 never empties and hence S_1 always attempts to transmit source bits with probability p_1. We observe that as long as S_1 never empties, the dominant system $\mathcal{M}^1$ and the original system behave the same. Therefore, we see that at saturation, the two systems are "indistinguishable". A similar "indistinguishability" argument can be applied to system $\mathcal{M}^2$. Thus we conclude that $\Re(p_1, p_2)$ is equal to $\bigcup_{i=1,2} \Re_i(p_1, p_2)$. Once the stability region for a fixed transmission probability pair (p_1, p_2) is found, the stability region for the system is the union of all such regions as (p_1, p_2) varies over $[0, 1]^2$. This concludes the proof of Theorem 5.1. ■

5.6.2 Proof of Lemma 5.1

Since we have obtained the stability region for a fixed probability pair (p_1, p_2) in Theorem 5.1, we can utilize the constrained optimization technique as in [19] to derive the stability region for all possible values of (p_1, p_2). To do this, for a fixed λ_1, we maximize λ_2 as (p_1, p_2) varies over $[0, 1]^2$, where λ_1 and λ_2 are constrained by the conditions in Eq. (5.2) and Eq. (5.3) (We can reverse the role of λ_1 and λ_2, i.e., fix λ_2, then maximize λ_1 over all possible values of (p_1, p_2). The thus obtained region can be shown to be the same).

By replacing λ_1 with x and λ_2 with y, the boundaries of the stability region

given by Eq. (5.2) and Eq. (5.3) can be written as

$$x = p_1 R_1 - \frac{p_1(R_1 - r_1)y}{R_2 - p_1(R_2 - r_2)}$$
$$\text{for} \quad 0 \le y \le p_2 R_2 - p_1 p_2 (R_2 - r_2) \tag{5.18}$$

and

$$y = p_2 R_2 - \frac{p_2(R_2 - r_2)x}{R_1 - p_2(R_1 - r_1)}$$
$$\text{for} \quad 0 \le x \le p_1 R_1 - p_1 p_2 (R_1 - r_1) \tag{5.19}$$

First we consider the constrained optimization problem suggested by Eq. (5.19) as

$$\max_{p_2 \in [0,1]} y = \max_{p_2 \in [0,1]} \left(p_2 R_2 - \frac{p_2(R_2 - r_2)x}{R_1 - p_2(R_1 - r_1)} \right) \tag{5.20}$$

Differentiating with respect to p_2 gives

$$\frac{dy}{dp_2} = R_2 - \frac{R_1(R_2 - r_2)x}{(R_1 - p_2(R_1 - r_1))^2} \tag{5.21}$$

Setting Eq. (5.21) to zero yields

$$p_2^* = \frac{R_1}{R_1 - r_1} - \frac{1}{R_1 - r_1}\sqrt{\frac{R_1(R_2 - r_2)x}{R_2}} \tag{5.22}$$

A simple calculation shows that the second derivative at p_2^* is negative; so p_2^* must

be the unique maximizer if it can be reached. Since p_2^* is a probability and it must satisfy $0 \leq p_2^* \leq 1$, it follows from Eq. (5.22) that

$$\frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)} \leq x \leq \frac{R_1 R_2}{R_2 - r_2} \tag{5.23}$$

The constraint in Eq. (5.19) is valid only for $p_2 < \frac{R_1 - x}{R_1 - r_1}$, then for p_2^* given by Eq. (5.22) to be reachable, x must satisfy

$$0 \leq x < \frac{R_1(R_2 - r_2)}{R_2} \tag{5.24}$$

Combining Eq. (5.23) and Eq. (5.24), for x in the range given by

$$\frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)} \leq x < \frac{R_1(R_2 - r_2)}{R_2} \tag{5.25}$$

It appears that p_2^* in Eq. (5.22) is the real maximizer. But caution is needed here. For the region expressed in Eq. (5.25) to exist, it must satisfy that

$$\begin{aligned} &\frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)} < \frac{R_1 R_2}{R_2 - r_2} \\ \Longleftrightarrow &\frac{r_1}{R_1} + \frac{r_2}{R_2} < 1 \end{aligned} \tag{5.26}$$

which is the condition as shown in Eq. (5.4) in Case (a) of Lemma 5.1. Thus, depending on whether $\frac{r_1}{R_1} + \frac{r_2}{R_2}$ is greater or less than one, we separately derive the stability regions corresponding to the two cases.

