Network Protocols for Ad-hoc Networks with Smart Antennas

A Thesis
Presented to
The Academic Faculty

by

Karthikeyan Sundaresan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
December 2006

Network Protocols for Ad-hoc Networks with Smart Antennas

Approved by:

Dr. Raghupathy Sivakumar, Advisor
School of Eletrical and Computer Engineering
Georgia Institute of Technology

Dr. Chuanyi Ji
School of Eletrical and Computer Engineering
Georgia Institute of Technology

Dr. Mary Ann Ingram
School of Eletrical and Computer Engineering
Georgia Institute of Technology

Dr. Thomas Pratt
Georgia Tech Research Institute

Dr. Douglas Blough
School of Eletrical and Computer Engineering
Georgia Institute of Technology

Date Approved: July 21, 2006

To my

Dad and Mom,

for their relentless support and selfless care!

ACKNOWLEDGEMENTS

Hmmm...probably one of the toughest chapters to write! The bottom line is that my entire journey through Ph.D has been a fairly smooth one. Of course, I do not think I will be able to do complete justice in thanking everyone who has been responsible for it.

Lets start with the Almighty! He has been simply phenomenal. He helped me remain focused for a not-so-short duration of five years. He also made sure that there weren't too many obstacles and hiccups along the way.

My parents have been really cool! In the past five years, not once have they asked me *how long is it going to take for me to get done* - a question that haunts every Ph.D candidate! They have always given me a great deal of freedom in making decisions in my career and supported me throughout the way.

During the past five years, I have learnt a lot from my advisor - not just on the technical front but also on how to mould oneself well in order to have a reasonable career in research. He has been one of my inspirations to carry on with a Ph.D after my masters. I also need to thank Dr. Ingram, one of the nicest people I have collaborated with, for all the fruitful discussions we have had in the projects and for allowing me to sneak in without an appointment and get questions clarified.

Now, we come to the more interesting part - my friends, in the absence of whom my whole Ph.D stint would have been a disaster! The numerous trips that I got to make with my friends back from undergrad during my Ph.D kept the spirit alive, especially when things were starting to get monotonous. Of course, the long hours of fruitless (a.k.a *vetti*) phone conversations with Girish and Chandru cannot be forgotten here. Ram, Aru and Sridhar have been a great source of company to me for watching movies and engaging in sports during my stay at Tech. I would also need to thank my colleagues in the lab, who have been a great deal of help in providing me valuable feedback on my research.

Last, but not the least, I need to thank myself for not messing up with my Ph.D! :-)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Multi-hop wireless networks or ad-hoc networks represent a class of wireless networks that operate without the need for an infrastructure. These networks face several limiting characteristics that make it difficult to support a multitude of applications. It is in this context that smart antennas find significant applications in these networks. Smart antennas represent a more sophisticated physical layer antenna technology with the ability to alleviate most of the limitations faced by ad-hoc networks. The focus of my research is thus to investigate the use of smart antennas in ad-hoc networks and hence efficiently design network protocols that best leverage their capabilities in communication. There are two parts to the proposed objective of designing efficient network protocols that pertain to the nature of the smart antenna network considered, namely, homogeneous and heterogeneous smart antenna networks.

Unlike heterogeneous smart antenna networks, where different devices in the network employ different antenna technologies, homogeneous smart antenna networks consist of devices employing the same antenna technology. Further, in homogeneous smart antenna networks, different antenna technologies operating in different strategies tend to perform the best in different network architectures, conditions and application requirements. However, it is a cumbersome task to go about designing communication protocols for each of these antenna technologies operating in different strategies. This motivates the need for developing a *unified* framework for designing efficient communication (medium access control and routing) protocols for homogeneous smart antenna networks in general. With the objective of designing such a unified framework, we start by designing efficient MAC and routing protocols for the most sophisticated of the smart antenna technologies, namely multiple-input multiple-output (MIMO) links. The capabilities of MIMO links form a super-set of those possible with other antenna technologies. Hence, the insights gained from the design of communication protocols for MIMO links are then used to develop unified MAC and

routing frameworks for smart antennas in general.

For heterogeneous smart antenna networks, we develop theoretical performance bounds by studying the impact of increasing degree of heterogeneity on network throughput performance. Given that the antenna technologies are already unified in the network, unified solutions are not required. However, we do develop efficient MAC and routing protocols to best leverage the available heterogeneous capabilities present in the network. We also design efficient cooperation strategies that will further help the communication protocols in exploiting the available heterogeneous capabilities in the network to the best possible extent.

CHAPTER I

INTRODUCTION

Ad-hoc networks are multi-hop wireless networks that operate without the need for infrastructure support. This made them extremely attractive for operation in military and disaster-relief environments. Recently, they have also gained significant interest in the form of commercial mesh networks. The limited battery resource capabilities of the nodes in these networks, restricts the communication range possible, thereby resulting in multi-hop nature of the flows in the network. Thus, every node not only acts as a source of data but also as a forwarder for packets generated by other nodes. The lack of a central coordinating entity as well as the multi-hop nature of the flows, poses several challenges for efficient communication in these networks. Some of the key limitations faced by these networks are: (i) low data rates on link as well as (end-end) flow basis restrict the usage of bandwidth intensive applications; (ii) fluctuating channel bandwidth affects link reliability and consequently quality of service; (ii) network partitions resulting from node mobility and sparse network density disrupt communication; (iv) power efficiency is critical for several mobile applications; and (v) simultaneous transmissions resulting from decentralized operations causes interference limited environments.

It is in this specific context that smart antennas find significant applications in ad-hoc networks. Omni-directional antennas have dominated the antenna technology used for communication in ad-hoc networks thus far. Of late, the advent of smart antennas has revolutionized the antenna technology. Smart antennas refer to a sophisticated physical layer antenna technology coupled with significant signal processing. These antennas provide significant performance improvements over omni-directional antennas. Their link and interference suppression gains can be used to alleviate some or all of the limitations faced by ad-hoc networks, thereby making them a valuable tool for operation in these networks. While these smart antennas have been significantly researched at the physical layer of the

protocol stack, it is not completely clear what advantages they can provide to the higher layers of the protocol stack and consequently to the end user in ad-hoc networks. Thus, the high level theme of my research is to (i) identify the various advantages of smart antennas that can be leveraged by the higher layers of the protocol stack, and (ii) design efficient network (medium access control [MAC] and routing) protocols to leverage the identified advantages. In trying to contribute towards the above theme, we decompose the problem domain into two components of *homogeneous* and *heterogeneous smart antenna networks*.

Homogeneous smart antenna networks refer to ad-hoc networks, where all devices in the network employ the same kind of smart antenna technology. Every smart antenna technology is capable of exploiting its different gains for different performance objectives through different strategies. However, no single antenna technology operating in a specific strategy performs the best across all possible network conditions and application requirements. Hence, different smart antenna technologies and strategies are well-suited for different applications and network architectures and conditions. However, it is a cumbersome task to go about designing communication protocols for each of these antenna technologies operating in different strategies. This motivates the need for developing a *unified* framework for designing efficient communication (medium access control and routing) protocols for smart antennas in general.

With the objective of designing such a unified framework, we begin by designing efficient MAC and routing protocols for the most sophisticated of the smart antenna technologies, namely multiple-input multiple-output (MIMO) links. The capabilities of MIMO links forms a super-set of those possible with other antenna technologies. Hence, the insights gained from the design of communication protocols for MIMO links is then used to develop unified MAC and routing frameworks for smart antennas in general.

- We start with the design of a medium access control protocol for ad-hoc networks with MIMO links, where we consider the following key questions: (i) What are the key optimization considerations that should be incorporated in the design of a MAC protocol designed for the target environment? (ii) How can the versatile properties of MIMO links be leveraged to effectively realize a practical distributed MAC protocol

with the optimal design? In the process of answering these questions, we use both results from related research at the physical layer, and detailed arguments to identify several optimization considerations. Based on these considerations, we first present a centralized MAC scheme, and then a distributed MAC protocol called SCMA (Stream-Controlled Medium Access) for ad-hoc networks with MIMO links.

- We then investigate and explore the various capabilities of MIMO links from the perspective of routing layer protocols. We identify two fundamental capabilities of MIMO links, namely spatial multiplexing and diversity that can be exploited by the routing layer protocols in their operations. However, since these two capabilities cannot be fully leveraged at the same time, we investigate and study the relative trade-offs between spatial multiplexing and diversity, in order to determine the optimal strategy of operation for improving the aggregate network throughput. A routing protocol called MIR (MIMO Routing) is then proposed with components built on the insights gained from the study and is capable of transparently adapting between the strategies based on perceived network conditions.

- Capitalizing on the insights gained from the design of MAC and routing protocols with MIMO links, we then proceed to design *unified, distributed* MAC and routing frameworks for use with smart antennas in general, including omni-directional antennas, switched-beam antennas, adaptive array antennas, and MIMO links. Specifically, (i) For the different antenna technologies, we identify the physical layer capabilities, their relevance to MAC/routing layer design, and MAC/routing layer design considerations specific to the physical layer capabilities, and capture them through a unified representation; (ii) We provide a unified formulation of the tractable MAC problem and derive unified centralized algorithms based on the problem formulation; and (iii) We then design unified, distributed MAC and routing algorithms that are of practical interest and can be deployed.

- Finally, using the algorithms derived from the unified framework, we substantiate our

earlier statement that different smart antenna technologies and their respective strategies tend to perform the best under different network conditions. For this purpose, (i) we comprehensively evaluate the different antenna technologies, and their respective strategies with respect to the different network parameters, and identify performance trends; (ii) we draw inferences on which strategy proves to be optimal for the different settings for each technology; and (iii) we then identify the technology that can deliver the best performance for a given network setting. The comprehensive study helps us understand the relative performance trade-offs of the different technologies, and identify key insights into how the technologies compare under varying conditions.

Heterogeneous smart antenna networks refer to ad-hoc networks, where different devices in the network could potentially employ different antenna technologies varying between omni, switched beam, adaptive arrays and MIMO links. While homogeneous networks form an essential first step towards understanding the true performance gains of the specific antenna technologies, there are several motivating reasons why it is important to consider a heterogeneous wireless antenna network, such as (i) economic feasibility (incremental deployments), (ii) commercialization of mesh networks, (iii) digital battlefields, (iv) zero-configuration community networks, etc. Such heterogeneous smart antenna networks present a more challenging and practical class of networks.

Unlike homogeneous networks, we do not need to develop unified solutions for heterogeneous networks since all antenna technologies are already unified in the network to start with. However, establishing efficient communication in such heterogeneous networks is an especially challenging task since one must design efficient communication strategies that will help learn, realize and exploit the available heterogeneous capabilities in the network to the best possible extent. We achieve this objective by first theoretically studying the impact of heterogeneity on the throughput performance of these networks. The insights gained from the study are then used to design efficient communication protocols and strategies that will help exploit and realize the available heterogeneous capabilities in the network.

- We begin by studying the theoretical throughput performance in heterogeneous networks. As a vital first step, we first identify the link layer capabilities when nodes with heterogeneous antenna technologies interact and incorporate this in the modeling of the throughput capacity. We study what the ideal performance bounds are in these networks. Perhaps more interestingly, through the bounds and feasible constructions, we provide insights into what percentage of the nodes in an ad-hoc network have to be "smart" in order to achieve a certain level of network performance. The study also reveals several interesting inferences that are of practical use to network practitioners. Further, since the homogeneous networks can be viewed as a special case of the heterogeneous networks, the performance bounds apply to homogeneous networks as well.

- Using the insights gained from the theoretical study, we design efficient MAC and routing protocols for heterogeneous smart antenna networks. One of the main components of the heterogeneity-aware smart MAC protocol (HSMA) is to enable the two nodes of a communicating link to realize the complete smart antenna capabilities of the link that can be formed between them. HSMA also incorporates a distributed contention resolution mechanism that takes into account the fact that different heterogeneous links provide different gains and access the medium (resources) differently, in trying to ensure a desired fairness model. The heterogeneity-aware smart routing protocol (HSR) is designed to be a reactive routing protocol and leverages the heterogeneous link capabilities available in the network in its routing mechanisms to provide improved network performance.

- Finally, we shift our focus towards designing efficient communication *strategies* that will aid the HSMA/HSR protocols in *further* exploiting the heterogeneous capabilities to the best possible extent. In this regard, we motivate the need for a simple form of node cooperation called *retransmit diversity*. While such a simple form of cooperation cannot bring significant benefits to *homogeneous* omni-directional antenna networks, we show that they can bring great gains to *heterogeneous* smart antenna networks.

However, the benefits are also accompanied by a fundamental trade-off between exploiting smart antenna gain and cooperation gain, that undermines the ability of these heterogeneous networks to leverage node cooperation to their maximum potential. To address this tradeoff, we then present an adaptive cooperation mechanism and also an accompanying MAC protocol to realize the mechanism.

In the following two chapters, we outline the motivation for the research problem along with the related research work in literature, followed by the specific contributions made in this thesis along with its organization.

CHAPTER II

BACKGROUND AND MOTIVATION

Smart antennas possess sophisticated signal processing capabilities that help them deliver significant performance benefits such as increased spectral efficiencies, reduced power consumption, interference suppression, increased communication reliability, better connectivity, etc., over conventional omni-directional antennas [32, 3, 45, 6, 54]. Not surprisingly, due to the above advantages, the use of such smart antennas in wireless networks has gained significant attention over the last few years. The term "smart antennas", in reality, represents a broad variety of antennas that consist of multiple-element arrays (MEAs) in hardware but differ significantly in their performance and transceiver complexity [21, 33, 42]. The different antenna technologies include switched-beam antennas, steered-beam antennas, adaptive array antennas, and multiple-input-multiple-output (MIMO) links.

In addition to the different technologies, smart antennas, even for a given technology, can be operated using different strategies to achieve different performance objectives by exploiting the available gains in different ways: (i) to increase the capacity of the link either by transmitting independent data streams from the different antenna elements [78, 16, 34] or by performing adaptive modulation while still maintaining the required level of signal-to-noise ratio (SNR) on the link [3, 6], (ii) to increase the transmission range to reduce the number of hops for the flow and to increase connectivity [73, 74, 14, 10], (iii) to increase the SNR or reduce the variance in SNR, thereby increasing the link reliability [52, 19, 1], and (iv) to perform power control and reduce the power consumption while still retaining the required SNR on the link [38, 59].

While smart antennas have been significantly researched at the physical layer [3, 76, 19, 75] and hence their applicability and performance benefits in the context of one-hop wireless cellular networks [53, 28, 36, 50, 79, 14] and to some extent wireless local area networks [46, 14, 69] has been well established, very little is understood in terms of the performance

of smart antennas in ad-hoc networks. Although several recent works have focused on network protocols (especially at the medium access control layer) in ad-hoc networks with smart antennas, the technology considered has been predominantly restricted to that of simple switched beam or directional antennas [10, 44, 15, 29, 4, 39, 51, 67] with some recent interest in adaptive arrays [35, 57, 58] and multiple-input multiple-output (MIMO) links [66, 23, 65]. This is because, switched beam antennas are the simplest of the smart antennas in terms of their sophistication and capabilities, thereby facilitating easy understanding of theory [41, 77, 60], protocol development [27, 48, 9, 26, 38, 51, 68] and implementation [70]. However, their simplicity also limits their performance and application to specific environments such as line of sight environments, unlike the more sophisticated adaptive arrays and MIMO links which can operate well even in multi-path environments. Hence, it becomes necessary to consider the applicability of the more sophisticated classes of smart antennas also, given that typical ad-hoc network settings such as office complex indoors, urban outdoors, etc. involve significant multi-path. Further, considering the literature in related research, it can be seen that most of the proposed protocols for switched beam antennas do not completely leverage their true potential. While switched beam antennas are capable of directing their energy in specific directions, thereby reducing the amount of interference in other directions and consequently increasing spatial reuse [41, 77], they are also capable of providing a directional gain on the link [3]. The directional gain can in turn be used to increase the communication range, rate of transmission, link reliability or even to reduce the transmission power. Most of the related works in literature, exploit only the directionality [44, 4, 39, 51, 27] and hence the increased spatial reuse property of switched beam antennas, but fail to efficiently exploit the directional gain on the link. Thus, there is a significant lack of complete understanding of the advantages of smart antennas to the higher layers of the network protocol stack and consequently the inability to exploit the true potential of smart antennas to improve network performance.

This in turn forms the main motivation for my thesis. The goal of my thesis is thus to first identify the true physical layer capabilities and gains provided by each the smart

antenna technologies. Based on the capabilities, the next step is to identify the optimization considerations that must be incorporated in the design of network protocols so as to completely leverage all the capabilities. The network protocols focused in this research are restricted to medium access control and routing protocols. This is because, when the current physical layer technology is replaced with a more sophisticated one, in order for the end-user application to see the benefits of the new technology, the higher layers of the protocol stack must be appropriately modified. But the amount of contribution provided by each of the layers in realizing the true potential of the antenna technology, decreases as we go up the protocol stack. Hence, MAC layer, being just above the physical layer, is the most important layer that needs adaptation in helping realize the capabilities, followed by the routing layer. However, as we go on to the transport layer, there is little impact due to a change in the antenna technology, thereby not requiring significant adaptation to the transport layer. We support the above objective with both algorithmic and theoretical foundations. The eventual goal of the thesis is to provide a comprehensive framework of theory, algorithms and protocols for communication in ad-hoc networks with smart antennas. The thesis will also provide a platform to understand the relative benefits of the different smart antenna technologies, which will help not only network designers in the appropriate design of smart antenna networks for real applications, but also researchers developing network protocols for the environment.

CHAPTER III

CONTRIBUTIONS AND ORGANIZATION

The focus of my research is to investigate the use of smart antennas in ad-hoc networks and efficiently design network protocols that best leverage their capabilities in communication. In the process of achieving this objective, my doctoral research and consequently my thesis makes the following principal contributions towards homogeneous and heterogeneous smart antenna networks:

1. **Homogeneous Smart Antenna Networks**

    - Design of efficient network (medium access control and routing) protocols with MIMO (multiple-input multiple-output) links;
    - Development of *unified* MAC and routing frameworks for smart antennas in general;
    - Design rules identifying the optimal antenna technology and strategy of operation for a given network architecture, condition and application requirements.

2. **Heterogeneous Smart Antenna Networks**

    - Theoretical analysis capturing the impact of increasing degree of antenna heterogeneity on throughput performance;
    - Design of efficient network (MAC and routing) protocols to best leverage the available heterogeneous capabilities in the network;
    - Design of simple but efficient cooperation strategies to further aid the communication protocols in the exploitation of network heterogeneity.

The above contributions have been categorized under the problem domains of homogeneous and heterogeneous smart antenna networks, which in turn form the two mains parts of my thesis. The rest of the thesis is organized as follows.

In homogeneous smart antenna networks, we first consider the most sophisticated of the smart antenna technologies, namely multiple-input multiple-output (MIMO) links. The capabilities and gains of MIMO links namely array, spatial multiplexing, diversity, and interference suppression gains, form a super-set of those possible by other smart antenna technologies like switched beam (directional and interference suppression gains) and adaptive arrays (array, diversity and interference suppression gains). Hence, identifying optimization considerations for the design of efficient protocols for MIMO links automatically helps us design efficient protocols for other antenna technologies as well. In this regard, we design a fair stream-controlled medium access control (SCMA) protocol in Chapter 4 and a MIMO routing (MIR) protocol in Chapter 5 that exploit the benefits of MIMO links to the best possible extent at the MAC and routing layers respectively.

Identifying the optimization considerations with respect to MIMO links helps us understand the relative benefits of other antenna technologies as well. This in turn gives us the potential to look at smart antennas as a whole and build theoretical foundations to better capture their communication capabilities in ad-hoc networks. Here, we design 'unified" frameworks that formulate and solve the MAC and routing problems for smart antennas in general in Chapter 6. The unified frameworks can be used to derive MAC and routing protocols for different kinds of smart antenna technologies like switched beam, adaptive arrays and MIMO links. The frameworks are then used to evaluate the different smart antenna technologies and their corresponding strategies in different network conditions in Chapter 7. The outcome of the evaluation helps us identify design rules indicating the optimal smart antenna technology and strategy to be used in specific network conditions depending on application requirements.

Building on the strength of the insights and inferences gained from protocol design in homogeneous environments, we then move on to consider ad-hoc networks with nodes having heterogeneous antenna technologies. Given the practical relevance of such heterogeneous smart antenna networks, we study the impact of increasing degree of antenna heterogeneity on the throughput capacity of the network and present results and inferences obtained from theoretical analysis in Chapter 8. We also design efficient MAC and routing protocols

that help exploit the heterogeneous link capabilities available in the network in Chapter 9. Finally, we design simple yet effective node-cooperation strategies in Chapter 10 to help the communication protocols further exploit the heterogeneous capabilities to the maximum extent.

Conclusions of the thesis as well as open research directions that require further investigation are presented in Section 11.

CHAPTER IV

MEDIUM ACCESS CONTROL IN AD-HOC NETWORKS WITH MIMO LINKS

4.1 Overview

This chapter presents the details of the proposed medium access control protocol for ad-hoc networks with multiple-input multiple-output (MIMO) links, referred to as the *stream-controlled medium access* (SCMA) protocol. The chapter begins with an outline of the physical layer characteristics of MIMO links, followed by a discussion on the optimization considerations needed to leverage the potential gains of MIMO links. A centralized protocol is then proposed to evaluate the ideal performance of a MAC protocol with MIMO links and in turn used to design a practical distributed protocol. The performance of the distributed protocol is then evaluated against the centralized version under varied network settings.

4.2 MIMO Background

4.2.0.1 Relevant PHY Layer Characteristics

A MIMO link employs digital adaptive arrays (DAAs) at both ends of the link, as shown in Figure 1. Such a link can provide three types of gain: array gain, diversity gain, and spatial multiplexing gain. Array and diversity gains primarily provide range extension, while spatial multiplexing gain primarily provides higher data rates.