1) Case (a) :

When it is true that $\frac{r_1}{R_1} + \frac{r_2}{R_2} < 1$, the region in Eq. (5.25) exists, and the optimal p_2^* expressed in Eq. (5.22) is the real maximizer. By substituting p_2^* into Eq. (5.19), the maximum y is

$$y_{\max} = \frac{\left(\sqrt{R_1 R_2} - \sqrt{(R_2 - r_2)x}\right)^2}{R_1 - r_1} \tag{5.27}$$

Now we consider those values of x that fall into the range of

$$0 \leq x < \frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)} \tag{5.28}$$

We calculate that

$$\frac{dy}{dp_2} > 0, \qquad \text{for } \forall x \in \left[0, \frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)}\right) \tag{5.29}$$

Therefore, y is a strictly increasing function of p_2 for $0 \leq x < \frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)}$, and the optimal p_2^* is equal to 1. The corresponding $y_{\max}$ is

$$y_{\max} = R_2 - \frac{(R_2 - r_2)x}{r_1} \tag{5.30}$$

Then we check for those x such that $x \geq \frac{R_1(R_2 - r_2)}{R_2}$. From Eq. (5.19), we know that x should be no greater than R_1. Hence, we investigate those x which fall into the range given by

$$\frac{R_1(R_2 - r_2)}{R_2} \leq x < R_1 \tag{5.31}$$

Further, Eq. (5.19) implies p_2 must satisfy $p_2 < \frac{R_1 - x}{R_1 - r_1}$. Under these conditions, we can easily check that $\frac{dy}{dp_2}$ is strictly positive in this case. So y strictly increases with p_2; but since p_2 is no greater than $\frac{R_1 - x}{R_1 - r_1}$, the optimal p_2^* that maximizes y is $\frac{R_1 - x}{R_1 - r_1}$. By inserting it into Eq. (5.19), the maximum y becomes

$$y_{\max} = \frac{R_1 r_2}{R_1 - r_1} - \frac{r_2 x}{R_1 - r_1} \tag{5.32}$$

Thus, under the condition $\frac{r_1}{R_1} + \frac{r_2}{R_2} < 1$, the stability region provided by Eq. (5.19) is completely described; its form is expressed by $\mathcal{L}_1 \bigcup \mathcal{L}_2 \bigcup \mathcal{L}_3$ given in Lemma 5.1. By applying the same procedure on the constrained optimization problem suggested by Eq. (5.18), we obtain the same form of the stability region. Hence, for the case $\frac{r_1}{R_1} + \frac{r_2}{R_2} < 1$, the stability region is as given in Case (a) of Lemma 5.1.

2) Case (b) :

When it is true that $\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1$, the subregion represented by $\mathcal{L}_2$ does not exist. In this case, the stability region is bounded by the union of $\mathcal{L}_1$ and $\mathcal{L}_3$, as expressed in Eq. (5.6) and Eq. (5.8) respectively.

As for the region given by $\mathcal{L}_1$, the optimal p_2^* is 1, and the feasible values of x are those for $x < p_1 r_1$, as according to the constraint in Eq. (5.19). Therefore, x is maximized by setting p_1^* to be 1. And the corresponding feasible range for x is $0 \leq x < r_1$. This is described as the region $\mathcal{I}_1$ in Eq. (5.11) of Lemma 5.1.

As for the region given by $\mathcal{L}_3$, it can be achieved from the conditions constrained by Eq. (5.18), after setting p_1^* to be 1 and p_2^* to be 1; and the values of x in this region fall into the range of $[r_1, R_1)$. This is described as the region $\mathcal{I}_2$ in

Eq. (5.12) of Lemma 5.1.

The convexity of the stability region in this case is shown as follows. First, it can be easily calculated that, the two straight lines that bound the subregions $\mathcal{I}_1$ and $\mathcal{I}_2$ intersect at the point (r_1, r_2). Then, if we form a straight line by connecting the two points $(R_1, 0)$ and $(0, R_2)$ at the x-axis and y-axis, and denote this line by L; the intersection point (r_1, r_2) lies above the line L when it holds that $\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1$. Therefore, the stability region characterized by Eq. (5.10) in this case is convex.

By combining the analysis in Case (a) and Case (b), we conclude the proof of Lemma 5.1. ■

Chapter 6

Conclusion

6.1 Summary of Contributions

In this dissertation, we focused on the issues of cooperative communication and rate control in wireless multi-access channels. By synthesizing properties of different layers, our work provided a "layerless" view of the wireless networks.

We first offered an innovative perspective for cooperation by implementing it at the network layer. In a wireless multi-access system with multiple source users, cooperation was exploited amongst users by taking into account physical-layer properties; as such, a packet is delivered to the destination through either a direct link or through cooperative relaying by intermediate source users, depending on the channel quality and instantaneous channel outcomes. We considered conflict-free, work-conserving transmission policies as well as plain TDMA policy at the MAC layer. By modeling bursty traffic arrivals as in a real network, the performance metrics of the stable throughput region and delay were evaluated for both MAC policies. We established that, the stable throughput regions under both classes of cooperative policies are the same, which strictly outer-bound the stable throughput regions achieved without cooperation. Moreover, the optimal policy that minimizes the average delay among the class of all cooperative work-conserving policies was determined. Then, in the case of two users, the closed-form delay expressions were

explicitly derived as well. Our results showed that cooperation substantially reduces delay for both users. A significant implication from our results is that, by relaying the packets of other users, the intermediate users also enjoy the throughput and delay benefit for their own packets. Besides cooperative diversity, we observed that the performance improvement is also attributed to the concentration of packets from disparate source queues into fewer virtual queues. These contributions are presented in Chapter 2 and in [57, 58, 59].

Next, in Chapter 3, we designed and analyzed several advanced network-layer cooperation techniques, with enhancements based on the physical layer. First, in a multi-access channel with a cognitive relay, we incorporated a dynamic decode-and-forward method which allows relaying assistance also during the source's transmission. The enhancement of cognitive relaying with DDF provides more cooperation opportunities which results in faster emptying of the user queues. Further, the cognitive cooperation was supported by an adaptive superposition scheme which allows the relay to simultaneously forward packets from different users. With these techniques that boost the relaying capability, we quantified the extent to which the enhanced cooperation schemes outperform the conventional cooperation scheme in terms of the stable throughput region. Subsequently, the composite effects of multipacket reception and relaying capability in a two-user cooperative channel were investigated. We showed that under certain channel conditions, simultaneous transmissions of different messages from the source user and the relay user should be permitted for yielding higher stable throughput rates. These results are also presented in [60, 61, 62].

Then we returned to the single-packet reception channel model as analyzed in Chapter 2, but considered only two users. For such two-user cooperative multi-access system, we made a thorough analysis on the stability and throughput regions. These two regions are conjectured to be identical in the non-cooperative multi-access system. After cooperation was performed, we revisited the relationship between these two regions, and contributed new and more complicated observations. We established that, the throughput region depends on the scheduling policies at the relay node, and may or may not be equal to the stability region, which was shown to be independent of the scheduling policies. This conclusion holds under both a centralized access scheme and a random access scheme. Furthermore, network coding was applied at the relay node as another form of cooperation; we concluded that network coding provides the same performance as the plain store-and-forward routing in the context of our single-relay cooperative system. These contributions appear in Chapter 4 and in [63, 64].