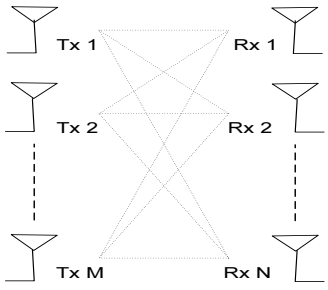


Figure 1: MIMO Illustration

Array gain can occur in an array receiver when the desired signal parts of each antenna output add coherently (coherent combining) and the noise parts add incoherently. Array gain makes the average signal-noise ratio (SNR) at the output of the combiner (the average with respect to random multi-path fading) N times greater than the average SNR at any one antenna element, where N is the number of antenna elements in the array. Array gain occurs even in the absence of multi-path.

Diversity gain relates to the reduction in the variance of the SNR at the output of the combiner, relative to the variance of the SNR prior to combining. The reduction in variance depends on the diversity order, which in turn depends on the degree to which the multi-path fading on the different antenna elements is uncorrelated. The maximum diversity order afforded by a MIMO link with M transmit antennas and N receive antennas is MN.

Since a wireless link is usually designed to have certain small probability that the SNR drops below some threshold value, both array and diversity gains contribute to range extension. In the presence of some channel state information, the factor of range extension d_f can approximately be given by [3],

$$d_f \approx \{(\sqrt{M} + \sqrt{N})^2\}^{\frac{1}{p}} \tag{1}$$

where p is the path loss component. While array gain continues to grow as more antennas are added, diversity gain tends to diminish, like the reduction in the variance of a sample mean. However, for a transmit array to provide either array or diversity gain, the data streams transmitted from the different antenna elements must be dependent.

In the presence of multi-path or rich scattering, the MIMO link can provide spatial multiplexing gain. This gain is defined as the asymptotic increase in the capacity of the link for every 3 dB increase in SNR [78, 8, 5]. This gain can be achieved when the transmit array transmits multiple independent streams of data. In the simplest configuration, the incoming data is demultiplexed into M streams, and each stream is transmitted out of a different antenna with equal power, at the same frequency, same modulation format, and in the same time slot. In fact, this approach is optimal in terms of capacity when the transmitter array has no channel state information (CSI) [17]; hence the approach is often

referred to as open-loop MIMO (OL-MIMO). At the receiver array, each antenna receives a superposition of all of the transmitted data streams. However, each stream generally has a different "spatial signature", and these differences are exploited by the receiver signal processor to separate the streams. When M = N = k, the capacity is given by the following equation [19],

$$C \approx k \log_2(1 + \rho) \tag{2}$$

where ρ represents the average SNR at any one receive antenna.

On the other hand, when the multiple antennas are used only for array and diversity gain, the asymptotic capacity is

$$C \approx log_2(1 + \rho') \tag{3}$$

where ρ' is a random SNR with a mean that increases only linearly with the array gain and a variance that decreases with the diversity order. Therefore, the capacity grows linearly with k with spatial multiplexing, but only logarithmically with array and diversity gain. *Thus, the focus in this work is to exploit the spatial multiplexing gain to increase the capacity of the system.* However, the range extension possible through the diversity gain can be intelligently leveraged to address some key problems at the MAC layer.

Another degree of classification of MIMO links is based on whether or not the transmitter uses CSI with respect to the receiver. If CSI is used at the transmitter, the MIMO link is referred to as closed-loop MIMO (CL-MIMO). CL-MIMO is known to outperform OL-MIMO under conditions of low SNR, correlated fading, and interference [7, 13, 12, 56]. Since we are considering improvements to network throughput where interference between MIMO links is allowed, we consider CL-MIMO. While we assume that OL-MIMO is used for MAC layer control packet exchanges (e.g. Request-to-send and Clear-to-send messages in the CSMA/CA framework) due to the absence of CSI at the transmitter initially, we leverage these packet exchanges to exchange the CSI, thereby enabling the use of CL-MIMO for the actual data packet transmission.

4.2.0.2 Abstraction

We use the following abstraction and assumptions for the physical (PHY) layer in our work. We assume that all the nodes in the network employ DAAs with the same number of elements (M = N = k), and operate on a single channel. The signals sent on the different modes of the channel represent the different streams transmitted. The total number of elements at a node correspond to the total available resources or *degrees of freedom* (DOFs) at the node. We assume that each receiver array treats all interference as noise. This is consistent with practical linear multi-user detection schemes such as minimum mean squared error (MMSE) [19].In a MIMO link, a receiver can isolate and decode all the incoming streams successfully as long as the total number of incoming streams (n) is less than or equal to its DOFs ($n \leq k$). This is especially true in flat-fading wireless channels. On the other hand, if the incoming streams overwhelm the DOFs at the receiver ($n > k$), it will not be possible to decode any of the desired signal streams, if the excess ($n - k$) streams degrade the k streams below their receive threshold. However, if the strength of the excess (say, interfering) streams is far weaker than that of the desired (k) streams such that the desired streams can still be received with atleast the receive threshold, then it may be possible to decode the desired streams [12]. In terms of transmission, a transmitter can use up all the available DOFs (taking into account DOFs available at nodes in the neighborhood after they have suppressed any interference) to spatially multiplex signals.

4.3 Motivation

Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) is the de-facto MAC protocol considered for use in ad-hoc network environments. Interestingly, a simple extension of CSMA/CA for MIMO links can be realized that can provide a k fold improvement in throughput performance through spatial multiplexing (k is the number of elements at each node), compared to a pure omni-directional environment.

We refer to the simple extension to CSMA/CA as CSMA/CA(k). Essentially, CSMA/CA(k) works in the exact same fashion as CSMA/CA except that *all* transmissions are performed using k streams to tap the spatial multiplexing gain. Such a protocol, when compared

to default CSMA/CA operating in the same network topology, but with omni-directional antennas, will achieve k times the throughput performance as the latter. While a k fold improvement is indeed quite attractive, the question that we answer in this subsection is: *Is it possible for a more intelligent MAC scheme to realize better performance?* More importantly, does the degree of improvement justify the development of a new MAC scheme, instead of using a protocol such as CSMA/CA(k). We argue that the answers to both the questions is *yes*. We discuss why CSMA/CA(k) does not truly leverage the capabilities of MIMO using simple toy topologies. More importantly, we show in Section 10.5 that the difference in performance between the "unaware" CSMA/CA(k) scheme and MIMO "aware" MAC scheme increases with increasing number of antenna elements.

In the rest of the subsection, we outline the key optimization considerations that need to be accounted for in the MAC design in order to effectively utilize the capabilities of MIMO.

4.3.0.3 Stream Control Gains

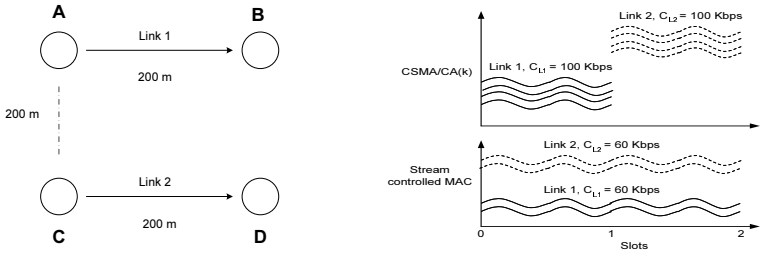


Figure 2: Stream Control Topology

When a link is allowed to use only a subset of the maximum possible number of streams (say m out of k), *it can distribute its transmit power over just the m strongest channel modes (streams).* Thus, when compared to two interfering links operating using time-division multiple access (TDMA) at the maximum number of streams k, letting the links operate simultaneously but with $\frac{k}{2}$ streams will result in improving the overall utilization in the network. We term the gains achievable through such simultaneous operation of interfering but stream controlled links as *stream control gains.*

In the simple toy topology shown in Figure 2 where the nodes have four-element DAAs each, consider transmissions from node A to node B, and from node C to node D. CSMA/CA(k)

allows only one transmission to take place in a given time slot but the transmission proceeds with all the four streams. On the other hand, consider a stream-controlled MIMO MAC where the two transmissions proceed simultaneously but the number of streams transmitted by each node is optimized (in this case to two streams) to give the maximum overall network throughput (Figure 2). For this simple two link topology, an improvement of 20% can be obtained in capacity over that of a TDMA scheme [13]. In general, as the number of mutually interfering links (l) increases, the subset of streams used by each of the links decreases ($\frac{k}{l}$), which in turn increases the gain obtained from performing stream control.

In a CL-MIMO system, there is a one-to-one mapping between streams and transmit array weight vectors; with the help of CSI each antenna element transmits a super-position of all (weighted) data streams. In the receiving node, there will be a different array weight vector for each stream. Therefore, there will be a channel gain for each stream, which is the stream gain. These *stream gains are not equal*, and for moderate-to-low SNR, they can *have quite large disparities* even in the presence of interference [13]. This in turn motivates the need for performing stream control in order to increase the network utilization, wherein the best possible channel modes (two in the above example) are selected for transmission. In the above example, the normalized gains of 1, 0.9, 0.7, and 0.6 were assumed on the four streams. Hence, during stream control, the two best streams with gains of 1 and 0.9 were chosen by the two links to provide an improvement of around 20%. The improvement of 20% is quite conservative for equal powered transmitters in the presence of low correlation between the desired and interfering MIMO streams based on the studies in [24]. In fact, [24] shows that upto a 65% performance improvement over a TDMA scheme can be obtained from performing stream control for measured indoor channels in the presence of low correlation. We summarize the above discussions by the following observation:

OBSERVATION 1: Multiple interfering links operating simultaneously using stream control achieve better overall throughput performance when compared to a scenario in which they operate using TDMA and k streams each.

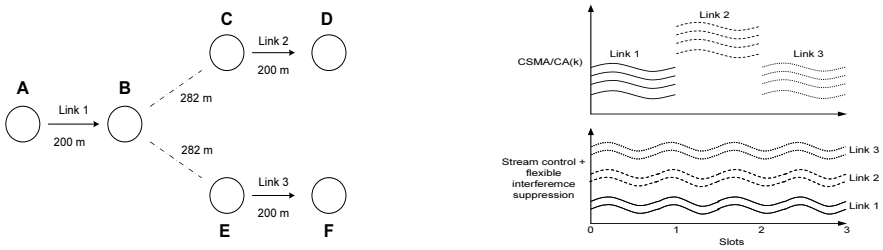


Figure 3: Partial Interference Suppression Topology

4.3.0.4 Partial Interference Suppression

When two interfering links are positioned such that the interference signals traverse a longer distance (R) than the desired signals (D) with $1 \leq \frac{R}{D} \leq 2$, the improvement of MIMO with stream control over CSMA/CA(k) can increase significantly. This is due to the flexible interference suppression capabilities provided by DAAs. The amount of resources that need to be sacrificed (expended) at a node to suppress an interference depends on the actual strength of the interference and the degree of spatial correlation [19, 13].

This is better illustrated through the toy topology in Figure 3, where the nodes have a four-element DAA each. Consider three transmissions, from node A to node B, from node C to node D, and from node E to node F. CSMA/CA(k) allows only one of the transmissions to take place in any time slot but on all four streams. Consider a stream-controlled MIMO MAC where the nodes operate with two streams each. Now the transmitters C and E are outside the receive range but within the carrier-sensing range of receiver B. Hence, the number of DOFs required to suppress interference at node B in this case might only be a fraction of the total number of interfering streams, which in turn depends on the strength of the interfering streams and their spatial correlation. Assuming this fraction to be half, this allows the three transmissions to take place simultaneously on two streams each (Figure 3). Hence, fewer resources are required to suppress interference when the interfering signals are from far away than when they are from close by. This, in turn results in more of the resources at a node being available for improving the performance of the desired transmissions/receptions.

But it must be noted that additional resources can be made available at any node due

to flexible interference suppression, only as long as the node operates on a subset of the maximum number of streams possible. This is because, if the node operates on all available streams then it will have to expend all its resources to receive desired signal streams from its intended receiver. Hence, no additional resources will be made available in this case. Thus, *the gain of flexible interference suppression can be obtained only in conjunction with stream control*. This explains why CSMA/CA(k) cannot exploit the advantage of flexible interference suppression, even if its mechanism of silencing nodes in the two hop neighborhood of any transmission, is extended to incorporate flexible interference suppression.

In the above example, the average number of streams/slot is six for a stream-controlled MAC, while it is only four for CSMA/CA(k). Also it has been shown in related work that the performance gain over a TDMA based approach, obtained by employing flexible interference suppression can be as high as 65% for a simple two link topology [24]. Based on the above discussions, we make our second observation:

OBSERVATION 2: The flexible interference suppression capabilities of DAAs helps create additional resources at a node that can be used in conjunction with stream control for additional transmissions (receptions) to provide additional gain.

4.3.0.5 Receiver Overloading

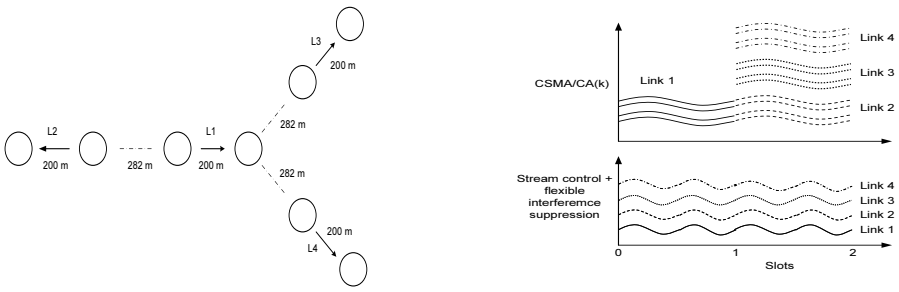


Figure 4: Receiver Overloading Topology

While the above two factors directly help a MIMO aware MAC achieve improved performance over CSMA/CA(k), there exists one facet of MIMO that can potentially *degrade* its performance when compared to CSMA/CA(k). In CSMA/CA(k), since there can be only one active transmitter in any *contention region*, the other passive receivers in the same

region can be *overloaded* with more streams than they can receive. This paves the way for spatial reuse. Hence, if a passive receiver belongs to more than one otherwise non-overlapping contention regions, then there can be an active transmitter in each of those contention regions.

On the other hand in a MIMO MAC employing stream control, all the transmitters within a contention region use the best subset of their streams such that no receiver in the region is overloaded. But if any of the receiver nodes also belong to other contention regions, then this prevents the nodes of those other contention regions from transmitting since this will overload the active receiver. This in turn reduces the advantage of spatial reuse and could potentially degrade performance.

For example consider the simple topology in Figure 4. There are four links, namely L1, L2, L3 and L4. The link L1 interferes with L2, L3, and L4, but the latter three links do not interfere with each other. If four element DAAs are used, CSMA/CA(k) can schedule L1 during one slot with four streams, and L2, L3, and L4, during the next slot with four streams each. Thus, the average throughput in the network in terms of streams per slot is eight. However, if a stream controlled MIMO MAC is used, all four links will operate exactly with one stream each, as any more streams will overload the receiver of link L1. Thus, the average throughput obtained is just four which is smaller than that of CSMA/CA(k) (Figure 4). Further, this degradation would increase as the number of passive receivers (belonging to multiple contention regions) increases, and also as the number of contention regions that a passive receiver belongs to increases. We attribute the above advantage of CSMA/CA(k) to its ability to perform *receiver overloading*, i.e. a *passive* receiver can be exposed to more than the maximum number of interfering streams. Thus, our final observation is:

OBSERVATION 3: The inability to overload a passive receiver because of performing pure stream control could result in a performance degradation that outweighs the gains from stream control.

4.4 Centralized SCMA

In this section, we present the centralized *stream-controlled medium access* (SCMA) protocol for ad-hoc networks with MIMO links. The design of a centralized algorithm has two potential benefits. (i) It provides a basis for the design of the distributed algorithm, and (ii) It serves as a benchmark against which the distributed algorithm can be compared.

The centralized algorithm has the objective of maximizing the network utilization subject to a given fairness model. The fairness model that we employ is the proportional fairness model. A good exposition on the motivation for the fairness model can be found in [37, 43]. While the primary goal is to come up with a channel allocation vector that is proportionally fair, the maximization of the network utilization can be achieved only by realizing the optimization considerations identified in Section 6.2. Thus, the centralized algorithm attempts to leverage the benefits of stream control and partial interference suppression, while at the same time enabling the passive receiver overloading possible in CSMA/CA.

4.4.1 Insights and Overview

The basis of the centralized SCMA algorithm rests on an observation about the (lack of) receiver overloading problem: there exists a specific subset of links in the network that contribute to the lack of receiver overloading when performing pure stream control. An example of such a link is Link 1 in Figure 4. We refer to such links as *bottleneck links*. An alternative description for bottleneck links is that they belong to *multiple contention regions* in the network. It can be observed in Figure 4 that Link 1 belongs to three contention regions with links 2, 3, and 4 respectively.

If such bottleneck links are scheduled in the non-stream controlled fashion (operating on all k streams), the links can essentially be removed from further scheduling considerations, leaving the scheduling algorithm with only independent contention regions within which pure stream control can be employed. Upon closer look, it can further be observed that the bottleneck links can be identified by identifying vertices in the *flow-contention graph*[1]

[1] A graph with vertices representing links in the underlying network, and an edge between two vertices existing if the two corresponding links contend with each other in the underlying network [37].

$INPUT$: Network Topology graph G = (V,E), and k
V = nodes in the network
E = pair of nodes within reception range
k = number of antenna elements at each node
Step1: Generate Flow Contention Graph G' = (V',E') from G based on neighborhood properties
Step 2: Color the vertices in G' : COLOR(G')
Step 3: Obtain the schedule : SCHEDULE(G")

COLOR(G')
1 Color all links white, add to set WHITE
2 Rank the links based on the tuple (d,s)
3 White links all have a rank of ∞ (some large value)
4 Choose highest rank link, color it red, add to set RED
5 Remove this link and all its edges for further coloring
6 Re-rank the remaining links after updating (d,s) values
7 If there are no links with rank $< \infty$ exit
Else goto line 4

SCHEDULE(G")
8 $\forall\ i \epsilon\ V'$,$service_i = 0$, $resource_i = k$, $allocation_i = 0$, $slot\_index = 0$
9 Do While $((min(service_i) \leq min(service_j))$ $i \epsilon$ RED, & $j \epsilon$ WHITE
10 I = Get_Red()
11 Do While (I $\neq\ \emptyset$)
12 Choose $i \epsilon$ I, s.t. $service_i = min(service(\text{I}))$
13 $service_i = service_i + k$
$resource_i = 0$, $allocation_i = 1$
14 $\forall\ j \epsilon\ Neighbor(i)$, $resource_j = resource_j - w_{ij}k$
15 I = Get_Red()
16 J = Get_White()
17 Do While (J $\neq\ \emptyset$)
18 Do Schedule_white
19 J = Get_White()
20 $slot\_index++$, $\forall\ i \epsilon\ V'$ $allocation_i = 0$, $resource_i = k$
21 Do While $(min(service_j) < max(service_i)$ $i \epsilon$ RED, & $j \epsilon$ WHITE)
22 J = Get_White()
23 Do While (J $\neq\ \emptyset$)
24 Do Schedule_white
25 J = Get_White()
26 $slot\_index++$
$\forall\ i \epsilon\ V'$, $allocation_i = 0$, $resource_i = k$

Schedule_white()
27 Choose the link $j \epsilon$ J, s.t. $service_j = min(service_j)$
28 $service_j = service_j++$ $resource_j = resource_j - 1$, $allocation_j = 1$,
29 $\forall\ p \epsilon\ Neighbor(j)$, $resource_p = resource_p - w_{jp}$

Get_Red()
30 Find I $\subseteq$ RED, s.t. $\forall\ i \epsilon$ I, $resource_i = k$ & $(resource_j \geq w_{ij}k$, $\forall\ j\ \epsilon Neighbor(i)$ & $allocation_j > 0)$

Get_White()
31 Find J $\subseteq$ WHITE, s.t. $\forall\ p \epsilon$ J, $resource_p \geq 1$ & $(resource_q \geq w_{pq}$, $\forall\ q\ \epsilon Neighbor(p)$ & $allocation_q > 0)$

Figure 5: Pseudo Code for Centralized Algorithm

of the underlying network that belong to multiple maximal cliques.

The centralized SCMA algorithm is designed based on the above insights, and has the following key elements: (i) identification of bottleneck links (link classification) - referred to as *red* links in the algorithm; (ii) scheduling of bottleneck (red) links in a non-stream controlled manner; and (iii) scheduling of the non-bottleneck (white) links in the network based on pure stream control.

4.4.2 Centralized Algorithm

We now present the centralized algorithm, the pseudo code for which is presented in Figure 28. We also use a running example of the network topology in Figure 6 (a) to illustrate the different stages of the algorithm.

a) Network Topology b) Flow Contention Graph

Figure 6: Graphs

4.4.2.1 ***Graph Generation***

Given the network topology, the *flow contention graph* G' = (V',E',W) is generated (Figure 6 (b)), where V' represents the set of links in the network (Step 1). E' represents the edges between any two vertices in G', whose links contend with each other in the underlying network, and the weight of the edge (ϵ W) represents the amount of interference caused by one link on the other.

Before being able to classify the vertices, it is necessary to identify all the non-overlapping contention regions in the link contention graph. This is equivalent to the problem of identifying all the maximal cliques in G'. We next explain how this can be achieved.