In Chapter 5, we studied the rate control problem in a multi-access channel. We made the distinction from the traditional networking approach by exploiting the bit-nature of a packet. And, instead of using the Shannon capacity rate at the network layer, the transmission rates are selected from a finite, discrete set such that the rates are achievable in a finite time slot. Under a class of partially distributed transmission policies, we explicitly characterized the stability region in terms of bits/slot. Under a certain channel condition, the transmission policy which activates both users with probability one when they are backlogged was shown to achieve the maximum stability region, and the resulting stability region is convex.

Finally, we studied the minimum delivery time problem where each user is allocated at the beginning some amount of data to be sent to the destination. For any initial queue size vector, we explicitly characterized the optimal rate control policy that empties the queues in the system within the shortest time. This part of work is also described in [65, 66].

6.2 Future Directions

The solid theoretical analysis in this dissertation provides useful insights for better understanding of the communication architecture in wireless networks and its ultimate performance limits. There remain a number of questions for future investigation.

A fundamental issue that naturally arises is the need for a distributed cooperative communication protocol. The proposed cooperation strategy for the multi-access system with N users, implicitly assumes that there exists a centralized controller which activates at most one user in a time slot, such that all other users can overhear the transmission and possibly relay the received packet. However, such centralized controller may not exist, or too costly to implement in a real wireless network. Further, the strategy requires all the users that capture the transmission to send back acknowledgements, upon which the "best" of them can be selected as the relay. This will result in the "feedback implosion" problem. As such, a distributed cooperation policy with feedback suppression mechanism will be of both theoretical and practical interest. The performance in our centralized policy can serve as an

upper bound to evaluate the effectiveness of the distributed policy.

Network coding was briefly addressed in a particular simple network model in Chapter 4. In a multi-access channel with single intended destination and unicast traffic only, we determined that applying network coding between the relayed packets and the source packets at the relay node will not introduce additional performance gains. However, if network coding is performed in different ways, for example, if we also allow inter-session network coding among the same streams of packets at the source user, it is possible to yield enhanced performance. And, if the network is extended to consist of multiple destinations, and hence, includes multicast traffic, network coding is known to outperform plain store-and-forward routing in multicast environment. Cooperative network with network coding has been investigated from the information-theoretic standpoint; we will continue with the "layerless" approach, and evaluate the impact of network coding with cooperation on the performance metrics of stable throughput and delay at the network level.

Throughout this dissertation we consider stationary time-invariant channels. An ultimate goal will be to formulate the problem over non-stationary time-variant channels and dynamic topologies. As such, opportunistic cooperation and network coding schemes that are adapted to the channel conditions will be necessary. The rate control problem can also be combined with the time-varying channel formulation. This extension will require rethinking of the performance objectives, as the mean throughput or mean delay may not exist in non-stationary network environment.

Bibliography

[1] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "On the construction of energy-efficient broadcast and multicast trees in wireless networks," *IEEE INFOCOM*, vol. 2, Tel Aviv, Israel, Mar. 2000, pp. 585–594.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, Mar. 1998.

[3] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part II: implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[6] 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.

[7] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 362–371, Feb. 2004.

[8] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[9] G. Kramer, I. Maric, and R. D. Yates, *Cooperative Communications*, vol. 1, no. 3–4. Foundations and Trends in Networking, Hanover, MA: NOW Publishers Inc., 2006.

[10] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Cooperative relaying with state available at the relay," *IEEE Information Theory Workshop (ITW)*, Porto, Portugal, May 2008, pp. 139–143.

[11] S. Mohajer, D. N. C. Tse, and S. N. Diggavi, "Approximate capacity of a class of Gaussian relay-interference networks," *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June/July 2009, pp. 31–35.