4.4.2.2 ***Clique Identification and Ranking***

Identifying all the maximal cliques in a graph is known to be an NP-Hard problem. Hence the centralized algorithm makes use of an algorithm that determines all the maximal cliques in chordal graphs (having less than 4 cycles)[2]. It first determines the *perfect elimination ordering* (PEO) using LexBFS (Lexicographic Breadth First Search) [47] for the chordal graph and then applies a *linear* time algorithm that detects all the maximal cliques given the PEO using a theorem by Fulkerson and Gross [18].

Once all the maximal cliques have been obtained, the vertices (in G') are then ranked. Every vertex has two attributes (*d,s*) : *clique degree d* (number of maximal cliques that

[2]Though the algorithm works for only graphs having cycles of size less than four, note that the graph in our case corresponds to the flow contention graph. Hence for a cycle of size four to be present in the flow contention graph, a cycle of atleast size eight must be present in the node graph with no nodes being present inside the cycle.

the vertex belongs to) and maximum size s of all possible cliques that it belongs to. The vertices are ranked lexicographically based on the tuple (d,s) with the vertex having the highest degree ranked first, and the maximum size s is used to break ties. However, it is not necessary to rank vertices that have a degree of one. The vertices are then colored (Step 2).

In the example in Figure 6, the different maximal cliques in Figure 6 (b) are *cdef*, *abc* and *acd*. Vertex c obtains the highest rank with a degree of 3, followed by d that as a degree of 2. The rest of the vertices all have a degree of one.

4.4.2.3 ***Coloring***

Initially all the vertices are colored white[3] (line 1, see Figure 28). Based on the tuple information (d,s) for each vertex, the vertices are ranked lexicographically as described before (lines 2-3). Then the vertex with the highest rank is recursively chosen, and colored red[4], following which, the particular red colored vertex and edges emanating from it are removed from G' (lines 4-5). The tuple (d,s) of the remaining vertices in G' are updated and the remaining vertices are re-ranked once again (line 6). The process repeats until no more vertices can be colored red (line 7). Once this is done the schedule for channel allocation is obtained (Step 3).

In the example, vertex c is colored red first, followed by vertex d. The rest of the vertices, with a degree of one, are colored white.

4.4.2.4 ***Red Vertex Allocation***

The allocation begins with the red vertices which are the bottleneck links in the underlying network. A red vertex can be scheduled in any slot[5] only if it can operate on all k (= number of elements) streams. The red vertices are scheduled based on their service with the minimum serviced link getting scheduled first (rank is used to break ties; a higher rank has more priority) as in line 12. At every slot, the algorithm also attempts to maximize the utilization, i.e. after scheduling a red vertex with k streams, the algorithm checks to see if any other red vertices can be scheduled in the same slot with k streams (line 15). If yes,

[3]A white vertex in G' corresponds to a white link in the network.
[4]A red vertex in G' corresponds to a red link in the network.
[5]Slot corresponds to the time for a packet transmission in the centralized approach.

all such red vertices are scheduled in the same slot. In addition to these red vertices, all the white vertices that can use at least one stream in that slot are also scheduled (lines 16-19). Then the algorithm attempts to increase the number of streams that these scheduled white vertices can use in the same slot. Note that while the red vertices can be scheduled only if they can use k streams, the white vertices can be scheduled even if they can use 1 stream. But among the white vertices that can be scheduled, the algorithm performs a fair allocation of streams (lines 27-29). The scheduling of white vertices in the same slot as that of red vertices (if possible) exploits spatial reuse and thereby helps maximize utilization. The scheduling of the red vertices is stopped when the red vertex with the minimum service (allocation) has received a greater service than that of the white vertex with the minimum service (line 9).

In the example, the switching happens when both the red vertices c and d have obtained a service of k streams each. In one slot, vertex c can alone be scheduled with k streams. But in the other slot when vertex d is scheduled with k streams, vertex b can also be scheduled with k streams.

4.4.2.5 ***White Vertex Allocation***

Once the scheduling switches to the white vertices, the allocation is done on a stream by stream basis to all the white vertices that can be scheduled in the same slot (lines 28-29). This results in a fair allocation of streams to all the white vertices that can be scheduled in the same slot. At a high level, the red vertices being the bottleneck links (responsible for utilization degradation) follow a schedule similar to that of the links in CSMA/CA(k) to avoid the degradation, while the white vertices perform stream control to exploit the advantages of MIMO. The scheduling switches back to the red vertices once the white vertex with minimum service has an allocation greater than or equal to the red vertex with maximum service (line 21). This condition to switch the schedule between red and white vertices ensures that all the vertices obtain an allocation of atleast k streams at the end of every l (= size of the largest maximal clique) slots. The switching conditions and the two level scheduling ensure that the resulting allocation vector is a proportionally fair one.

In the example, white vertices a, b belong to one clique, while e, f belong to another independent clique. Hence, these vertices perform stream control operating on $\frac{k}{2}$ streams each in their independent cliques simultaneously. The switching would occur after two slots when each of these vertices would have obtained an allocation of k streams.

4.4.3 Time Complexity and Optimality

4.4.3.1 Time Complexity

The time complexity incurred in the processing of the centralized SCMA algorithm can be categorized under the two main components of (i) classification of the red and white links, and (ii) obtaining the schedule per slot. The classification component does not vary per slot but instead varies at a much coarser granularity of the changes in the interference topology. Thus, the time complexity involved in the classification component is amortized over several time slots. The schedule component however, varies on a per-slot basis. If the network is represented as a graph, $G = (V, E)$, where V is the set of nodes and E is the set communication edges, then the worst-case time complexity of the link classification component can be bounded by $O(E \log E)$, while the worst-case time complexity of the schedule-per-slot component can be bounded by $O(E)$. The proof for the time complexity results is presented in Appendix A.

4.4.3.2 Optimality

The optimal solution to the exploitation of MIMO's stream control advantages can be formulated as an optimization (network utility maximization) problem. The solution to the problem will output at every link (i) the set of resources to be used for stream control [1,K], and (ii) how the remaining resources should be used for choosing the set of interferers to suppress interference from. The whole problem with the combination of the two components can be formulated as a mixed integer problem, which is hard to solve in polynomial time. More importantly, since the optimal solution to the first component is dependent on stream gains, which vary at the granularity of packets, the optimal solution will also vary at the granularity of packets. This makes it prohibitive to solve the optimization problem on a per-packet basis and consequently does not provide any insights into a feasible implementation

for the solution as well. On the other hand, if the optimization considerations/constraints for the purpose of maximizing utilization are derived based on the interference topology pattern, which remains constant over several packets, as opposed to instantaneous stream control gains, then the solution exploiting such optimizations to maximize utilization will also be a feasible solution for implementation. Thus, given such a desirable feature of having the constraints/optimizations dependent only on the interference topology and not on instantaneous gains, centralized SCMA essentially formulates and optimally solves such a problem to maximize the aggregate network utilization, while also providing a feasible solution for implementation.

However, considering only the second component of the problem, namely the allocation of available resources (degrees of freedom) at a node for suppressing interferers, the constraints in this case are inherently dependent only on the interference topology, thereby motivating the need for solving the optimization problem in its true form. Given that the problem is a mixed integer program, we have obtained efficient approximation algorithms, both centralized and distributed, with approximation guarantees of $\frac{4}{3}$ and 2 with respect to the optimal respectively. Details of these algorithms can be obtained from [63].

4.5 Distributed SCMA

In this subsection we present the distributed SCMA MAC scheme that aims to approximate the centralized algorithm. The distributed scheme achieves this goal in a purely localized manner without the requirement of any large scale coordination in the network. We first identify the key challenges involved in realizing the centralized algorithm in a distributed fashion. Then we use design elements and mechanisms that address these key challenges as building blocks, and present the distributed algorithm. Finally, we describe the details of the distributed mechanisms, concluding the subsection with some practical considerations. For convenience, we shall refer to the distributed SCMA scheme as simply the SCMA scheme, and the centralized SCMA scheme as simply the centralized scheme in the rest of this chapter.

4.5.1 Overview

The basic components inherent in the design of SCMA are outlined below:

- The SCMA MAC scheme is a variant of CSMA/CA. Though CSMA/CA is widely deployed over ad-hoc networks with omni-directional antennas, its two key elements namely *carrier sensing* and *collision avoidance*, are still necessary for the target environment employing MIMO, although for a slightly different purpose. While carrier sensing in the omni-directional environment refers to a node sensing the state of the channel thereby deciding whether to transmit or not, its significance is more in the case of a MIMO environment. The node has to sense the channel in order to determine the number of resources (streams) that it has to sacrifice to suppress the interference, and consequently determine the remaining resources that can be used for transmission. Collision avoidance for the MIMO environment has a similar significance. The transmitter is unaware of the channel state around its receiver and other neighbors, which in turn necessitates the exchange of control information between the nodes. Although the motivation for collision avoidance is the same as in the omni-directional environment, the purpose is to obtain information of available resources at the receiver, and other neighbors, and thereby make a decision on the number of resources to be used for transmission.

- Although SCMA performs collision avoidance, the contention resolution no longer happens in the back-off domain. Instead SCMA, performs contention resolution in the persistence domain similar to [37], which makes it easier to achieve a proportional fairness model. Further, it has been shown in [37] that if the persistence parameters (x_i) of the flows are adapted according to

$$\dot{x_i} = \alpha - \beta p_i x_i \tag{4}$$

  then the system converges to the optimal point of maximizing the aggregate network utilization for a proportional fairness model. α and β are system parameters, p_i is the

loss probability experienced by the flow, and $\dot{x_i}$ is the rate of change of persistence. Hence, if the persistence parameter is assumed to be indicative of the rate experienced by the flow, equation 9 says that by performing a linear increase proportional decrease (LIPD) adaptation of the rate, one can achieve proportional fairness. Proportional decrease in this context refers to a decrease of rate that is proportional to the loss experienced by the flow, which in turn is dependent on the number of contention regions it belongs to as well as the number of flows within each of those contention regions.

- However, the above adaptation assumes a single level scheduling. Hence, to extend the adaptation to the dual scheduling (red and white links) employed in SCMA, the persistence value (P_{old}) obtained from the basic adaptation is translated into a new persistence value, P_{new}. This P_{new} is the same as that of P_{old} for the red links, while it is scaled up for the white links. Since the white links in a clique operate simultaneously, using only a subset of the streams ($K_{new} < k$), their P_{new} value will be a scaled version of P_{old} as in,

$$P_{new} = (P_{old} * K_{old})/K_{new}, \tag{5}$$

  The entire adaptation process still happens on the P_{old} values, but the links thereafter appropriately identify their P_{new} value based on their color and use it to determine access to the channel. This ensures that the resulting channel allocation vector is still a proportionally fair one. Further details on the adaptation mechanism are provided in Section 4.5.2.2.

4.5.2 Distributed Algorithm

We now present the components of the distributed algorithm that address these challenges in the process of approximating the centralized algorithm. In addition, the design of these components also helps leverage the advantages provided by the optimization considerations. The state diagram and the pseudo code for the algorithm are presented in Figures 7 and 51

respectively.

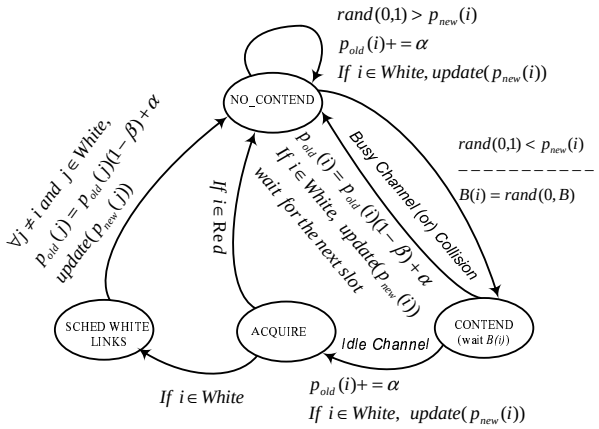


Figure 7: SCMA State Diagram

4.5.2.1 Coloring

Coloring is necessary to distinguish between the red and white links, in order to leverage the optimization considerations. It is sufficient if the links are able to identify whether they belong to multiple contention regions or not, instead of actually identifying the number of contention regions they belong to, for the process of coloring. Further, only a red link will be overwhelmed in terms of resources due to members of the different cliques that it belongs to, transmitting at the same time. This fact is exploited in aiding the transmitting and receiving nodes determine the color of their respective links.

Every transmitter initially starts transmitting on all the streams (with $P_{new} = P_{old}$) like CSMA/CA(k) until it determines its color. It locally determines the color of its link in conjunction with its receiver. Specifically, the transmitter observes the usage of resources between every two slots that it has gained access to the channel. If the transmitter or the receiver observe more than k streams during any of the slots (duration of a RTS/CTS/DATA/ACK exchange in the distributed approach) that it has not obtained access to the channel, then the link is automatically colored red (line 16). It is possible for a link to be red even if the transmitter and the receiver individually do not observe more than k streams. This is because there could be an active link within the interference range of the transmitter but not within that of the receiver or vice versa. The transmitter and

```
CONTEND( )
1 ∀ Slot S, node i
2       state = NO_CONTEND
3       If uniform(0,1) ≤ P_new,i
4           state = CONTEND
5           B_i = uniform(0,B)
6           Defer(B_i)
7           If (Check_resources( ) == Available)
8               Acquire_channel()
9               If (Acquire_status( ) == Collision)
10                  P_old,i = P_old,i * (1 − β)
11                  If i ϵ WHITE
12                      P_new,i = (P_old,i * K_old,i) / K_new,i
13              else state = ACQUIRE
14                  ∀ j ϵ Neighbor(i)
15                      Update Resources
16                      If (resources_j < 0), color(j) = RED
17                  Recolor(i)
18                  If (i ϵ WHITE), Co-ordinate_Schedule(i)
19                  If (Remaining_resources(i))
20                      K_new,i = K_new,i + Remaining_resources(i) / white_count
21          else P_old,i = P_old,i * (1 − β)
22              If i ϵ WHITE
23                  P_new,i = (P_old,i * K_old,i) / K_new,i
24      P_old,i = P_old,i + α
25      If i ϵ WHITE
26          P_new,i = (P_old,i * K_old,i) / K_new,i

Co-ordinated_Schedule(i)
27 ∀ j ϵ Neighbor(i) && color(j) == WHITE
28 P_old,j = P_old,j * (1 − β)
29 P_new,j = (P_old,j * K_old,j) / K_new,j
Recolor(i)
30 Check slot history from previous service slot
   in conjunction with receiver to color
31      If i ϵ WHITE, K_new,i = K_old,i / white_count
```

Figure 8: Pseudo Code for Distributed Algorithm

receiver in this case would not be able to "independently" identify the correct color. Hence the transmitter during its RTS/CTS exchanges with the receiver compares its version of the winner list (IDs of nodes that have obtained channel access in its neighborhood) with that of the receiver for the slots in between their successive channel accesses. If the list happens to be different in at least one of the slots, then the link is colored red, since this effectively means that this link as a whole is exposed to more than k stream transmissions at the same time (line 17). Otherwise the link is colored white.

4.5.2.2 Contention and Channel Access

Once the nodes have successfully colored their respective links, they adopt a channel contention mechanism that is tuned to their color. There are four possible states in which a node can be, namely *Contend*, *No_Contend*, *Acquire* and *Sched_White_Links* (Figure 7). Every node having a packet to transmit, first decides to contend for the channel with a probability of P_{new}. This persistence probability P_{new} is the same as P_{old} for the red links, while it is scaled for the white links (line 12). If the node succeeds, it moves from the *No_Contend* state to the *Contend* state (lines 2-4), where it chooses a waiting time uniformly distributed from the interval (0,B). B is a constant and set to 32 in the simulations as advocated in [37]. The node then waits for the back-off period (in slots), after which it tries to access the channel to see if the channel is busy (lines 5-7). The busy state of the channel in our case corresponds to a lack of sufficient amount of resources at the transmitter or the receiver or the two-hop neighbors.

If the node finds the channel to be busy, it gives up the slot and decrements its persistence by $\beta * P_{old}$. Similarly if the channel is idle but if the node faces or detects collision, it decrements its persistence by a factor of β. In addition to decrementing the value of P_{old}, if the node belongs to a white link then it also has to update its P_{new} value (lines 8-12). On the other hand if the node finds the channel to be idle, and does not experience any collision then it moves to the *Acquire* state where it transmits. Every node in the two-hop neighborhood of this transmission would automatically expend the appropriate number of resources to suppress this transmission (line 15). At the end of the slot, all the nodes having a packet to transmit in the next slot increase their persistence P_{old} by α with the white links also updating their P_{new} value (lines 24-26). The values of α and β are chosen to be 0.1 and 0.5 based on the rationale provided in [37].

In being able to determine if the channel is busy, a node needs to know about the resource availability at nodes in its two hop neighborhood. This can be achieved by piggybacking the amount of resources remaining at a node in its control packet transmissions. However, to make these control packets decode-able in the two hop neighborhood, the reception range needs to be extended by a factor of two. This in turn can be achieved by transmitting the

control (RTS/CTS) packets as multiple copies (dependent signals) on atleast four streams (using space-time block codes). In this case, we exploit the diversity gain of MIMO to provide us with this range extension factor of two (with 4 element DAAs and a path loss exponent of 4), instead of its spatial multiplexing gain. Note that this range extension will not be the same in all directions and will depend on the radiation pattern currently used by the transmitting node. This range extension mechanism has the additional benefit of aiding the white links in their stream control process, which we explain subsequently. One might think that the disadvantage of not performing spatial multiplexing for the control packets might lead to a degradation in performance. But our contention is that the benefits of diversity gain for control packets is much more significant in terms of addressing several of the issues associated with distributed operations.

4.5.2.3 White Link Adaptation

For the white links to be able to perform stream control and hence determine the appropriate persistence (P_{new}) with which to contend for the channel, they need to estimate the fair share of resources to use in their contention region. While the computation of fair share of the red links is relatively easy (k streams), the fair share estimation for the white links is non-trivial.

Every node advertises the color of its link (if colored) in its transmissions. During the initial phase, when a white link may not be aware of the other white links in its contention region, it will not be able to arrive at the correct fair share. Hence for this purpose, every node transmits for one more slot on all k streams (even after it has colored itself white) along with its color information to inform the other members of the clique about its newly colored link. Since the control packets are decode-able within a range that is twice the normal reception range due to the diversity gain, the other members of the clique will receive this information. This helps the white links keep track of the number of white links in the same clique (say w) and hence help them arrive at the fair share in the clique k_{new} ($=\frac{k}{w}$) (line 31).

4.5.2.4 Co-ordinated Scheduling

Distributed execution of the stream control mechanism by the white links is a challenge that has to be accomplished. This requires that the white links operate simultaneously on a subset of the maximum allowable number of streams. Since persistence is used to ensure proportional fairness, it is possible that only some of the white links in a clique actually contend for a slot. Hence, even if one of the white links gains access to the channel it must be ensured that all the other white links in the same clique are also scheduled in the same slot, failing which the advantage of stream control cannot be leveraged.

Accordingly, when the first white link in a clique gains access to the channel, it also co-ordinates the other white links in the clique to transmit in that slot using their own estimated fair share (line 18). This corresponds to the node entering the *Sched_White_Links* state in Figure 7. Specifically this link's RTS/CTS messages will contain a flag ordering the schedule of all the white links in the clique. Since the control messages can be decoded within the two-hop neighborhood (owing to the range extension factor of two), all the white links in the clique will be able to listen to the command of this initiating link and thereby schedule themselves in the same slot, irrespective of whether they contended for channel access in that slot or not. However, the contending white links that were not the initiator of the co-ordinated scheduling will still have their P_{old} values decremented by the factor β to be in conformance with the normal adaptation algorithm (lines 27-29).

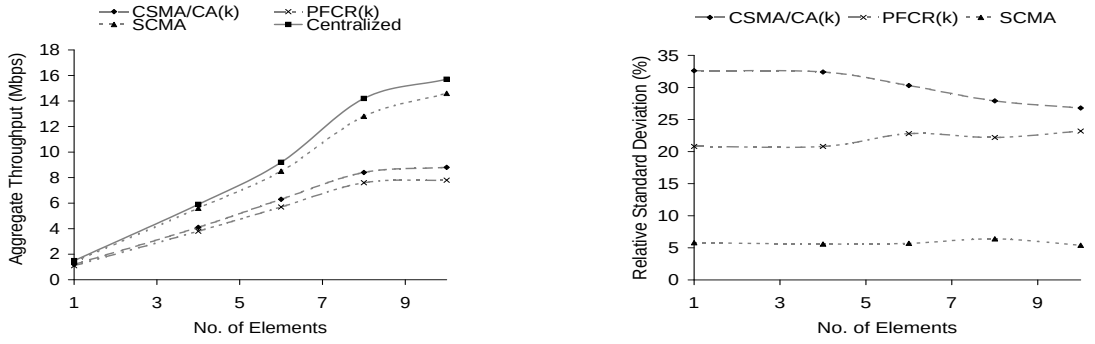
In addition, to be able to leverage the advantages of flexible interference suppression, the nodes belonging to white links observe if their fair share can be increased based on the remaining resources available at the end of their transmission (line 20). If so, the node increases its fair share only by a fraction of the remaining resources to allow other white links in the clique to increase their fair share as well. Thus the resources are fully utilized in the clique, thereby leveraging the advantage of flexible interference suppression.

4.6 Performance Evaluation

In this subsection, we present a small subset of the performance results for SCMA obtained through simulation studies. We use an event-driven packet level simulator for recording

the results. We use user datagram protocol (UDP) as the transport layer protocol and constant bit rate (CBR) as the traffic generator. The packets are generated at a rate of 100 packets/sec and are of size 1KByte. The number of flows and elements in the antenna array are parameters that vary from one experiment to another. We extend the distributed proportional fair contention resolution (PFCR) mechanism [37] to PFCR(k) and CSMA/CA to CSMA/CA(k) for ad-hoc networks with MIMO links and consider these as baseline protocols in our simulation study. CSMA/CA(k) and PFCR(k) are variants of CSMA/CA and PFCR respectively that use k streams for their transmissions and receptions. Further, we use a simple model that conservatively assumes only about 20% gain from performing stream control. The statistical gains on the different streams could have significant disparities resulting in a much larger gain [13]. We compare the performance of SCMA, CSMA/CA(k) and PFCR(k) with that of the centralized protocol and thereby present results to highlight the benefits of the mechanisms involved in our distributed protocol.

The metrics we use for comparing the various schemes are *throughout* and *relative standard deviation*. Since the algorithm provides proportional fairness that is location dependent, standard deviation or normalized standard deviation would not be able to capture well the degree of fairness provided by the scheme. Hence we have chosen to use relative standard deviation as the fairness metric. Specifically, the distribution of throughput of the various schemes is compared against that of the centralized scheme and the deviation normalized to the mean is obtained.



(a) Throughput (b) Unfairness

Figure 9: Random Network Topologies

We consider a random network topology consisting of 50 nodes distributed uniformly

over an area of 750m by 750m. The random scenarios are generated using the *setdest* tool and the results are averaged over several seeds and also across different number of elements present in the antenna array. Mobility is not considered in these scenarios. Figures 9(a) and (b) present the results for the various schemes in comparison with the centralized scheme. The utilization in Figure 9(a) shows an increasing trend with the number of elements for all the schemes. However, for CSMA/CA(k) and PFCR(k) the improvement in utilization obtained for every extra element employed starts to decrease due to the absence of stream control. Hence the scalability of SCMA in terms of utilization is better when compared to CSMA/CA(k) and PFCR(k). The fairness results are presented in Figure 9(b). Although we are not able to conclude any relation about the trend in fairness with respect to the number of elements, the result clearly shows that SCMA provides an improvement of around 15% to 25% when compared to CSMA/CA(k) and PFCR(k).

CHAPTER V

ROUTING IN AD-HOC NETWORKS WITH MIMO LINKS

5.1 Overview

While the previous chapter outlined the design of a MAC protocol for use with MIMO links, the focus of this chapter is to go one step ahead and explore the various capabilities of MIMO links from the perspective of routing layer protocols. We identify whether and how each of the capabilities can translate to improved performance at the routing layer. More specifically, this chapter deals with the following aspects:

- We identify the capabilities of MIMO links and capture their relevance to routing layer protocols.
- We analyze both theoretically and practically the relative tradeoffs of exploiting the different capabilities of MIMO links.
- We propose a reactive routing protocol whose components are built on the insights gained from the analysis, and hence leverage the PHY layer characteristics in their operations to improve network performance.

Briefly, we identify two fundamental capabilities of MIMO links, namely spatial multiplexing and diversity that can be exploited by the routing layer protocols in their operations. However, since these two capabilities cannot be fully leveraged at the same time, it becomes necessary to investigate the relative trade-offs between spatial multiplexing and diversity, in order to determine the optimal strategy of operation for improving the aggregate network throughput. To this end, we analytically study the benefits and drawbacks of both the strategies from the perspective of routing layer protocols. The study also incorporates

practical considerations in determining the optimal strategy of operation. A routing protocol is then proposed with components built on the insights gained from the study. The corresponding cross-layer support required from the MAC layer is also identified and accommodated in the design. The effectiveness of the components in the proposed solution is then comprehensively evaluated through simulation studies.

5.2 Gains and Strategies

As outlined in Chapter 4, a MIMO link is capable of operating in two modes (strategies), namely spatial multiplexing and diversity.

The gain associated with spatial multiplexing is referred to as the spatial multiplexing gain. This multiplexing gain can provide a linear increase (in the number of elements) in the asymptotic link capacity, which is given by the following equation [19],

$$C \approx min(M, N) \log_2(1 + \rho) \tag{6}$$

where M and N correspond to transmitter and receiver antenna elements and ρ represents average SNR at any one receive antenna.

The gain associated with diversity is referred to as the diversity gain. In practical lossy channels, diversity can help significantly reduce the bit error rate (BER) on the link. At high SNR, this reduction in BER (p) as a function of the diversity order (d) can be given as [78],

$$p \approx \frac{1}{SNR^d} \tag{7}$$

An increase in diversity gain comes at the cost of a reduction in rate and hence the multiplexing gain, thereby leading to a fundamental trade-off between the two strategies [78]. The increased reliability provided by diversity gain can be used in one of three ways, (i) reduce p on the link, (ii) for a required p on the link, the increased reliability provided by diversity can be translated to an increase in the SNR at the output of the combiner, which can then be used for increasing the communication range of the link, and (iii) for a given p and SNR requirement, the transmit power consumption can be minimized. Since

power consumption minimization is an orthogonal direction of optimization, we focus only on the first two methods of leveraging diversity gain in this work with aggregate network throughput being the parameter of optimization.

In the rest of this chapter, we shall refer to the spatial multiplexing strategy as MUX, diversity with reduced BER as DIV-BER and diversity with increased communication range as DIV-RANGE for simplicity. Specifically, at the PHY layer all nodes in the network employ spatial multiplexing (eg. VBLAST [16]) for communication in MUX, while they employ diversity (eg. space-time block codes, STBC [1]) in both DIV-RANGE and DIV-BER. The difference in the strategies arises from how and which specific gains are leveraged. In MUX, the routing protocol uses omni-directional communication range for route discovery, resulting in routes with all omni-links on which spatial multiplexing is performed (referred to as "rate" links) during data transfer. In DIV-RANGE, diversity and hence SNR gain is exploited by the routing protocol in route discovery itself, resulting in routes composed of links with larger communication ranges ("range" links). However, in DIV-BER, omni-directional communication range is first used to obtain routes with omni links, with diversity being exploited later on for data transfer, thereby reducing the BER on the link ("reliable" links). While DIV-BER and DIV-RANGE both exploit diversity, the difference arises from whether diversity is exploited in the route discovery component or not. Further, the area of inhibition due to a transmission is lesser in DIV-BER than in DIV-RANGE. While the same transmit power is used in both cases, since the links are much shorter in DIV-BER (higher SNR), the ability (probability) of capturing a packet at the receiver is much larger in DIV-BER, thereby providing better spatial reuse. Finally, none of the strategies assume feedback of any channel state information (CSI) from the receiver to transmitter on the links.

5.3 Theoretical Analysis

In this section, we analytically explore the trade-offs between using spatial multiplexing and diversity. The basic model assumed for this analysis is similar to the one used for the protocol model in [20]. Let us consider a random network operating on a single channel as a

unit area disc with n nodes independently and uniformly distributed. Assume the nodes to possess omni-directional antennas initially. Each node is capable of generating $\lambda(n)$ bits/sec of traffic to a random destination. Let us consider a generic single path routing protocol. Further, let the bandwidth of the wireless channel be W bits/sec. From the seminal work in [20], we know that when edge effects are incorporated, the following inequality holds in a planar disk under the protocol model,

$$n\lambda(n)h \leq \frac{16W}{\pi\Delta^2 r^2} \tag{8}$$

where h represents the average hop count of a flow and r the transmission/reception range. Δ represents the guard zone parameter which roughly determines the region silenced by a transmission. The term on the left side in inequality 75, $n\lambda(n)h$ determines the total amount of traffic (bits/sec) that needs to be carried by the entire network due to multi-hop relaying burden. The term on the right side determines the total number of simultaneous transmissions in the network (or spatial reuse), which in the presence of edge effects has been shown to be bounded by $\frac{16A}{\pi\Delta^2 r^2}$, where A is the area of the network. Hence, for the generated traffic to be carried by the network, we need to ensure that equation 75 holds good.

The importance of diversity arises in the presence of lossy links. Hence, the average probability of bit error (BER) p on the link is an important factor in the analysis. A more useful parameter is the packet error rate (PER) P, which under the assumption of no forward error correction (FEC), is given by,

$$P = 1 - (1-p)^L, \tag{9}$$

where L is the size of the packet in bits. Now, when p and hence P are extremely small but non-zero, the links can almost be considered to be loss-less. In such cases, using diversity to directly reduce p further on the link does not contribute much. Hence, the diversity gain can then be exploited to increase the communication range and hence reduce the number of hops and thereby the multi-hop burden. Thus, we consider two cases of (i) non-lossy (p being very small but non-zero) and (ii) lossy links, for the analysis. Also assume the nodes to now possess a multiple-element array with M elements each.

5.3.1 Non-lossy links

In the case of MIMO links, spatial multiplexing (MUX) can increase the capacity of the link by a factor that is linear in the number of antennas elements employed. This factor is given by the minimum of the transmitter and receiver antenna elements. This increased capacity of the links can directly be translated to a linear increase in the network and per-node throughput as well. In effect, the spatial reuse in equation 75 can be visualized to be scaled by the linear factor. The linear increase in throughput $\lambda_s(n, M)$ can now be captured by the following inequality,

$$\lambda_s(n, M) \leq \frac{16W \cdot M}{\pi n h \Delta^2 r^2} \tag{10}$$

where M represents the number of antenna elements at both transmitter and receiver by our assumption. The case of diversity being used to directly reduce the bit error rate on the link (DIV-BER) does not contribute much when the links are almost loss-less (p being extremely small). However, it has the drawback of losing out on rate when compared to MUX and hence the throughput ($\lambda_{db}(n, M)$) can be given by,

$$\lambda_{db}(n, M) \leq \frac{16W}{\pi n h \Delta^2 r^2} \tag{11}$$

Hence, DIV-BER is not a beneficial option when the links are almost loss-less. Let us hence, consider the case of diversity gain being exploited for range extension (DIV-RANGE). If the function that characterizes the range increase relative to omni-directional transmission range is represented by $f(p, d)$ (where d denotes diversity order = MN; f depends on relation 7 at high SNR), then as the range of transmission increases from r to $f \cdot r$, the average hop length of routes decreases from h by a factor of f to $\frac{h}{f}$, while the inhibited zone due to a transmission is scaled up by a factor of f^2, which has the dominant impact. Hence, the corresponding throughput $\lambda_{dr}(n, M)$ and hence inequality becomes,

$$\lambda_{db}(n, M) \leq \frac{16W}{\pi n \frac{h}{f} \Delta^2 (r \cdot f)^2} = \frac{16W}{\pi n h \Delta^2 r^2 f} \tag{12}$$

Thus, it can be seen from equations 10, 11 and 12 that in the case of non-lossy links, *while increasing rate through spatial multiplexing (MUX) helps increase the per-node and hence*

aggregate throughput, increased range (DIV-RANGE) degrades performance and increased reliability (DIV-BER) does not provide any significant gains.

5.3.2 Lossy links

Let us now consider scenarios where the links are lossy, each with an average BER of p with the associated PER being given by P as in equation 9. We assume end-end reliability semantics with a semi-reliable MAC layer (eg. IEEE 802.11) at each of the nodes. Recall that, when sources generate traffic at the rate of $\lambda(n, M)$ bits/sec, the total amount of traffic generated due to multi-hop relaying is $n\lambda(n, M)h$. In the presence of time-correlated fading, when a link i fails on a path, the resources used by the packet till hop i are wasted, even in the presence of semi-reliable MAC layers and FEC. This is because a semi-reliable MAC layer (like 802.11 using a fixed number of retransmissions) or FEC can help recover link errors in the presence of time-uncorrelated fading where the bit errors can be assumed to be independent. However, this is not possible in time-correlated (Rayleigh) fading; occurring in several practical environments and hence considered in this work. In time-correlated Rayleigh fading, errors occur in bursts and last typically from tens to hundreds of milliseconds [71], which in turn corresponds to tens of packet transmissions at a channel rate of 2 Mbps with 1KB packet size and even more at higher channel rates.

Thus, additional transmissions are required to transport a packet successfully in the lossy case. Consequently the throughput capacity decreases. The additional transmissions are equivalent to increased multi-hop relaying which can be captured by finding the equivalently increased average hop length over which a packet needs to be carried before reaching the destination successfully. We shall refer to this hop length as the lossy hop length and the original hop length as the non-lossy hop length in the rest of our discussion. In this context, we have the following theorem:

Lemma 1 *The average lossy hop length $\bar{h}$ over which packets travel in the case of lossy links with a PER of P (> 0) is given by,*

$$\bar{h} = h + \frac{[1-(1-P)^h][1-(1+Ph)(1-P)^h]}{P(1-P)^h} \tag{13}$$

where h is the average non-lossy hop length or the true average hop length in the underlying topology.

Proof: Proof available in Appendix B. ■

5.3.3 MUX

In the case of MUX, link reliability is not impacted and hence the increased average hop length $\bar{h_s}$ is the same as $\bar{h}$ in equation 13. Hence, the per-node throughput ($\lambda_s(n, M)$) can now be given by,

$$\lambda_s(n, M) \quad \leq \quad \frac{16W \cdot M}{\pi n \bar{h_s} \Delta^2 r^2} \tag{14}$$

5.3.4 DIV-BER

When diversity is exploited to directly reduce the BER on the link, p reduces to p^{M^2} as given by equation 7, where p is the BER in the case of multiplexing. Hence, the packet error probability R also decreases as,

$$R = 1 - (1 - p^{M^2})^L \tag{15}$$

This leads to an increase in the success probability of a packet reaching the destination, while at the same time also helping reduce the average lossy hop length $\bar{h_{db}}$ as,

$$\bar{h_{db}} = h + \frac{[1 - (1 - R)^h][1 - (1 + Rh)(1 - R)^h]}{R(1 - R)^h} \tag{16}$$

The reduction in BER and hence the decrease in multi-hop burden comes at the cost of not being able to use the elements to increase the rate of transmission as in MUX. Hence, the throughput when diversity is used to reduce the BER on the link ($\lambda_{db}(n, M)$) is now given by,

$$\lambda_{db}(n, M) \leq \frac{16W}{\pi n \bar{h_{db}} \Delta^2 r^2} \tag{17}$$

5.3.5 DIV-RANGE

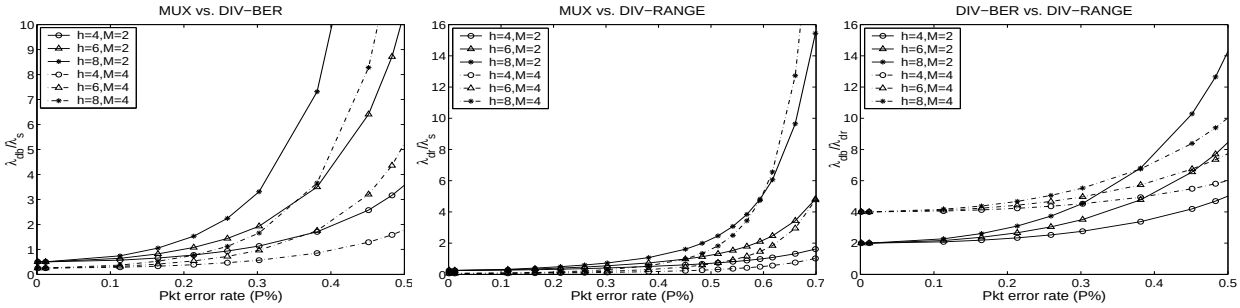
On the other hand, in DIV-RANGE the gain in SNR is used to increase the communication range of the link. Though the BER and hence the PER on the links are maintained to be the same p and P respectively as in MUX, the reduction in the number of hops for each flow,

results in a reduced multi-hop relaying burden and consequently an increased probability of successful packet reception at the destination. The average lossy hop length $\bar{h_{dr}}$ is now given by,

$$\bar{h_{dr}} = \frac{h}{f} + \frac{[1-(1-P)^{\frac{h}{f}}][1-(1+\frac{Ph}{f})(1-P)^{\frac{h}{f}}]}{P(1-P)^{\frac{h}{f}}} \tag{18}$$

where f represents the function that characterizes the range increase factor (over omni-directional range) as a function of the number of elements. As in DIV-BER, DIV-RANGE also incurs the cost of not being able to use the elements to increase rate. The throughput in this case $\lambda_{dr}(n, M)$ can be given as,

$$\lambda_{dr}(n, M) \leq \frac{16W}{\pi n \bar{h_{dr}} \Delta^2 r^2 f^2} \tag{19}$$



a) MUX vs. DIV-BER b) MUX vs. DIV-RANGE c) DIV-BER vs. DIV-RANGE

Figure 10: Theoretical Comparison of Strategies

5.3.6 Inferences

In the case of diversity, it can be seen from equations 17 and 19 that while there is a decrease in rate when compared to MUX, there is also a relative increase in throughput due to a decrease in the multi-hop relaying burden. Hence, it becomes necessary to compare the relative gains provided by the different strategies.

We use the limiting values of the inequalities 14, 17 and 19 for relative comparison purposes. The rationale for using the limiting values is as follows. As long as the transmission range is more than that required for critical connectivity, the above inequalities hold. Further, even if the constructive lower bounds are considered, the three fundamental components impacting throughput, namely n, r and $\bar{h}$, remain to be the same although

the specific constants may change. This is true for MIMO links as well. Hence, relative comparisons of the different strategies with respect to these three components will lead us to the same inferences irrespective of whether upper or lower bounds or average throughputs are considered. Thus, our relative comparison based on the upper bounds serves as a reasonable starting point to explore the benefits of the different strategies and to gain insights into their applicability. Further, the strategies are comprehensively evaluated again under practical network conditions in Section 5.4 to verify the insights gained from the analysis.

We consider a packet size of 1KB, number of elements of 2 and 4 and average hop lengths h (in the underlying topology) of 4, 6 and 8 in the comparisons. We have the following three cases:

5.3.7 MUX vs. DIV-BER

In order for DIV-BER to provide higher gains than MUX, we have

$$\frac{\lambda_{db}(n,M)}{\lambda_s(n,M)} = \frac{\bar{h_s}}{\bar{h_{db}}M} \geq 1 \tag{20}$$

$$\implies \frac{\bar{h_s}}{\bar{h_{db}}} \geq M \tag{21}$$

Figure 10(a) plots the inequality $\frac{\lambda_{db}(n,M)}{\lambda_s(n,M)}$. It can be seen that DIV-BER serves to be a better strategy at packet error rates from about 15% for $h = 4$, and at even lower packet error rates for larger hop lengths. The gain over MUX is more than linear with increasing packet error rates. Further, this gain increases as the hop length is increased since the vulnerability of route failures is more in MUX. However, as the number of elements increases, it can be seen that the gain of DIV-BER over MUX is decreased. This is because, the usage of 2 elements is sufficient in this case (for the BER considered) to reduce the BER to a negligible value. Hence, the usage of any further elements to reduce BER does not bring much benefits. On the other hand, the loss of rate due to using extra elements towards diversity continues to contribute a linear degradation factor. We elaborate further on this in the context of practical considerations in Section 5.4.

Thus, under lossy conditions, it is essential to devote only the appropriate number of elements to diversity such that link BER is significantly reduced, while the rest of the elements

should be devoted to multiplexing to maximize the rate on the link.

5.3.8 MUX vs. DIV-RANGE

In order for DIV-RANGE to provide higher gains than MUX, we have

$$\frac{\lambda_{dr}(n,M)}{\lambda_s(n,M)} = \frac{\bar{h_s}}{\bar{h_{dr}} M f^2} \geq 1 \quad (22)$$

$$\implies \frac{\bar{h_s}}{\bar{h_{dr}}} \geq M f^2 \quad (23)$$

Figure 10(b) plots the inequality $\frac{\lambda_{dr}(n,M)}{\lambda_s(n,M)}$. It can be seen that DIV-RANGE serves to be a better strategy only at higher packet error rates from about 50% for $h = 4$, and at lower packet error rates for larger hop lengths. The other observations are similar to the DIV-BER case with one interesting exception. When the number of elements is increased from 2 to 4, the impact of rate loss due to diversity and the reduced spatial reuse due to increased communication range, reduce the gains causing DIV-RANGE to serve as a better strategy only from error rates of 70% for smaller hop lengths, $h = 4$. While this is expected, as the hop length increases, the degradation due to increased elements is compensated by a significant reduction in the average lossy hop length (and hence multi-hop burden) compared to MUX to result in a higher gain for $h = 8$ than in the 2 element case.

Thus, DIV-RANGE provides better gains than MUX only at very high packet error rates and the gain decreases with increasing elements unless the hop lengths of the flows are large.

5.3.9 DIV-BER vs. DIV-RANGE

For DIV-BER to outperform DIV-RANGE, we have the following condition,

$$\frac{\lambda_{db}(n,M)}{\lambda_{dr}(n,M)} = \frac{\bar{h_{dr}} f^2}{\bar{h_{db}}} \geq 1 \quad (24)$$

$$\implies \frac{\bar{h_{dr}}}{\bar{h_{db}}} \geq \frac{1}{f^2} \quad (25)$$

Figure 10(c) plots the inequality $\frac{\lambda_{dr}(n,M)}{\lambda_s(n,M)}$. It can be seen that DIV-BER outperforms DIV-RANGE for all packet error rates and for all hop lengths. Further, the improvement increases with increasing packet error rates and with increasing hop lengths. This is because, DIV-RANGE, in addition to the loss of rate as in DIV-BER, also suffers a loss of throughput

due to increased communication range (larger inhibition zone), which has to be compensated by the decrease in the multi-hop burden. An interesting observation to be made is that as the number of elements increases from 2 to 4, while the gain for lower hop lengths increases as expected due to the term f^2 in equation 24, the gain for larger hop lengths reduces as can be seen in Figure 10(c). This can be related to our earlier discussion that at increased number of elements and larger hop lengths, reduction in multi-hop burden in DIV-RANGE is significant enough to compensate for the reduction in spatial reuse.