[12] A. K. Sadek, K. J. R. Liu, and A. Ephremides, "Cognitive multiple access via cooperation: protocol design and performance analysis," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3677–3696, Oct. 2007.

[13] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Stable throughput of cognitive radios with and without relaying capability," *IEEE Transactions on Communications*, vol. 55, no. 12, pp. 2351–2360, Dec. 2007.

[14] I. Krikidis, J. N. Laneman, J. S. Thompson, and S. McLaughlin, "Protocol design and throughput analysis for multi-user cognitive cooperative systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4740–4751, Sep. 2009.

[15] A. Ephremides and B. Hajek, "Information theory and communication networks: an unconsummated union," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2416–2434, Oct. 1998.

[16] R. Loynes, "The stability of a queue with non-independent inter-arrival and service times," *Proc. Camb. Philos. Soc.*, vol. 58, pp. 497–520, 1962.

[17] R. R. Rao and A. Ephremides, "On the stability of interacting queues in a multiple-access system," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 918–930, Sep. 1988.

[18] V. Anantharam, "The stability region of the finite-user slotted ALOHA protocol," *IEEE Transactions on Information Theory*, vol. 37, no. 3, pp. 535–540, May 1991.

[19] V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite-user slotted ALOHA with multipacket reception," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2636–2656, July 2005.

[20] J. Luo and A. Ephremides, "On the throughput, capacity and stability regions of random multiple access," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2593–2607, June 2006.

[21] B. Shrader and A. Ephremides, "Random access broadcast: stability and throughput analysis," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2915–2921, Aug. 2007.

[22] 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.

[23] B. Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," *Proc. ACM/IEEE MobiCom*, Boston, MA, Aug. 2000, pp. 243–254.

[24] S. Biswas and R. Morris, "Opportunistic routing in multi-hop wireless networks," *ACM SIGCOMM*, vol. 34, no. 1, pp. 69–74, Jan. 2004.

[25] W. Szpankowski, "Stability conditions for some distributed systems: buffered random access systems," *Advances in Applied Probability*, vol. 26, pp. 498–515, 1994.

[26] D. Bertsekas and R. Gallager, *Data Network*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[27] J. Baras, D. J. Ma, and A. M. Makowski, "K competing queues with geometric service requirements and linear costs: the μc-rule is always optimal," *System & Control Letters*, vol. 6, no. 3, pp. 173–180, Aug. 1985.

[28] C. Buyukkoc, P. Varaiya, and J. Walrand, "The $c\mu$ rule revisited," *Advances in Applied Probability*, vol. 17, no. 1, pp. 237–238, Mar. 1985.

[29] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 466–478, Mar. 1993.

[30] M. Sidi and A. Segall, "Two interfering queues in packet-radio networks," *IEEE Transactions on Communications*, vol. COM-31, no. 1, pp. 123–129, Jan. 1983.

[31] L. Kleinrock, *Queueing Systems, Volume 1: Theory*, New York: Wiley, 1975.

[32] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[33] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[34] A. Murugan, K. Azarian, and H. El Gamal, "Cooperative lattice coding and decoding in half-duplex channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 268–279, Feb. 2007.

[35] K. R. Kumar and G. Caire, "Coding and decoding for the dynamic decode and forward relay protocol," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3186–3205, July 2009.

[36] S. Ghez, S. Verdú, and S. C. Schwartz, "Stability properties of slotted ALOHA with multipacket reception capability," *IEEE Transactions on Automatic Control*, vol. 33, no. 7, pp. 640–649, July 1988.

[37] S. Ghez, S. Verdú, and S. C. Schwartz, "Optimal decentralized control in the random access multipacket channel," *IEEE Transactions on Automatic Control*, vol. 34, no. 11, pp. 1153–1163, Nov. 1989.