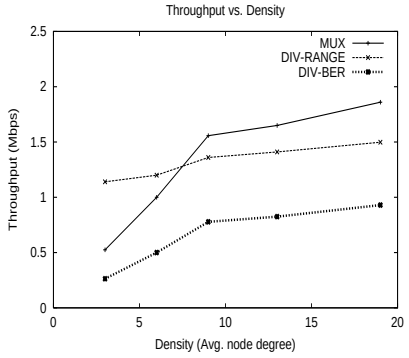
The DIV-BER scheme provides better gains than the DIV-RANGE scheme under conditions of lossy links.

In summary, we have seen that while MUX outperforms DIV-BER and DIV-RANGE in the case of non-lossy links, it is possible for DIV-BER to outperform MUX and DIV-RANGE in the case of lossy links even for moderate loss rates.

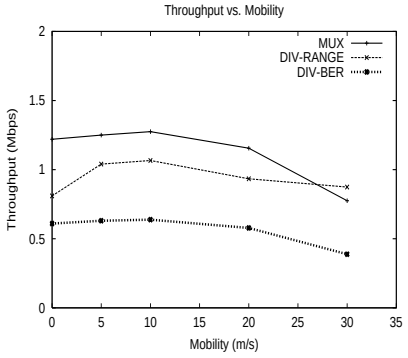
In either case, we find that DIV-RANGE exhibits the lowest performance (except for very high loss rates). Does this mean that DIV-RANGE is not an option to be considered? Note that, the analysis considered only static ad-hoc networks with sufficient network density for connectivity (using omni-directional transmission range) in the presence/absence of channel (link) errors. There exist other aspects of ad-hoc networks, namely mobility and sparse density that are rather cumbersome to model analytically. In the next section, we re-evaluate the different strategies under practical network conditions of varying mobility and density (in addition to the static connected scenarios considered in the analysis) through simulations.

5.4 Practical Considerations

In this section, we shall discuss the impact of several practical considerations on the different strategies. Specifically, in addition to the lossy nature of links, we shall consider varying density (average node degree) and mobility in ad-hoc networks. Simulations results from *ns2* are used for the study with aggregate network throughput being the metric of interest. Briefly, 100 nodes are considered in a 2-d grid whose area is varied in order to vary density

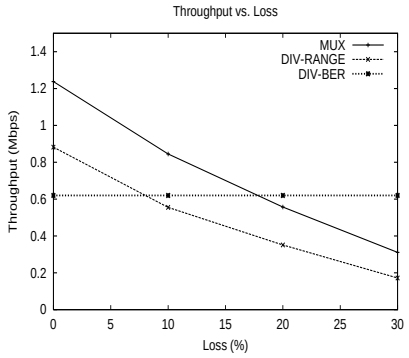


a) Density (Load = 30 flows)

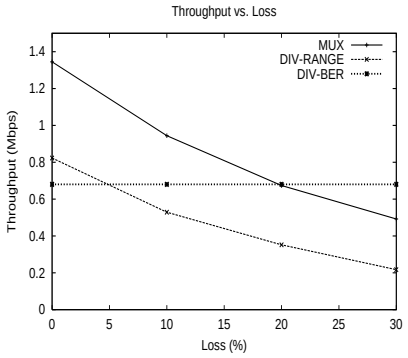


b) Mobility

Figure 11: Impact of Practical Components on Network Throughput (1)



a) Loss (Load = 10 flows)



b) Loss (Load = 15 flows)

Figure 12: Impact of Practical Components on Network Throughput (2)

with the default being 1000m by 1000m. The default number of flows, mobility and packet loss probability on the links are 30, 0 m/s and 0% respectively unless specified otherwise. Each flow is a CBR flow using UDP and generating data of packet size 1KB at the rate of 100Kbps on a 2Mbps channel. Each node is equipped with two element array. DSR is used as the reactive routing protocol with MUX, DIV-RANGE and DIV-BER being incorporated as outlined in Section 8.2. The range extension function f is assumed to be linear[1] in the rest of this chapter. Random waypoint mobility model is used for generating scenarios involving mobility. IEEE 802.11b in the DCF mode is used as the MAC protocol.

[1]It is more than linear only for very high SNR's and is a reasonable assumption for moderate-high SNR's. Additional details can be found in Appendix B.

5.4.1 Density

Partitions are a likely possibility in sparse ad-hoc networks which disrupt communication due to the lack of a route to the destination. They may be caused due to the sparse nature (density) of the network or due to mobility of the nodes. In cases of mobility, there is a finite probability for the partition to be bridged back again due to node mobility itself. However, the case of partitions in static sparse networks is much more severe since it disrupts communication permanently. Hence, in this subsection we shall focus on partitions resulting in sparse static networks and postpone the impact of mobility to the next subsection.

While MUX can increase throughput, it cannot increase the communication range and hence operates using the omni-directional communication range. Thus, in static sparse networks, the network graph may not be connected in MUX, resulting in the possibility of no routes existing for some of the flows. This leads to a degradation in aggregate network throughput. DIV-BER still uses the same omni-directional communication range and hence suffers from the same problem as MUX. DIV-RANGE on the other hand, has the potential to bridge partitions through its increased communication range. However, it suffers from the reduced spatial reuse resulting from the increased communication range. Thus, while DIV-RANGE is a better option at low network densities where the ability to bridge network partitions outweighs the reduced spatial reuse, this is not true at high network densities. This can be observed from the results in Figure 11(a). Since the links are assumed to be non-lossy, this makes DIV-BER a poor candidate as outlined in Section 5.3.1.

Hence, the optimal strategy with respect to varying node density would be to use DIV-RANGE at lower network densities and MUX at higher network densities. More specifically, DIV-RANGE must be used only by those nodes that would not be able to deliver the packet to the next hop towards the destination using MUX. This would keep the reduction in spatial reuse due to increased communication range to a minimum while at the same time helping all sources obtain a route to their respective destinations in the network.

5.4.2 Mobility

Mobility is an inherent component of ad-hoc networks. Increased speeds increase the number of route errors, thereby increasing the duration a flow spends without a route to its destination. This in turn leads to a degradation in throughput. In addition, reactive routing protocols such as DSR might purge all packets which are bound to use a recently failed link during the propagation of the route error towards the source. Thus, the bandwidth resources used by the packets before being purged are wasted. Once again, to isolate the impact of mobility on the different strategies, we assume the links in the network to be almost lossless. In this context, the following can be shown.

Lemma 2

$$P_{l\_db} < P_{l\_dr} < P_{l\_s} \tag{26}$$

where P_{l_db}, P_{l_dr} and P_{l_s} are the probabilities of a link going down due to mobility in DIV-BER, DIV-RANGE and MUX respectively.

Proof: Proof available in Appendix B. ■

This implies that the impact of mobility is greatest on MUX at higher speeds followed by DIV-RANGE and DIV-BER. The links in DIV-BER operate on BERs well below the threshold and can hence sustain increased communication ranges during mobility without bringing the link down. In DIV-RANGE, the increased communication range gives the node more area to move about without breaking the link when compared to MUX.

However, the throughput results presented in Figure 11(b) indicate trends contrary to expectation, with DIV-BER performing the worst. Since the links are considered to be non-lossy, DIV-BER suffers a degradation in rate compared to MUX. However, the increase in reliability helps sustain the links longer during mobility, but then results in routes with increased range per hop (number of hops remaining unaltered), thereby resulting in highly sub-optimal routes. These sub-optimal routes tend to degrade performance more than the case wherein route failures occur, however resulting in new (better) routes to be formed. It is also interesting to note that MUX outperforms DIV-RANGE for most of the mobilities. This is because the region silenced by a transmission is more in DIV-RANGE due to the

increased communication range. This in turn increases the number of contending nodes to any node and hence increases the number of collisions due to distributed MAC operations. This would also trigger more route failures, thereby overcompensating for the increased robustness to mobility. However, this increased number of collisions is not dependent on mobility and is a constant impact for a fixed load. Hence, as mobility increases, DIV-RANGE's increased robustness to mobility helps it outperform MUX.

MUX, performing the best under most mobility conditions, serves to be the optimal base strategy of operation. However, diversity's ability to provide increased communication ranges should be used in tandem with MUX to alleviate its vulnerability to route failures.

5.4.3 Link quality

In ad-hoc networks, due to the multi-hop nature of the flows, packets that are received in error and dropped (due to time-correlated fading) closer to the destination, end up wasting bandwidth resources utilized along the path. This leads to a degradation in the end-end throughput.

In order to determine the impact of wireless channel effects in isolation, we do not consider mobility in these evaluations. The results for the different strategies are plotted as a function of the packet loss rates in Figures 12(a) and (b). It can be seen that both MUX and DIV-RANGE suffer degradation in throughput with increasing packet error rates since neither of them is targeted towards increasing the reliability of the link. DIV-BER increases the reliability of the link, albeit at the cost of rate as in DIV-RANGE. The result in Figures 12(a) and (bG) indicate that DIV-BER provides a steady performance with increasing packet error rates. This is because the decrease in BER resulting from the diversity order of 4 (2 elements) is sufficient to handle packet error rates as high as 30% easily. DIV-BER shows better performance than MUX at packet error rates higher than 20%, although it suffers at lower packet error rates.

Hence, the appropriate strategy would be to use MUX as long as packet error rates are negligible; have mechanisms to detect persistent channel errors and switch to DIV-BER to increase reliability.

In summary, we observe that our conclusions from Section 37 hold good with respect to lossy and non-lossy links in static networks with high density. In addition, we also find that DIV-RANGE is a beneficial option to be considered in the presence of low network density, high mobility and small loss rates, with DIV-BER being considered for significant loss rates.

5.5 MIR Routing Protocol

In this section we present the MIMO routing protocol called *MIR*, which is an on-demand (reactive) routing protocol. The goal of MIR is to exploit the benefits of the different MIMO strategies in an optimal manner to provide the best network performance. In the previous section, we saw that no single strategy performs the best over different network conditions of varying density, mobility and link quality and also identified the optimal combination of spatial multiplexing and diversity strategies for the different network conditions. MIR is thus an adaptive routing protocol that adapts between the different strategies based on the network conditions and does so in a transparent fashion. Note that, switching between strategies at a node can be achieved through software adaptation without additional hardware circuitry. However, in realizing the adaptation, there arise several challenges which are outlined below followed by the description of the protocol itself.

5.5.1 Challenges

Since a specific combination of strategies is required for different network conditions, the challenges that arise in each of them are also varied and are hence considered separately.

5.5.1.1 Density

We had seen that while MUX is the appropriate strategy for dense networks, DIV-RANGE is a more favorable option for sparse networks where it might not be possible to obtain routes with omni-directional communication ranges. Furthermore, not all regions of the network have the same density of nodes; while some regions might be densely populated, others may be sparsely populated. Also, the goal is to obtain routes that support high rate and allow for maximum spatial reuse in the network. This indirectly means that the number

of range links in the route should be kept to a minimum. So the key question is that, *how should the combination of the two strategies be used in the determination of routes?*

The simplest way would be to issue a route request using the MUX scheme. The routes discovered in this scheme would consist of all rate links. However, if such a route does not exist, the next route request can be issued using DIV-RANGE. There are two problems with this scheme when there exists no route to the destination with pure rate links; (i) we would be unnecessarily doing a MUX based route request while we could have directly done a DIV-RANGE based route request. This increases the delay in discovering a route, which in turn is dependent on the route request timeout value (minimum of $2\frac{T}{K}h$, where T is the omni-directional transmission time for a packet, h is the average hop length and K is the number of elements); and (ii) the resulting route will be one with predominantly range links and not the one that provides maximum spatial reuse and high rate. Thus, the main challenge here is to obtain the route that provides maximum spatial reuse and rate while at the same time trying to keep the route discovery latency as close as possible to that of the MUX scheme.

5.5.1.2 Mobility and Link Quality

We have seen that diversity strategy by virtue of using increased communication ranges (as in DIV-RANGE) or increased link reliability (as in DIV-BER), reduces the probability of link failures due to mobility and channel degradation respectively. Hence, MIR's goal is to detect link breakages due to mobility and channel degradation proactively and switch from multiplexing to diversity. The increased communication range or the increased reliability will be exploited to increase the longevity of the link during mobility and channel degradation respectively, which can in turn be used to find an alternate path even before the link breaks. However, there are several associated challenges. It becomes necessary to (i) identify and react to link losses resulting from mobility and persistent channel degradation since switching to diversity for a contention loss would only reduce rate unnecessarily, (iii) prevent unnecessary route switches due to short-term (transient) mobility and channel

degradation, (iii) determine the appropriate number of elements to be used towards diversity, since the additional elements used would only further reduce rate without any benefits, and (iv) ensure that an alternate route is obtained before the existing route actually fails.

5.5.2 MAC Layer Support

The routing protocol in isolation or in co-ordination with the MAC protocol, decides on the strategy to be employed by the node based on the network conditions. However, it is the MAC protocol that is responsible for communicating the decision to the PHY layer and ensuring that the communication takes place using the appropriate strategy for maximum benefits. Irrespective of the strategy employed, the PHY layer receiver must be aware of the strategy being used by the transmitter. Only then will it be able to decode the packet using the appropriate decoding strategy. This information exchange is achieved with the help of MAC layer. The MAC transmitter always transmits the preamble of the packet using diversity (STBC) in which it conveys the strategy to be employed for the actual packet transmission. Since a route can be comprised of both rate (multiplexing) and range/reliable (diversity) links, a receiver cannot determine which strategy to receive a packet on apriori. This is because, the receiver does not know which of its neighboring transmitters is going to transmit and also the strategy that is going to be used by the transmitter. To overcome this problem, the transmitters and receivers use STBC as the invariant strategy for the preamble in order to account for both rate and range/reliable links. Further, the preamble would convey the actual strategy that will be used by the transmitter for the rest of the packet. In order for the receiver to be able to estimate the channel and use it in its receiver processing, a training sequence is added to the front of the preamble.

Another important role of the MAC layer is to determine the minimum range extension factor with which a packet can be received on a range link. This would help the routing protocol in choosing routes with maximum spatial reuse as we shall see later in this section. Recall that the function f that characterizes range extension factor is linear in the number of elements. Thus, when a packet is transmitted through DIV-RANGE using K elements, it can reach upto K hops. However, at the end of the packet, the node adds $K-1$ short

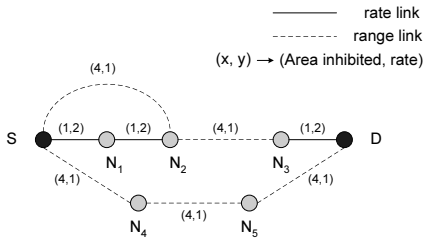


Figure 13: Illustration

preambles, each being transmitted-received at unity rate similar to the packet but using lesser number of elements ($[1, K-1]$) and hence lower diversity order. These preambles correspond to the different $K-1$ range extensions and carry the respective extension information in them. Only a node within the i^{th} hop will be able to decode the preamble that was transmitted-received using i elements with a diversity order of i^2. By this method, each node keeps track of the minimum range extension ([1,K]) with which the packet was received from the upstream node and stamps it on the packet to be used by the routing layer.

5.5.3 Routing Protocol Components

The routing strategies in MIR apply to on-demand (reactive) routing protocols that involve route discovery (involving route request and response phases) and maintenance phases. While the theoretical and practical inferences drawn from Sections 37 and 5.4 apply to proactive routing protocols as well, the design of the routing mechanisms themselves have to be tuned in order to be applicable. We present MIR as a source-driven reactive routing protocol (such as DSR). However, the protocol components are also applicable to table-driven reactive protocols such as AODV, etc. with minor modifications. Route discovery and maintenance form the two main components of a reactive routing protocol, under which we consider the following five components, namely, route metric, route request, route response, route failure detection during mobility and channel degradation, and route maintenance.

5.5.3.1 ***Route Metric***

The goal of the route discovery component is to obtain "quality" routes through a careful use of the different strategies irrespective of the density in the network. We define "quality"

of a route as its ability to allow for maximum spatial reuse in the *network* while at the same time incurring low multi-hop relaying burden and providing high rate for the *flow* itself. Hence, we consider a two-tuple route metric $Q(R)$ (=($Q_n(R)$,$Q_f(R)$)) for a route R as the combination of network ($Q_n(R)$) and flow ($Q_f(R)$) metrics. This is similar to the routing metrics considered in the widest-shortest routing paradigm in the Internet. We consider $Q_f(R)$ to be the minimum of the rates used by the strategies employed on the links in the path. $Q_n(R)$ can be captured by the total area inhibited (along with duration of inhibition) when a packet travels *along the path* from source to destination. Using $Q(R)$, routes are chosen lexicographically, based on a low network metric; if the network metrics are the same, a high flow metric is used. Though $Q(R)$ biases towards the network metric, optimizing $Q_n(R)$ indirectly means that the number of range links as well as the number of hops in the route should be kept low. This, in turn also indirectly favors the flow in obtaining a higher rate on the path. The choice of the routing metric itself is not closely tied to the routing strategy and is hence open to further research.

$$\begin{aligned} Q_n(R) &= \sum_{i=1}^{h} \frac{Area\ inhibited\ by\ i^{th}\ hop\ transmission}{Rate\ of\ transmission\ at\ i^{th}\ hop} = \sum_{i=1}^{h} \frac{f_i^2}{r_i} \\ Q_f(R) &= min\{r_i\} \end{aligned} \tag{27}$$

where f_i and r_i are the range extension factor and rate of transmission at the i^{th} hop of a h hop route respectively, both being normalized to the corresponding values of an omni-directional transmission. Thus, as long as a "quality" path to the destination exists with all rate links or minimal range links, it should be determined by the routing protocol. In the simple topology shown in Figure 13 where two elements (extension factor of two) are used, we have three possible routes from source S to destination D, namely, $\{S, N_1, N_2, N_3, D\}$, $\{S, N_2, N_3, D\}$ and $\{S, N_4, N_5, D\}$. The two-tuple metric for the 3 routes are (5.5,1), (8.5,1) and (12,1) respectively. Hence, while several routes exist, the routing protocol should identify the route with the best metric (minimal impact from range links), namely $\{S, N_1, N_2, N_3, D\}$ in our example that has a single range link.

MIR uses a route discovery procedure whose two main components are (i) propagation of route request that keeps the route discovery time small, and (ii) route reply propagation

and route selection that help obtain quality routes.

5.5.3.2 ***Route Request***

Every node on receiving a route request packet, goes ahead and propagates the request on the maximum range possible using DIV-RANGE. Whenever a route request is transmitted using DIV-RANGE, it is done on all elements (K), and can thus reach upto K hops. With support from the MAC layer through the use of short preambles, each node keeps track of the minimum range extension ([1,K]) with which the request was received from the upstream node. Whenever a node forwards a request, in addition to adding itself to the source route, it also adds the extension factor of the link with its upstream node. The source route carried on the request packets will thus "initially" consist of predominantly range links (eg. $\{S, N_2, N_3, D\}$ in Figure 13). However, when the destination receives the request packet with a range factor of say i ($i\epsilon[1, K]$), it waits for approximately ϕiT seconds (where T is packet transmission duration on an omni link, ϕ is a constant that accounts for the additional time taken to access the channel due to contention) to help the nodes in the last set of i hops including itself to bridge/patch their local upstream range links with rate links. The time spent by the destination in waiting coupled with route response propagation time is sufficient to helps the other upstream nodes to have their local upstream range links bridged (replaced) with rate links if permitted by the underlying topology (eg. $\{S, N_2\}$ can be bridged as $\{S, N_1, N_2\}$). The bridging is achieved as follows. Each intermediate node (say N_2) after forwarding a route request keeps track of the reverse path to the source along with its metric ($\{N_2, S\}$ with metric (4,1)). Later on, if it receives another request with a much better route back to the source ($\{N_2, N_1, S\}$ with metric (1,2)), then it replaces the local route it has stored for the source with the better route.

5.5.3.3 ***Route Response***

When a route reply finds its way back to the source using the information on the source route along with the extension factors to be used on the different links, it is checked by every intermediate node along the path. If the locally cached route to the source has a better route metric than that currently being used, then it replaces the portion of the route

from itself to the source with the locally stored route. Thus, when the route reply finds its way back to the source, the nodes update the route in the packet with their locally cached route that is made up of minimal range links as determined by the metric. Hence, when the source finally gets back the route reply, the route contained in it consists of all rate links if such a quality route exists or with the minimal impact from range links otherwise. When a flow is considered in isolation, the worst case delay incurred by MIR in getting the request delivered to the destination is $T \cdot (l + K - 1)$ neglecting the short preambles, with $T \cdot (l - 1)$ constituted by the propagation of request until the last set of i hops and TK constituted by the worst case waiting time at the destination to allow for local bridging; $l = \frac{h}{K}$, h is assumed to be an integral multiple of K without any loss of generality. This is reasonable compared with the best case delay of Tl seconds in MUX. The route reply propagation would however incur the same latency in both cases.

MIR is thus able to obtain quality routes with the minimal impact from range links at the cost of a small additional delay. The main benefit of this mechanism is that, it does not require the routing protocol to be aware of the network density in order to be able to change strategy in determining routes but can obtain the quality routes provided by the underlying topology for any given density in a transparent fashion. Also, by helping obtain routes with high rate (predominantly rate links), it provides more potential to alleviate link breakages due to mobility and channel effects by switching to diversity techniques, which are explained subsequently.

While the sparse nature of the network is a steady-state feature of the network, mobility and channel errors are network dynamics that need to be handled by the routing protocol through its route maintenance component to avoid throughput degradation.

5.5.3.4 ***Route Failure Detection during Mobility and Channel Degradation***

The challenge is to detect that a link is going to break due to mobility or persistent channel degradation. The MAC in MIR addresses this challenge by switching from MUX to diversity after four trials of the RTS packet (RTS has a maximum retry limit of seven in IEEE 802.11b). However, depending on the nature of the loss (mobility or channel degradation),

the gain in SNR automatically provides increased range for a far-by receiver (DIV-RANGE), or increased reliability for a close-by receiver (DIV-BER) to appropriately recover from the loss. Hence, it is not necessary for us to differentiate between mobility and channel error losses which is a very useful feature. When switching to diversity, the number of elements used towards diversity is initially two, since in most cases the range extension or increased reliability resulting from a diversity order of four is easily sufficient to sustain the link from breaking even at high speeds (30 m/s) and high packet error rates (30-40%). This keeps the reduction in rate and spatial reuse (if increased communication range is exploited) due to diversity to a minimum and exploits the remaining elements to increase the rate of transmission through MUX. It also keeps the overhead due to short preambles very small. If the link is already operating in diversity, then an increase in number of elements exploited for diversity by one will still serve the purpose.

If the increased communication range or reliability due to diversity is able to get the packet across, then there are two possible cases: (i) receiver was actually moving away and/or the channel quality was bad; or (ii) neither the receiver was moving away nor the channel quality was bad. The second case is a false alarm (negative) but can be handled easily. The change of strategy is known to the receiver through the normal preamble that is always transmitted using diversity (STBC). Once the receiver receives the RTS, it sends back the CTS using the same strategy used by the transmitter. However, the transmitter on receiving back the CTS packet can determine the false alarm (assuming symmetric channel conditions) if it can successfully decode the short preamble that was transmitted using a lesser (previously used) diversity order than that used by it on its RTS. On the other hand, if the transmitter cannot decode short preambles of lower diversity orders, then a "proactive" route error can be generated confidently. However, when the extended communication range or increased reliability is also not able to get the packet across, we again have two cases: (i) neither the receiver was moving nor the channel quality was bad but there is contention; or (ii) receiver was moving and/or channel quality was bad and there is contention. In both cases, we would appropriately infer congestion and would not generate a proactive route error.

5.5.3.5 ***Route Maintenance***

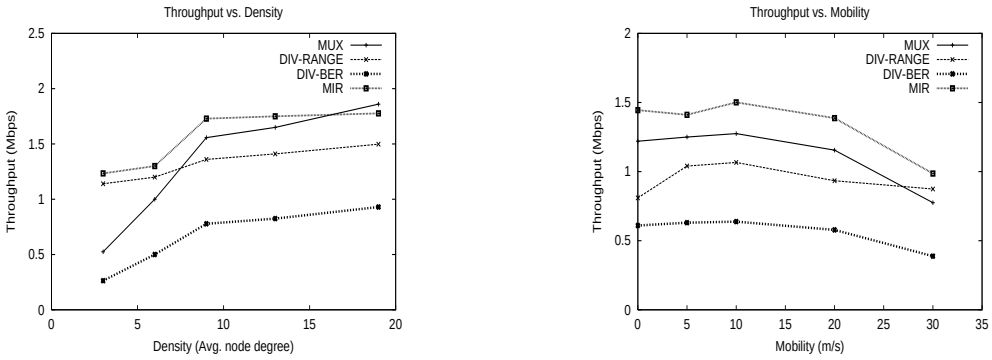
Now, once the link failure due to mobility or persistent channel errors has been "proactively" detected, the routing protocol at the node is informed. MIR at the detection node recognizes the proactive link failure and hence does not purge packets from the queue that are using the link. It generates a proactive route error to the source. The intermediate nodes delete routes from their caches (but do not purge packets) that contain the detected link as in a normal route error so that they do not respond back to the new proactive route discovery process with the stale route. Once the source receives the proactive route error, it initiates a new route request for the destination to obtain a better route with minimal impact from range links through the route discovery component. However, it does not purge the routes that contain the detected link until a new route is obtained or an official route error is received and uses the existing route for sending the packets currently in its buffer. The MAC at the node that detected the proactive route error, though it has switched to diversity to enable communication on the link, it continues to keep track of the diversity order that is supported by the link in conjunction with its receiver through the short preambles. Once it finds that the diversity order used prior to switching can be supported again by the link within four transmissions in succession, it generates a "route error cancel" notification to the source to prevent the unnecessary route change. This would automatically take care of the case wherein the receiver moves out of the current communication range and then comes back into it within a very short duration, and also the case where the channel degradation is short-term. This is because when the receiver moves back in or the channel quality becomes good, the transmitter would receive the short preamble corresponding to a lower diversity order.

When a new route is obtained, the source switches its route to the new route and purges the old routes that contain the detected link. The reaction to an official route error is the same as in a conventional source-driven reactive protocol such as DSR. Further, the granularity of the time taken for a link to break due to mobility or persistent channel errors under increased diversity is large when compared to the route discovery latency. This is because the route discovery component in MIR obtains a route in a time close to that of

the high rate MUX scheme. Hence, the proactive route is obtained before the detected link actually fails. Thus, proactive switching from MUX to diversity helps reduce the number of route errors due to mobility and persistent channel errors that occur in time-correlated fading, thereby reducing the degradation in throughput, while also addressing all the challenges outlined in Section 5.5.1.2. Further, exploiting diversity is different in principle from the typical topology control algorithms where transmission power is used to control the communication range and increase the link reliability, since it does not require any change in transmit power and hence does not complicate the MAC protocol operation.

In summary, the components in MIR exploit multiplexing to increase the rate of a flow but intelligently use diversity in their route discovery and maintenance components based on the perceived network conditions to obtain quality routes and to prevent degradation in throughput. The routing protocol components in MIR can also be applied to table-driven routing protocols such as AODV, whose details can be found in [62].

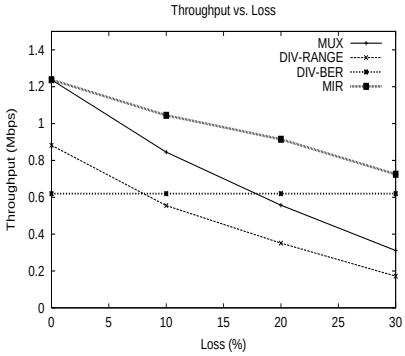
5.6 Peformance Evaluation



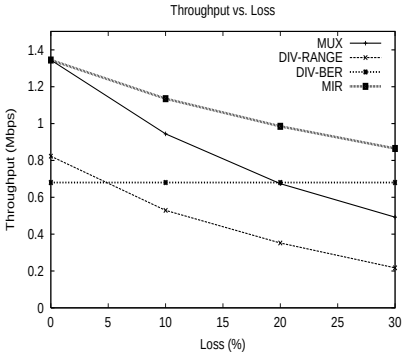
a) Density b) Mobility

Figure 14: Individual Impact of Components on Aggregate Throughput (1)

In this section, we evaluate the performance of MIR against MUX, DIV-RANGE and DIV-BER strategies. We evaluate the different strategies in the *ns2* simulator. The density, load, mobility and loss characteristics of the link are varied from one experiment to another. We assume the range extension function f to be linear. The constant ϕ in the route discovery component is empirically set to 5. The number of elements is fixed at two initially for fair



a) Loss (Load = 10 flows)



b) Loss (Load = 15 flows)

Figure 15: Individual Impact of Components on Aggregate Throughput (2)

comparisons between the different strategies since we do not want the range extension resulting from DIV-RANGE to be limited by the topology size. We also do not want the usage of additional elements for increasing reliability in DIV-BER when the existing elements themselves have reduced the link error probability to a negligible value. Later on, we do consider the impact of number of elements on the different strategies. We represent density of the network by the average node degree parameter ρ. ρ is varied from 3 (sparse networks, 100 nodes in 2500m*2500m) to as high as 19 (dense networks, 100 nodes in 1000m*1000m) with the transmission range being 250m. Mobility and link loss rates are varied upto 30 m/s and 30% respectively. The default values of load, mobility and loss rate are 30, 0 m/s and 0% unless varied or specified otherwise. Each flow uses CBR as the traffic generating application at a rate of 100Kbps on a 2 Mbps channel with a packet size of 1KB and UDP serving as the transport protocol. MIR is implemented by effecting the necessary changes to the DSR routing protocol, with support from IEEE 802.11b MAC protocol in DCF mode using standard specifications. In addition to the two-ray ground propagation model supported in *ns2*, we incorporate the impact of time-correlated Rayleigh fading on packet errors through a new collision model. The collision model captures the probability of packet errors for various configurations (locations and MIMO strategies used) of desired transmitters and interferers in the presence of time-correlated Rayleigh fading. This is in turn derived from the BER statistics obtained from bit-level Matlab simulations of detailed physical layer modeling of spatial multiplexing and diversity in the presence of Rayleigh

fading with time correlation. The scenarios are generated using the random waypoint mobility model. Aggregate throughput is the primary metric of comparison and each of the data point in the results presented is averaged over 10 seeds with each seed running for 100s.

5.6.1 Individual impact of components

5.6.1.1 Varying density

Figure 14(a) presents the aggregate throughput results for the different strategies as a function of node density (average node degree) for a load of 30 flows. No mobility of nodes or losses on the links are considered. It can be seen that irrespective of the density, MIR is able to track the performance of the best strategy, which is DIV-RANGE in the sparse network case and MUX in the dense network case.

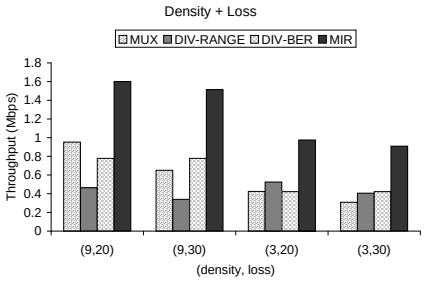
5.6.1.2 Varying mobility

In these experiments, the load in the network is maintained at 15 flows and the links are assumed to be almost non-lossy. The mobility is varied till 30m/s and the results are presented in Figure 14(b). Recall that, DIV-RANGE is outperformed by MUX at lower speeds due to the increased contention experienced from extended range that outweighs its robustness to mobility. MIR delivers the best performance since it emulates the MUX scheme while at the same time increases robustness by reducing the number of route errors experienced by an intelligent use of diversity.

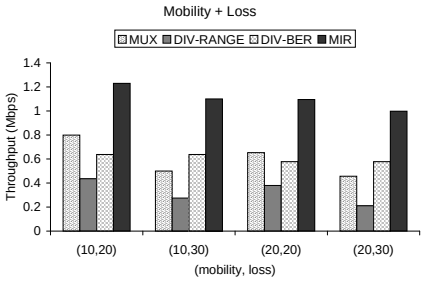
5.6.1.3 Varying loss characteristics

The loss rates on the links are varied in this set of experiments from 0% to 30%. The nodes are assumed to be static and the network size considered is a dense environment with 100 nodes in 1000m*1000m. Figures 15(a) and (b) present the results where aggregate throughput is recorded as a function of loss percentage for two different loads of 10 and 15 flows respectively. MUX is able to outperform DIV-BER only at low error rates (10%) since the degradation due to losses in MUX is not significant when compared to the rate that is sacrificed in DIV-BER. DIV-RANGE not only suffers from reduced rate as in DIV-BER,

but also from the decreased link reliability due to longer links, thereby exhibiting poor performance. In both these cases, MIR is able to adapt the MUX scheme to incorporate DIV-BER, thereby increasing the robustness of the link based on the link conditions and is hence able to deliver performance improvement.

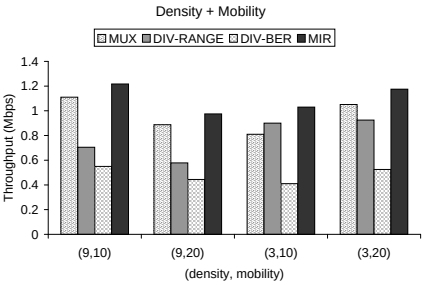


a) Density + Loss (No mobility)

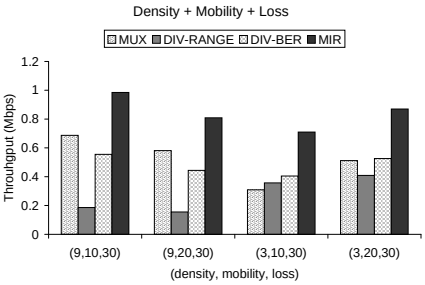


b) Mobility + Loss (density = 19)

Figure 16: Combined Impact of Components (1)



a) Density + Mobility (No loss)



b) Density + Mobility + Loss

Figure 17: Combined Impact of Components (2)

5.6.2 Joint impact of components

In the following sets of experiments the load in the network is fixed at 15 flows, each flow having a rate of 100 Kbps.

5.6.2.1 Density + loss

We now evaluate the joint performance gains of the route discovery and link error components in MIR. We assume static nodes but consider densities of 3 and 9 and loss rates of 20% and 30% in four possible combinations. The results for the different schemes are presented in Figure 16(a). First let us consider MUX, DIV-BER and DIV-RANGE alone for comparison. At higher densities and lower loss rates, MUX performs the best while DIV-BER

suffers from reduced rate and DIV-RANGE suffers from reduced rate as well as reduced spatial reuse. However, at higher loss rates (30%), DIV-BER performs the best. DIV-RANGE exhibits the best performance in lower densities and lower loss rates. In the case of lower densities but higher loss rates, the choice could be between DIV-BER and DIV-RANGE depending on how the benefits of increased range and increased reliability outweigh the respective drawbacks in the two cases. Finally, MIR incorporates the advantages of MUX at lower loss rates and higher densities, the advantages of DIV-BER at higher loss rates and the benefits of DIV-RANGE at lower network densities in a transparent manner to provide significant gains of about 100%.

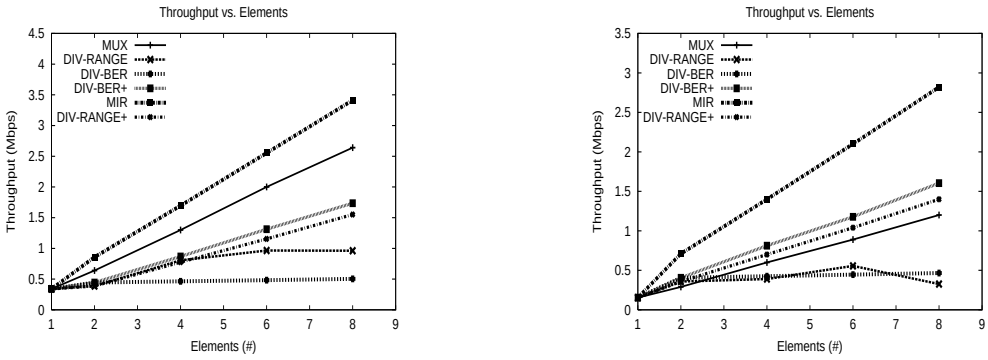
5.6.2.2 Mobility + loss

Figure 16(b) presents the performance gains of the mobility and link error components of the different strategies. Considering MUX, DIV-BER and DIV-RANGE alone initially, it can be seen that MUX provides higher gains than the other two strategies for lower speeds and lower loss rates. At higher loss rates, as expected DIV-BER delivers better performance. However, the magnitude of gain is decreased. This is because link (channel) errors which might be recovered by DIV-BER in the static case, might well be lost in the combined presence of mobility and channel errors. The benefit of DIV-RANGE in mobile scenarios appears only at large speeds close to 30 m/s. For the scenarios considered here, the increased communication range (increased contention) and loss, and reduced rate and spatial reuse in DIV-RANGE all contribute to its worst performance. Finally, MIR combines the benefits of increased reliability (DIV-BER) for handling link errors and increased range (DIV-RANGE) for handling mobility gracefully, with MUX serving as its base strategy of operation to provide gains close to 100%.

5.6.2.3 Density + mobility

Figure 17(a) presents the results for the joint performance of the route discovery and mobility components in MIR. It can be seen that MUX performs better than DIV-RANGE for higher densities. The robustness of DIV-RANGE to mobility over MUX does not contribute much when compared to the reduction in spatial reuse and rate at higher densities.

While DIV-RANGE does perform better than MUX at lower densities and low mobility, it is interesting to note that MUX performs better than DIV-RANGE at lower densities of $\rho = 3$ in the presence of high mobility. This can be explained as follows. It can be seen from the result that increasing mobility from 10m/s to 20m/s actually helps MUX obtain better performance at low density. This is possibly due to the fact that flows that have no routes in static (or low mobility) sparse network condition, may be able to form routes due to high mobility. Hence, the benefit of DIV-RANGE to be able to form routes that do not exist in MUX is decreased. Furthermore, the gain of MIR over MUX is also only moderate. We would have expected MIR to track the performance of MUX or DIV-RANGE depending on the density and also provide the benefits of diversity for handling mobility. But as pointed out before, mobility has not degraded MUX's performance significantly. In fact it has aided MUX at lower densities. Hence, the gain of the mobility and route discovery components in MIR is decreased in this case. However, the gain is still more than 25%. Finally, the gains provided by all the components in MIR for different combinations of (density, mobility, loss) have also been considered and the results are available in [62] with a representative result presented in Figure 17(b).



a) (9,20,20) b) (3,10,30)

Figure 18: Impact of # Elements

5.6.3 Impact of elements

Figures 18(a), and (b) present the results for the different strategies for different combinations of (density, mobility, loss), with elements being the parameter of variation. These results give us an idea of the ability of the different strategies to scale with the number of

antenna elements. Recall that for a fixed loss rate in the network, increasing the number of elements beyond a certain value may not bring any additional gains in DIV-BER if the BER's are already significantly reduced on the links. However, a loss in rate would occur. Similarly in the case of DIV-RANGE, for a fixed network density, increasing the number of elements beyond a certain value may not help. In fact, it might degrade performance due to the reduced spatial reuse resulting from increased communication range in addition to the loss in rate compared to MUX. Hence, in addition to MIR, MUX, DIV-BER and DIV-RANGE, we have also considered two other strategies namely, DIV-BER+ and DIV-RANGE+ that are combinations of DIV-BER and MUX, and DIV-RANGE and MUX respectively. We consider two elements at a node being devoted for diversity and use the rest of the elements for spatial multiplexing in these two cases.

It can be seen from the results that DIV-BER and DIV-RANGE are not able to scale with the number of elements as expected unlike DIV-BER+ and DIV-RANGE+. DIV-BER and DIV-RANGE are able to outperform MUX at higher loss rates and lower densities only when number of elements is low (two), but suffer from rate loss at higher number of elements. On the other hand, DIV-BER+ and DIV-RANGE+ are able to sustain the gains over MUX even for higher number of elements under similar conditions. This suggests that *diversity techniques when needed, should be employed in tandem with multiplexing, such that only as many number of elements contribute to diversity as required for the purpose.* Finally, MIR exhibits the best scalability properties amongst all strategies considered.

CHAPTER VI

UNIFIED FRAMEWORK FOR MEDIUM ACCESS CONTROL AND ROUTING IN AD-HOC NETWORKS WITH SMART ANTENNAS

6.1 Overview

While the previous chapters presented MAC and routing protocol design with MIMO links, this chapter will extend the optimization considerations identified to help develop unified MAC and routing frameworks for smart antennas in general.

MIMO links are the most sophisticated of the various antenna technologies thus far. They are capable of adaptive resource usage, flexible interference suppression and also provide several gains such as multiplexing, array and diversity gains. Other antenna technologies such as omni-directional, switched beam and adaptive arrays are capable of delivering only a subset of the capabilities and gains provided by MIMO links. Hence, the optimization considerations needed to leverage the potential benefits of MIMO links as well as the MAC and routing protocols proposed for MIMO links in Chapter 4 and 5 respectively can be easily extended to generalize the problem of medium access control and routing with smart antennas. In this direction, unified MAC and routing frameworks for solving the MAC and routing problems respectively in ad-hoc networks with the different class of antenna technologies is presented in this chapter.

The chapter begins by systematically identifying the different physical layer capabilities of the different antenna technologies as well as their corresponding MAC layer considerations under specific categories. This paves the way for the design of a unified framework for solving the MAC problem. The framework is then used for deriving both centralized and distributed MAC solutions for the different antenna technologies which are eventually evaluated under varying network conditions. The chapter then shifts focus to the design of

a unified distributed routing framework for smart antennas in general. Unified design rules for the operation of the different strategies common to different smart antenna technologies are identified and then incorporated in the design of routing components of a reactive routing protocol.

6.2 Cross-layer Considerations

In this section, we first describe, for each of the antenna technologies, key PHY layer characteristics that are relevant to the MAC layer design. We present the details in terms of the communication pattern, potential gains and interference suppression capabilities. We then highlight key optimization goals that need to be accounted for in the MAC layer design in terms of leveraging gains, resource allocation and utilization, and scheduling. The problem formulation and algorithms presented in the later sections, accommodate these optimization considerations.

6.2.1 Switched Beam Antennas

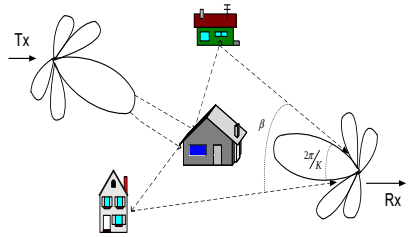


Figure 19: Switched Beam Illustration

Relevant PHY layer characteristics

Communication pattern: Switched beam antennas rely on directionality for their communication, where the transmitter and the receiver are capable of forming directional beams towards each other (Figure 19).