[38] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[39] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[40] S. C. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Trans. Networking*, vol. 13, no. 3, pp. 636–647, June 2005.

[41] E. C. van der Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, pp. 120–154, 1971.

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

[43] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Transactions on Information Theory*, vol. 28, no. 3, pp. 536, May 1982.

[44] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1815–1817, May 2005.

[45] Y. H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.

[46] Y. E. Sagduyu and A. Ephremides, "Cross-layer optimization of MAC and network coding in wireless queueing tandem networks," *IEEE Transactions on Information Theory*, vol. 54, no. 2, pp. 554–571, Feb. 2008.

[47] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[48] P. Pakzad, C. Fragouli, and A. Shokrollahi, "Coding schemes for line networks," *IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 1853–1857.

[49] H. Takagi, *Queueing Analysis, Volume 3: Discrete-Time Systems*, Amsterdam: North-Holland, 1993.

[50] E. M. Yeh and A. S. Cohen, "Throughput and delay optimal resource allocation in multiaccess fading channels," *IEEE International Symposium on Information Theory (ISIT)*, Yokohama, Japan, June/July 2003, pp. 245.

[51] E. M. Yeh, "Minimum delay multiaccess communication for general packet length distributions," *the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep./Oct. 2004, pp. 1536–1545.

[52] N. Ehsan and T. Javidi, "Delay optimal transmission policy in a wireless multiaccess channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3745–3751, Aug. 2008.

[53] J. Yang and S. Ulukus, "Delay minimization in multiple access channels," *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June/July 2009, pp. 2366–2370.

[54] A. Eryilmaz, R. Srikant, and J. R. Perkins, "Throughput-optimal scheduling for broadcast channels," *Modelling and Design of Wireless Networks, Proceedings of SPIE*, vol. 4531, July 2001, pp. 70–78.

[55] K. Seong, R. Narasimhan, and J. M. Cioffi, "Queue proportional scheduling via geometric programming in fading broadcast channels," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1593–1602, Aug. 2006.

[56] J. Luo and A. Ephremides, "On rate control of packet transmission over fading channels," *IEEE Communications Letters*, vol. 11, no. 12, pp. 982–984, Dec. 2007.

[57] B. Rong and A. Ephremides, "The effect of cooperation at the network protocol level," *the 4th International Wireless Internet Conference (WICON)*, Maui, Hawaii, Nov. 2008.

[58] B. Rong and A. Ephremides, "Protocol-level cooperation in wireless networks: stable throughput and delay analysis," *the 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, Seoul, Korea, June 2009, pp. 1–10.

[59] B. Rong and A. Ephremides, "Cooperative access in wireless networks: stable throughput and delay," *IEEE Transactions on Information Theory*, submitted.

[60] B. Rong, I. Krikidis, and A. Ephremides, "Network-level cooperation with enhancements based on the physical layer," *IEEE Information Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5.

[61] I. Krikidis, B. Rong, and A. Ephremides, "Cooperation for multiple-access channel via dynamic decode-and-forward," *IEEE Transactions on Information Theory*, submitted.

[62] B. Rong and A. Ephremides, "On opportunistic cooperation for improving the stability region with multipacket reception," *the 3rd Intl. Conference on Network Control and Optimization (NET-COOP)*, Eindhoven, The Netherlands, Nov. 2009, pp. 45–59.

[63] B. Rong and A. Ephremides, "Cooperation above the physical layer: the case of a simple network," *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June/July 2009, pp. 1789–1793.

[64] B. Rong and A. Ephremides, "On stability and throughput for multiple access with cooperation," *IEEE Transactions on Information Theory*, submitted.

[65] B. Rong and A. Ephremides, "Stable throughput, rate control, and delay in multi-access channels," *the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, Avignon, France, May/June 2010, pp. 204–211.

[66] B. Rong and A. Ephremides, "Layerless optimal rate control for stability and delay in wireless multi-access channels," *Performance Evaluation*, submitted.