Potential gains: The ability of the switched beam antennas to concentrate energy in a particular direction, provides an increase in SNR with respect to the desired signal resulting in a *directional gain* G_d. This gain can be bounded by,

$$G_d = K^2 \tag{28}$$

in the case where both the transmitter and receiver knowing the direction of transmission to each other [74]. K represents the number of elements at either of the ends (assumed to be the same at both ends for simplicity) and p represents the path loss component. However, the above equation is true only for an LOS environment. In the case of multi-path environments which are characterized by rich scattering, the bound on the gain is determined by the scattering angle as the number of elements at both the ends increases [74]. Now, if the number of elements at both the ends of the link is assumed to be K, and if the scattering angles at the transmitter and receiver are α and β respectively, then

$$\begin{aligned} G_d &= G_t G_r,\ where \\ G_t &= K,\ if\ \frac{360}{\alpha} \geq K \\ &= \frac{360}{\alpha},\ if\ \frac{360}{\alpha} < K \\ G_r &= K,\ if\ \frac{360}{\beta} \geq K \\ &= \frac{360}{\beta},\ if\ \frac{360}{\beta} < K \end{aligned} \tag{29}$$

The scattering angle saturates the gain beyond a certain number of elements because for higher number of elements, the antenna gain resulting from the increased elements is compensated by the loss of energy outside the main beam but within the scattering angle. Unlike in cellular environments, where the higher elevation of the base station helps keep the scattering angle low, ad-hoc networks - especially in an indoor setting - will be characterized by large scattering angles [74].

Interference suppression: These antennas do not provide flexible interference suppression owing to the use of pre-determined beam patterns. While they are considered to suppress interference along the non-active beams, the presence of side lobes contributes to accumulation of noise along these non-active beam directions and hence brings the down the SNR of the desired signal.

MAC layer considerations

Leveraging gains: The ability to focus energy in specific directions increases the spatial

reuse in the network, thereby increasing the number of simultaneous parallel transmissions. We know that the directional gain represents the increase in SNR over that of an omni-directional element. But *how does this gain help improve the network performance?* The directional gain can be used in one of two ways: (i) for a given probability of error on the link, the gain can be used to increase the range of transmission/reception. This would in turn reduce the number of hops for multi-hop flows and hence increase the throughput. However, the interference range of a single transmission will also be increased, and the degree of spatial reuse in the network thereby reduced. The range extension factor (R_f) can be bounded by ([74]),

$$R_f = (G_t G_r)^{\frac{1}{p}} \tag{30}$$

and (ii) the gain can be used to increase the capacity of the link (with no obvious negative side-effect). We know that the basic Shannon's upper bound on channel capacity (normalized to bandwidth) is given by

$$C = log_2(1 + \rho) \tag{31}$$

Now, when the average SNR (ρ) is increased by a factor G_d, there is a relative logarithmic increase in capacity (C_f) to C' which can be approximated as,

$$C_f = \frac{C'}{C} \approx 1 + \frac{\log_2 G_d}{\log_2 \rho} \tag{32}$$

This increase in turn is achieved by means of using higher modulation rates for transmission/reception, while providing the same probability of error. In effect, more bits are transmitted per symbol (than in the case without the directional gain) while ensuring the same probability of error on the link.

Resource allocation and utilization: Every node has a single resource (for transmission) which simplifies resource allocation and utilization to the binary decision of gaining access to the channel or not.

Scheduling: The availability of a single resource eliminates the necessity to perform intelligent scheduling of links to leverage the advantages of location based contention prevalent in shared wireless channels [37].

6.2.2 Fully Adaptive Array Antennas

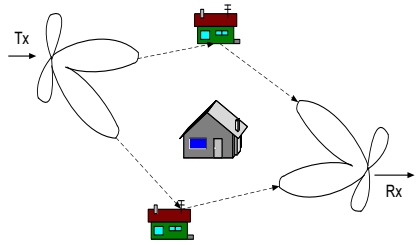


Figure 20: Adaptive Arrays Illustration

Relevant PHY layer characteristics

Communication pattern: As the name suggests, these antennas are capable of adapting their beam pattern to maximize the SNR of the desired signal. Hence, the notion shifts from *directionality* to *availability of resources* for communication. Each transmitter with K elements, uses a single resource to transmit the signal, while using the remaining $K-1$ resources (DOFs) for suppressing other transmissions in its neighborhood (Figure 20).

Potential gains: The maximization of SNR achieved through the use of maximal ratio combining (MRC) at the receiver results in an *array gain*. Assuming that CSI is available at both the transmitter and receiver, the resulting array gain G_a can be bounded by ([2]),

$$G_a = MN \tag{33}$$

where M and N are the number of antennas at the transmitter and receiver respectively. While the gain from adaptive array antennas does not degrade with an increase in the degree of multi-path (scattering angle) unlike switched beam antennas, yet, when the angular spreads are significantly large at the transmitter and receiver, the low correlation existing between the different signal components bounds the gain as ([2]),

$$G_a = (\sqrt{M} + \sqrt{N})^2 \tag{34}$$

In the case when all the DOFs are not used towards suppressing interference, then the remaining DOFs can be used to provide diversity gain. The diversity gain reduces the probability of the signal experiencing deep fades. While the array gain relates to the increase in the mean of the SNR at the output of the combiner, the diversity gain relates to the

reduction in the variance of the SNR at the output of the combiner, relative to its variance prior to combining. The reduction in variance depends on the diversity order, which in turn depends on the degree to which the multi-path fading on the different antenna elements is uncorrelated. The maximum diversity order afforded by a link with M transmit antennas and N receive antennas is MN. While array gain continues to grow as more antennas are added, diversity gain tends to saturate [19].

Interference suppression: Adaptive array antennas are capable of flexible interference suppression. While the number of DOFs (resources) required for suppressing interference depends predominantly on the number of interfering streams transmitted, the strength of the interference and their spatial correlation also plays a notable role. Weak and highly correlated interference signals may not require as many DOFs to be sacrificed for suppression as the strong ones, thereby helping the node potentially accommodate more transmissions in its neighborhood.

MAC layer considerations

Leveraging gains: As in the case of switched beam antennas, the array and diversity gains of adaptive arrays can be used toward range extension or capacity increase. However, if the resources at a node are all used up in suppressing interference, then we can consider array gain to be the dominant contributor, in which case, we obtain the bound on range extension as [74],

$$R_f = (G_a)^{\frac{1}{p}} \tag{35}$$

Since the diversity gain deals with the variance in SNR (unlike the array gain that deals with the mean SNR), the relation between diversity gain and number of elements is not direct. However, a measure of diversity, namely the diversity order can be directly related to the number of elements. On the other hand, if capacity increase is desired, then the average SNR (ρ) can be increased by a factor G_a, thereby resulting in a relative logarithmic increase in capacity (C_f),

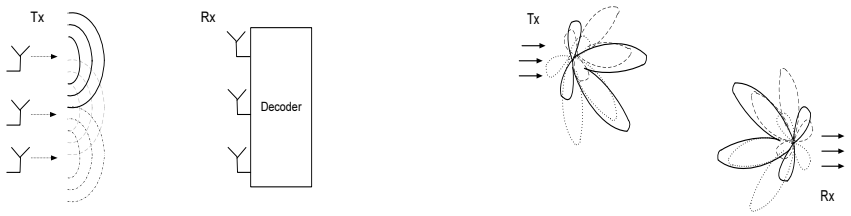
$$C_f = \frac{C'}{C} \approx 1 + \frac{\log_2 G_a}{\log_2 \rho} \tag{36}$$

Resource allocation and utilization: The channel can be visualized to be composed of

min(M,N) independent and parallel channels (called as "channel modes") in the presence of rich multi-path. The data sent on each of these channels is generally referred to as a *stream*. Each of these channel modes does not have the same gain and hence, whenever a transmitter has a resource for transmission, the strategy is to transmit the signal only on the strongest channel mode (single stream) with all its power.

Scheduling: When a link belongs to more than one otherwise non-overlapping *contention* regions (also termed *bottleneck link*), the total number of transmissions in all those contention regions is limited by the amount of resources at the bottleneck link which is active. On the other hand, if the bottleneck link is passive, then its resources can be *overloaded* to increase the number of transmissions in each of those contention regions (if possible), thereby increasing the utilization. We refer to this as the *passive receiver overloading* problem. Hence, careful scheduling of the bottleneck links is essential for increasing the aggregate utilization of the network. Note that, this problem is implicitly addressed in the case of omni-directional and switched beam antennas, since there can be only one transmission in every contention region.

6.2.3 MIMO Links



(a) OL-MIMO (b) CL-MIMO

Figure 21: MIMO Illustration

Relevant PHY layer characteristics

Communication pattern: Communication using MIMO links can also be modeled with the notion of "availability of resources". The difference from adaptive arrays is that the available resources (= number of elements, say K) can not only be used by receivers for suppressing interference, but can also be used by the transmitter to send multiple streams. The objective is still to maximize the SNR of the link. However, the decoding strategy

employed at the receiver involves significant complexity (Figure 21).

Potential gains: A MIMO link can provide three types of gain: array gain, diversity gain, and spatial multiplexing gain. If the modulation rates are assumed to be fixed, then array and diversity gains primarily provide range extension as in adaptive array antennas, while spatial multiplexing gain primarily provides higher data rates. While for a transmit array to provide either array or diversity gain, the data streams transmitted from the different antenna elements must be dependent, this is not true for spatial multiplexing. MIMO links provide spatial multiplexing gain in the presence of multi-path or rich scattering, by simultaneously transmitting independent data streams. This gain is defined as the asymptotic increase in the capacity of the link for every 3 dB increase in SNR [17].

Interference suppression: Similar to adaptive arrays, MIMO links are also capable of flexible interference suppression, with the difference being that the resources saved from suppressing weak and highly correlated interference signals may not only be used to accommodate more transmissions but may directly be used to increase the number of streams being transmitted as well.

MAC layer considerations

Leveraging gains: The spatial multiplexing gain can be achieved when the transmit array transmits multiple independent streams of data. When $M = N = K$, the capacity bound of a MIMO link is given by the following equation [19],

$$C \approx K \log_2(1 + \rho) \tag{37}$$

where ρ represents the average SNR at any one receive antenna. On the other hand, when the multiple antennas are used only for array and diversity gain, the asymptotic capacity is

$$C \approx log_2(1 + \rho') \tag{38}$$

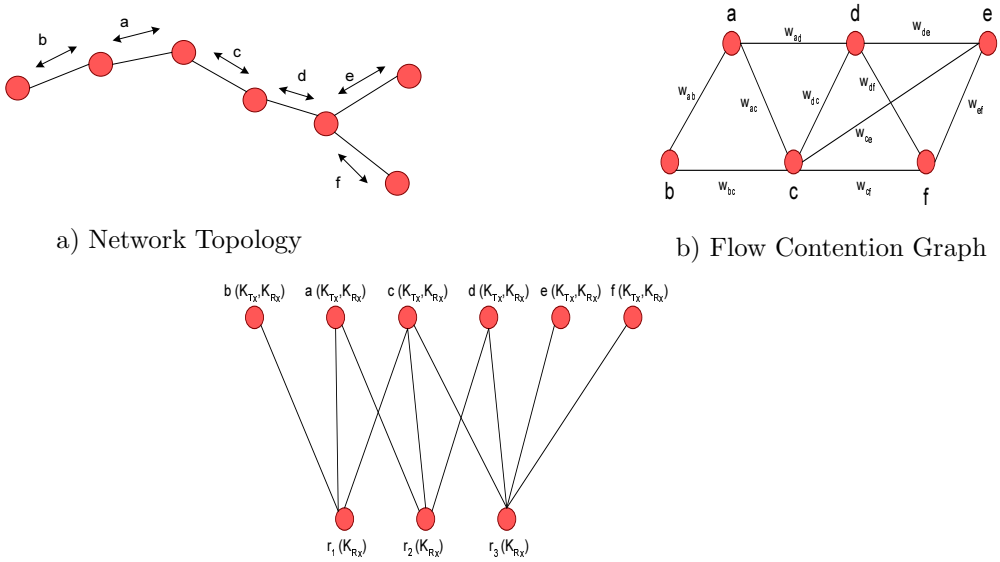
where ρ' is a random SNR with a mean that increases only linearly with the array gain and a variance that decreases with the diversity order. Therefore, capacity grows linearly with K with spatial multiplexing, but only logarithmically with array and diversity gain.

Resource allocation and utilization: Unlike adaptive arrays, MIMO links operate on several streams during spatial multiplexing. Note that the different channel modes in a MIMO link do not experience the same gain. Hence, when a link is allowed to use only a subset of the maximum possible number of streams (say m out of K), *it can distribute its transmit power over just the m strongest channel modes (streams).* Thus, when compared to two interfering links operating using TDMA at the maximum number of streams K, letting the links operate simultaneously but with the stronger $\frac{K}{2}$ streams will result in improving the overall utilization in the network. We term the gains achievable through such simultaneous operation of interfering but stream controlled links as *stream control gains* [12].

Scheduling: MIMO links also suffer from the "passive receiver overloading" problem. However, the problem manifests itself in the context of stream control. In a TDMA scheme using all streams ($= DOFs$) for transmission, if a *passive* receiver belongs to more than one otherwise non-overlapping contention regions, then there can be an active transmitter in each of those contention regions, overloading the passive receiver with more streams that it can receive. On the other hand in a MIMO MAC employing pure stream control, all the transmitters within a contention region use the best subset of their streams such that no receiver in the region is overloaded. But if any of the receiver nodes also belong to other contention regions, then this prevents the nodes in those other contention regions from transmitting since this will overload the active receiver. This in turn reduces the advantage of spatial reuse resulting in a potential performance degradation and must hence be addressed.

6.3 MAC Problem Formulation

A good way of looking at the problem of fair medium access is to first associate each flow with a utility function. The utility function represents the satisfaction measure of the flow corresponding to a parameter of interest. The parameter could be service (throughput), delay, etc. If the utility functions are assumed to be concave, then there exists a mapping between the maximization of the sum of all utility functions and a system-wide notion of



a) Network Topology

b) Flow Contention Graph

c) Resource Constraint Graph

Figure 22: Graphs

fairness. Furthermore, the nature of the utility function chosen determines the model of fairness achieved [25]. Hence, the problem of fair channel access equivalently reduces to the optimization problem of maximizing the aggregate utility of the network subject to the transmission constraints with the appropriate choice of the utility function. We adopt the *proportional fairness* model in this work, wherein a logarithmic (or diminishing returns) utility function is applied to each flow [37]. A good exposition on the motivation for the fairness model can be found in [25]. The transmission constraints are in turn determined by the physical layer properties and the optimization considerations that are specific to the antenna technology.

The focus and contribution of this section is to extend the analytical framework for omni-directional antennas [37] to handle different antenna technologies, and most importantly show how the problem formulation with respect to the different technologies can be accommodated within the same analytical framework. We begin by briefly explaining how the channel access problem in ad-hoc networks can be formulated as a four-phase process generically, and then specifically indicate how the four phases can be adapted with respect

to the specific technologies.

Phase 1: Node Graph Generation

Given the network topology, the *node graph* G = (V,E) is generated as in Figure 22 (a), where V represents the set of nodes in the network, and E represents the set of edges between all those pairs of nodes that are within transmission/reception range of each other. This does not include edges between nodes that are within carrier-sense range of each other. The adjacent nodes in Figure 22(a) are assumed to be separated by about 250m.

Phase 2: Flow Contention Graph Generation

From the node graph, the *flow contention graph* G' = (V',E',W) is generated as in Figure 22 (b), where V' represents the links in G that have a packet for transmission (*active links*). E' represents the edges between any two vertices in G' (active links in G) that are within contention range of each other. We define two vertices to be within contention region of each other, if either of their transmissions can cause interference at the other. The $V' \times V'$ matrix W represents the weights on the links [0,1] which are representative of the strength of interference between the corresponding active links. We can approximate the weights of the edges in the flow contention graph as,

$$w_{ij} = \frac{S_{i,j}}{\frac{S_j}{SNR_{thresh}} - N_j} \tag{39}$$

where $S_{i,j}$, S_j, SNR_{thresh} and N_j represent the signal strength of the interfering link i on j, signal strength of the communicating link j, required SNR threshold, and the noise level at the receiver of the desired link j. Note that this model can be made more rigorous to include the spatial correlation between the interfering signals in the case of adaptive arrays and MIMO links.

Phase 3: Resource Constraint Graph Generation

The *resource constraint graph* G" = (V",E") that captures the various contention regions in the network topology based on the flow contention graph, is then generated as in Figure 22 (c). It is essentially a bipartite graph with the two sets of vertices being V' and R, where V" = V' + R and R represents the set of resource vertices, one for each contention region.

Each resource vertex could be considered as the resource server for a contention region indicating the total amount of resources (K_{Rx}, without flexible interference suppression) available in the particular contention region. The links in the set V' have two parameters (K_{Tx}, K_{Rx}) indicating the amount of resources available for transmission and reception respectively. The edges in G" correspond to links going from the set V' to set R indicating the membership of the active links in the various contention regions. These contention regions can be obtained by identifying the various *maximal cliques* in the flow contention graph. Figure 22 c) represents the resource constraint graph for the example considered. It can be seen that there are three maximal cliques {r_1, r_2, r_3} that can be identified from Figure 22 b), namely {abc, acd, cdef}.

Phase 4: Problem Formulation

Given the resource constraint graph, the channel allocation problem can be modeled as a utility maximization problem subject to the transmission constraints.

For each flow, consider a utility function *U(s)*, for a service (allocation in streams) s that is continuous, differentiable, increasing and strictly concave over the range $s \geq 0$. Also consider some time window $[t - \Delta T, t + \Delta T]$. Let $C_{i,j}(t)$ be the instantaneous channel allocation indicator as a function of time. If $C_{i,j}(t) = m$, then the vertex i ($i \epsilon V'$) in the flow contention graph has obtained an allocation of m streams in the contention region j ($j \epsilon R$) at time t. Let $a_i(t)$ be the channel allocation (service) for flow i in time [0,t].

The model should incorporate the optimization considerations that are specific to the antenna technology. This can be achieved by subjecting the considerations as constraints to the channel access problem. Let $T_i(l)$ denote the time for service of l streams for the link i. The problem of media access can now be modeled as,

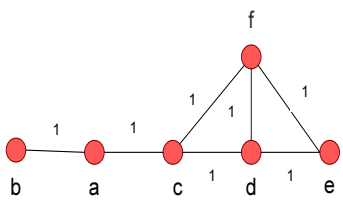
$$Maximize \sum_i U(s_i),\ where\ s_i = \frac{\Delta a_i(t)}{\Delta t}, \tag{40}$$

with the allocation being,

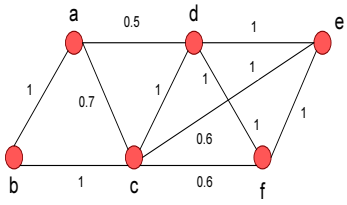
$$\forall t,\ \forall i,\ a_i(t) = a_i(t-1) + C_i(t), \tag{41}$$

$$where\ C_i(t) = min\{C_{i,j}(t)\} \tag{42}$$

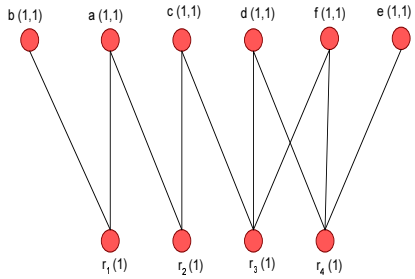
subject to the constraints specific to the antenna technology. The above allocation indicates that for a link i to gain channel access in a slot, it must receive non-zero allocation in each of



a) Flow Contention Graph



b) Supplement Graph



c) Resource Constraint Graph

Figure 23: Switched Beam Graphs

the resource (contention) regions, and the allocation obtained would be the minimal of that obtained over all the contention regions. Since the objective function remains the same, we focus only on the constraint formulation when discussing the last phase with respect to each of the antenna technologies.

6.3.1 Omni-directional Antennas

The *node graph* is obtained as mentioned before. Figure 22(a) is used as a running example to work through the phases.

The *flow contention graph* is generated from the node graph. The vertices (links) in V' that are within the reception range of each other have their link weights set to 1. Links that are out of reception range but inside the carrier-sensing range have a weight between 0 and 1, while those lying outside the carrier-sensing range have their weights set to 0. Since only one element (resource) is available, as long as two links have an edge of non-zero weight between them, then only one of them can be active in any time slot. Hence, all the edges present in the graph have a weight of 1. Figure 22(b) represents the flow contention graph for the node graph in Figure 22(a) with all the edge weights set to 1.

The *resource constraint graph* is obtained such that each of the resource servers has a

resource of only one, with the amount of resources available to the links also being limited to one ($K = K_{Tx} = K_{Rx} = 1$). Furthermore, there can be only one winner in any contention region since a link can be allotted only an integral number of resources. Figure 22(c) represents the corresponding resource constraint graph for the example considered with $K_{Tx} = K_{Rx} = 1$.

The utility maximization problem is thus subject to the following constraints,

$$\forall j,\ \forall t\ \sum_{i} C_{i,j}(t)\ \leq\ 1, \tag{43}$$

$$\begin{aligned} where\ C_i(t) &= 1,\ if\ \{C_{i,j}(t) = 1\},\ \forall\ j\ \epsilon\ R \\ &= 0,\ otherwise \end{aligned}$$

The above constraint implies that there can be only one winner in each contention region and a link can gain access to a slot only if it wins in all the contention regions that it belongs to.

6.3.2 Switched Beam Antennas

The *node graph* is generated as described before in Figure 22(a). However, if range extension is exploited, then the number of edges between nodes in the node graph will accordingly increase.

The *flow contention graph* is significantly different from Figure 22(b) due to the directionality involved. It is presented in Figure 23(a). For a given traffic pattern, the active links in the network are determined. Now, two links are said to have an edge between them, if either of the receivers is within the reception/sensing range of the transmitters of both the links on the same beam, in the case of uni-directional communication. In the case of bi-directional communication, both the ends of a link must be considered as a receiver in deciding if it contended with another link. Further, if the directional gain is to be used for range extension, then the corresponding range extension factor can be obtained from equation 61 and the flow contention graph can be generated taking into account the extended range of communication and interference in the node graph. The weights on the edges between links are still {0,1} due to the lack of fine grained interference suppression.

The incorporation of directionality in the generation of the flow contention graph, exploits the additional degree of spatial reuse resulting from directional transmissions by reducing the number of edges between links. However, the flow contention graph does not account for the effect of side lobes, to model which we use another flow contention graph which we refer to as the *supplement graph* hereafter. The supplement graph is essentially the omni-directional flow contention graph for the same network graph (see Figure 23(b)).

The *resource constraint graph* is generated in the same manner as in the omni-directional case. However, to model the effect of side lobes, we consider every node to have a main beam with a peak gain of unity and side lobes of a smaller but uniform gain along the other $K-1$ beam directions (a conical main lobe and a spherical side lobe) with a *front-side lobe ratio,* f ($f > 1$). . Hence, whenever a link is scheduled its neighbors that lie along the beam direction in the supplement graph are updated with an increased noise power in a diametrically opposite beam. Whenever a link is to be scheduled in the flow contention graph its total noise power along all the beam directions is accumulated from the supplement graph to check if the SNR threshold is still maintained (see equation 45). If so, then the link is scheduled. The constraints to the problem can now be formulated as,

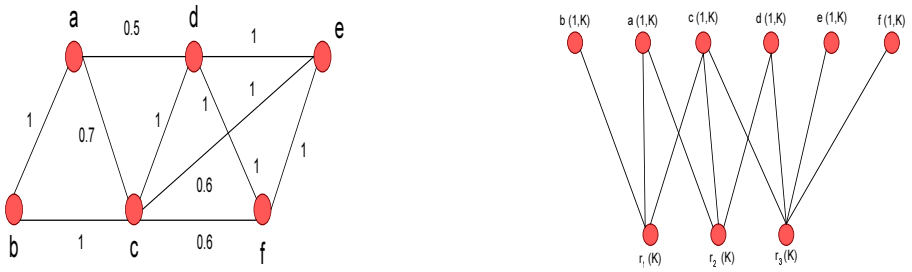
$$\forall j,\ \forall t\ \sum_{i} C_{i,j}(t)\ \leq\ 1, \tag{44}$$

$$\begin{aligned} where\ C_i(t) &= 1,\ if\ \{C_{i,j}(t)=1\},\ \forall\ j\ \epsilon\ R \\ &\quad \&\&\ \sum_{m=0}^{K-1} \frac{N_i(m)\cdot w_{ij}}{f}\ \leq\ \frac{1}{SNR_{thresh}} \\ &= 0,\ otherwise \end{aligned} \tag{45}$$

where K is the total number of elements (beams), w_{ij} is the interference weight between the links i and j, and $N_i(m)$ is a variable corresponding to link i such that,

$$N_i(m) \quad = \quad l,\ if\ \exists\ l\ transmissions\ along\ beam\ direction\ m$$

If instead of range extension, a direct increase in capacity is required by switching to higher modulation rates, then the resource in each contention region can be viewed to be scaled by a factor (C_f, given in equation 32) with each link's allocation being limited to $\{0,C_f\}$.



a) Flow Contention Graph b) Resource Constraint Graph

Figure 24: Adaptive Array Graphs

6.3.3 Fully Adaptive Array Antennas

The *node graph* is the same as in the other antenna technologies (Figure 22(a)). But if range extension is employed, the number of edges would increase.

The *flow contention graph* is generated in the same way as for the omni-directional antennas. However, since the adaptive array antennas are capable of fine grained interference suppression, the edges between the links can have weights between 0 and 1 ($0 \leq w_{ij} \leq 1$, see Figure 24(a)). Also, the range extension resulting from the array gain (see equations 33 and 34) can be incorporated in the generation of the flow contention graph.

The procedure for the *resource constraint graph* generation remains the same. However, the resource servers have a total of K (= number of elements) resources each, with only one resource being used by any link for its own transmission. The remaining resources are used by the link for flexible interference suppression ($K_{Tx} = 1, K_{Rx} = K$). The resource constraint graph is presented in Figure 24(b).

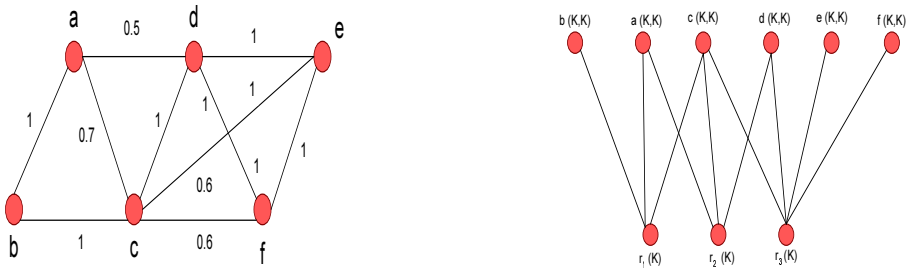
The constraints for adaptive array antennas are thus,

$$\forall j,\ \forall t\ \sum_{i} C_{i,j}(t)\ \leq\ 1, \tag{46}$$

$$\begin{aligned} where\ C_i(t) &= 1,\ if\{C_{i,j}(t)\ =\ 1\},\ \forall j\ \epsilon\ R \\ &= 0,\ otherwise \end{aligned}$$

$$G(s_1) \geq G(s_2) \geq \cdots \geq G(s_K) \tag{47}$$

$$\forall t,\ \forall i\ \sum_{l} w_{li} C_l(t)\ \leq\ K,\ if\ C_i(t) \geq 0, \tag{48}$$



a) Flow Contention Graph b) Resource Constraint Graph

Figure 25: MIMO Links Graphs

$$where\ l \epsilon L,\ the\ set\ of\ links\ such\ that \tag{49}$$

$$\exists\ some\ j \epsilon R,\ for\ which\ \{(l,j) \epsilon E\text{"}\ \&\&\ (i,j) \epsilon E\text{"}\} \tag{50}$$

If the streams are arranged in the descending order of their gains $G(s_i)$, then the constraint in equation 47 ensures that the strongest stream will always be chosen. Constraint in equation 48 incorporates the flexible interference suppression possible through adaptive array antennas. This in turn, creates more resources and hence more transmissions in the same contention region if the interference caused by some links is weak. Thus, while the total number of resources used in any contention region can potentially be greater than K, the total number of resources used up by any link in any contention region cannot exceed K. If the array gain is to be used towards increasing the capacity directly rather than increasing the range, then the resource in each contention region (K) can be viewed to be scaled by a factor (C_f, given in equation 36) with each link's allocation being limited to $\{0,C_f\}$.

6.3.4 MIMO Links

The *node graph generation* remains the same as in the other cases (Figure 22(a)).

The *flow contention graph* is generated in the same manner as in the case of adaptive array antennas (Figure 25(a)).

The *resource constraint graph* is similar to the one in adaptive array antennas, except that every link can potentially use all its K resources for transmission as shown in Figure 25(b) ($K_{Tx} = K_{Rx} = K$). Coloring is once again required to enable passive receiver

overloading. Only the white links perform stream control, while the red links operate on all available K streams.

The problem of media access now has the following constraints,

$$\forall j,\ \forall t\ \sum_{i} C_{i,j}(t)\ \leq\ K, \tag{51}$$

$$where\ C_i(t) = min\{C_{i,j}(t)\} \tag{52}$$

$$0\ \leq\ C_{i,j}(t)\ \leq\ K\ \ if\ i\ \epsilon\ WHITE,\ and \tag{53}$$

$$C_{i,j}(t) = \{0, K\}\ \ if\ i\ \epsilon\ RED \tag{54}$$

$$G(s_1) \geq G(s_2) \geq \cdots \geq G(s_K) \tag{55}$$

$$\forall t,\ \forall i\ \sum_{l} w_{li} C_l(t)\ \leq\ K,\ if\ C_i(t) \geq 0, \tag{56}$$

$$where\ l\ \epsilon\ L,\ the\ set\ of\ links\ such\ that \tag{57}$$

$$\exists\ some\ j\ \epsilon\ R,\ for\ which\ \{(l,j)\ \epsilon\ E\text{"}\ \&\&\ (i,j)\ \epsilon\ E\text{"}\} \tag{58}$$

The primary concern of the model is to perform the two-level colored scheduling to eliminate the under-utilization that results due to the bottleneck links. This is achieved by allowing the red links to receive all K streams as allocation when they get scheduled. On the other hand the white links that get scheduled can receive service anywhere between 0 and K streams depending on the fair share in their maximal clique, due to stream control. If the streams are arranged in the descending order of their gains $G(s_i)$, then the constraint in equation 55 ensures that the best streams are chosen by the white links. Note that the amount of resources used up at the resource server is determined by the edge weights in the flow contention graph and hence the gain of flexible interference suppression is automatically leveraged by way of more resources being available at the resource servers due to weak interference.

6.4 Unified MAC Framework

The objective of the unified algorithm is to maximize the aggregate utilization of the network subject to the transmission constraints and conforming to the fairness model. The proposed solution consists of three components: (i) The first component generates the flow contention

graph (Step 1) and determines all the contention regions (maximal cliques) in the flow contention graph and arrives at the minimum clique cover. Maximal cliques define distinct contention regions in the network where there can be only as many winners in any contention region as to not exceed the number of resources in the region. (ii) The second component *colors* the links. The fairness model that we employ is proportional fairness. A property of proportional fairness in shared wireless medium access is that the solution tends to favor links that belong to a single contention region (non-bottleneck links) from those that belong to multiple contention regions (bottleneck links). Hence, it becomes vital for the algorithm to identify the bottleneck links from the non-bottleneck links and schedule them appropriately to ensure proportional fairness. We refer to this process as *coloring* (Step 2). (iii) Once the maximal cliques are obtained and the vertices of the flow contention graph are colored, the final component of the algorithm determines the proportional fair allocation and computes a schedule of the links to achieve the required proportional fair allocation (Steps 3 and 4).

The four steps mentioned above are the same as those outlined for MIMO links in the previous section. However, the number and manner in which the resources are used for transmission and reception (K_{tx} and K_{rx}) vary from one technology to another. In the rest of this section, we outline the coloring procedure, and how the transmission (K_{tx}) and reception (K_{rx}) resources are utilized in each contention region with respect to each of the antenna technologies and also show how the resource constraints are met and the optimization considerations identified in Section 6.2 are leveraged.

6.4.1 Instantiations

6.4.1.1 Omni-directional Antennas

Every link can transmit and receive on only one resource ($K_{tx} = 1$) and every contention region also has only one resource ($K_{rx} = 1$). The weights of all the edges in the flow contention graph are one, and hence transmission by any link will reduce the resources of all the other links in the same contention region to zero. For this reason, it is not essential to distinguish between the red and white links. Hence, all the links can be colored red

without going through the elaborate process of finding all the maximal cliques.

6.4.1.2 Switched-beam Antennas

The flow contention graph is generated for the given traffic pattern taking directionality into account. For a K element array, K beams of $\frac{360}{K}$ degrees each are considered. However, the presence of scattering limits the gains obtained from increasing K, and must be accounted for in the generation of the flow contention graph. Every link and contention region in a switched-beam environment also has only one resource ($K_{tx} = K_{rx} = 1$). Thus the slot sequence is again given by the size of the largest maximal clique in the flow contention graph. As before, all the links in switched-beam can also be colored red. Since switched-beam links do not perform flexible interference suppression, the weights of the edges in the flow contention graph are all set to one. Hence, the transmission by one link reduces the resource at each of the neighboring links in the flow contention graph to zero (line 34 with $w_{ij} = 1$ and $K_{tx} = 1$). The side lobes are accounted for in the supplement graph. Each link j in the supplement graph also maintains a resource variable $resource_{j,sup}$ which is initially set to $\frac{1}{SNR}$. Whenever a link i is scheduled, the resource of all the neighboring links ($N(i)$) in the supplement graph is diminished by the appropriate noise power ($\frac{w_{ij}}{f}$, f being the front-side lobe ratio) as in,

$$resource_{j,sup} = resource_{j,sup} - \frac{w_{ij}}{f},\ \forall\ j\ \epsilon\ N(i) \tag{59}$$

Thus, whenever a link i is a candidate to be scheduled in the flow contention graph due to the availability of resource ($resource_i = 1$), then its corresponding resource ($resource_{i,sup}$) in the supplement graph is also checked. If it happens to be greater than zero then the link is scheduled.

If range extension is desired, then this is incorporated in the generation of flow contention graph that is fed to the algorithm. On the other hand, if capacity increase is desired, then the resource of each link and contention region ($K_{tx} = K_{rx} = 1$) can be viewed to be increased by the factoral increase in capacity (C_f, given by equation 32). But the resource in the supplement graph remains unchanged and is totally decoupled from the resource in the flow contention graph.

6.4.1.3 Adaptive Array Antennas

The flow contention graph that is fed as input to the algorithm is the same as in the omni-directional case. However, the number of resources handled by each link changes. Every link is capable of transmitting on only one resource as in omni-directional and switched beam antennas, but can handle upto K resources during reception. Though the transmission resource per link (K_{Tx}) is one, the resources per contention region (K_{Rx}) is K. The possibility of multiple winners in a contention region calls for coloring and two level scheduling to ensure a proportionally fair vector as well as to enable passive receiver overloading. The algorithm to determine the proportional fair allocation presented in the beginning of this section applies to the case when $K_{tx} = K_{rx}$. However, when the transmission and reception resources are different, obtaining the proportional fair allocation becomes intractable. In this case we adopt the following heuristic. We first obtain the proportional fair allocation and schedule assuming $K_{tx} = K_{rx} = 1$. Then, in each slot we ensure that the available K_{rx} resources within each contention region are efficiently utilized. Hence, it is possible for white links to be scheduled along with red links in the same slot and vice versa. However, in trying to utilize the resources within a contention region, the white links are given preference to red links. Within white links, the preference is based on minimal service. When red links can also be scheduled in the same slot due to availability of resources, the preference is given to the red link with a lower potential degree. The flexible interference suppression characteristic of adaptive arrays is taken care of in the update of resources where the appropriate edge weights (w_{ij}) are used in deciding the amount of resources required to be sacrificed for suppressing interference. Further, since the algorithm tries to use up all the resources in a contention region by appropriately scheduling transmissions, the gains from diversity are not going to be significant. However, the array gain can be still be used for range extension or capacity increase similar to the directive gain in switched beam antennas. In the case of capacity increase, the resource at each link and contention region (1,K) can be viewed to increase by the factoral increase in capacity (C_f, as in equation 36).

6.4.1.4 MIMO Links

MIMO links can transmit as well as receive on multiple resources ($K_{tx} = K_{rx} = K$) and solution is the same as centralized SCMA protocol.

Detailed exposition on the unified centralized and distributed algorithms can be found in [61, 64].

6.5 MAC Performance Insights

The goal of this section is to evaluate the performance of the different antenna technologies under different network configurations and derive insights from the study. The evaluation is carried out predominantly using the centralized protocols in order to eliminate the distributed inefficiencies in identifying the trends.

We first present results for the technologies evaluated using the centralized protocols for a network configuration of 100 nodes in 1500m * 1500m area. The results for other network configurations are omitted due to lack of space. We consider the following components: (i) impact of scattering, (ii) impact of number of elements, and (iii) impact of load. Each of these components are viewed from the perspectives of, when the available antenna gains are used towards range extension, and when they are used for capacity (rate) increase.

We use the network simulator *ns2* [22]. In addition to free space, two-ray and shadowing path loss models considered in *ns2*, we incorporate the notion of Rayleigh fading into the channel model through the addition of a new collision model [55]. We consider different angles of scattering from 45 degrees to as high as 180 degrees to study the impact of different degrees of multi-path in the network. The number of elements used is from the set {1, 4, 6, 8, 12}. The load on the network can be anywhere from 5 flows to 40 flows from the set {5, 10, 20, 30, 40}. We primarily use throughput (transmissions/slot) as the metric of evaluation. We use the antenna pattern model used in [44] for the main lobe, and side lobe gains of switched beam antennas. Further, the SNR on each of the link is assumed to be 10db in order to achieve a specific level of probability of error on the link. Each of the simulations is run for 500 time slots. Every data point in the graphs presented subsequently is averaged over 10 random seeds. Mobility is not considered in our simulations. The flows

in the network are all multi-hop flows and are considered to be back-logged throughout the simulation. Each of the flows is generated with a random source-destination pair. The routes for the flows are obtained using a shortest path routing protocol and fed as input to the algorithm to perform scheduling for channel access.

6.5.1 Impact of Scattering

We first consider the performance of switched beam antennas in the case of LOS. In this case, MIMO links should change their strategy to operate as an adaptive array since spatial multiplexing cannot be leveraged in LOS. Adaptive arrays gain over the switched beam antennas due to their flexible interference suppression capabilities and also due to the degradation suffered by switched beam antennas from *scalloping loss*. The degradation due to scalloping loss would be even more significant in mobile environments.

Observation 1: In the presence of LOS, adaptive beam-forming provides the best performance. MIMO links cannot leverage spatial multiplexing; hence change of strategy is required for MIMO.

In the presence of multi-path, adaptive arrays are not significantly impacted as long as each link is able to control the number of transmissions in its neighborhood. However, switched beam antennas suffer significantly with increasing multi-path. In the presence of multi-path, beyond a certain number of elements (determined by the scattering angle) the performance tends to saturate, and the knee of the saturation point keeps shifting to the left with increasing multi-path.

Observation 2: Switched beam antennas suffer from multi-path unlike adaptive arrays and MIMO links.

6.5.2 Impact of Number of Elements

The environment considered is a multi-path environment with a scattering angle of 60 degrees. The load in the network is maintained a constant of 20 flows while the number of elements is varied. Each of the flows is a multi-hop flow with an average of 4-5 hops each. The results are presented in Figures 26(a) and 27(a). The following observations

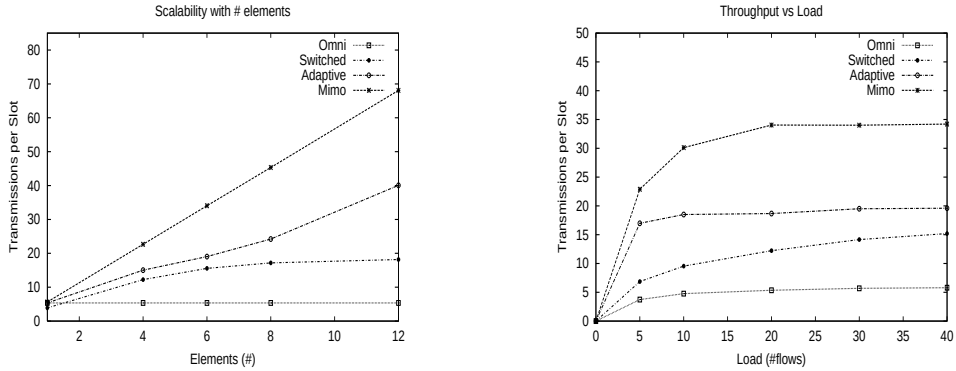
can be made: (i) The MIMO links outperform all the technologies under consideration, owing to the potential linear increase in capacity and the optimization considerations. The improvement over the adaptive array antennas is due to the fact that MIMO links employ spatial multiplexing to exploit all available resources efficiently. A potential reason for the inefficient resource utilization in adaptive array antennas stems from cases where the number of links in a contention region is less than the number of elements. (ii) Switched beam's performance tends to saturate beyond the use of six elements due to the scattering angle of 60 degrees.

When range extension is exploited, with an increase in the number of elements, the number of active neighboring links to any link considered also increases due to the increased range. Hence, it is possible for the per-source throughput to scale (with number of nodes) as long as the number of elements is sufficient to hit full network connectivity and accommodate all transmissions in the network. While this is potentially true for adaptive arrays and MIMO links, the presence of strong multi-path could saturate switched-beam's throughput without helping it to scale. In the case where rate increase is leveraged, the number of active neighboring links does not change. But employing higher rates increases the link capacity logarithmically (with elements) in the case of switched beam and adaptive, and linearly in the case of MIMO links. Again, in the case of switched beam antennas, strong multi-path would limit its performance.

Observation 3: Average number of active neighboring links is an influential parameter in performance. In range extension, increasing elements increases the number of active neighboring links, unlike in rate increase where the gain comes purely from employing higher rates.

6.5.3 Impact of Load

The number of elements is fixed at six and the load in the network is increased upto 40 multi-hop flows. The scattering angles are considered only from 90 degrees onwards, since the impact of scattering on a six element link is felt only from 90 degrees onwards. The results are shown in Figures 26(b) and 27 (b). The MIMO links exhibit the best performance. Since

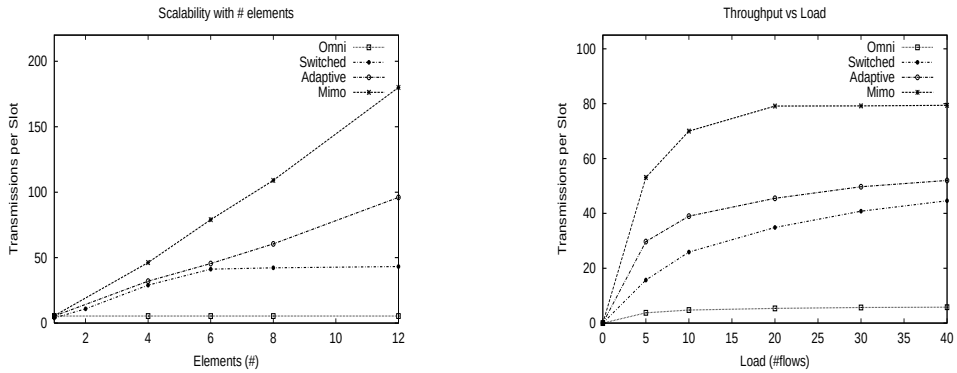


a) Varying Elements: Range Extension b) Varying Load: Range Extension

Figure 26: General Performance Trade-offs (Range Extension)

the number of elements is fixed, the transmission ranges are fixed. However, increasing the load increases the number of active neighboring links to any link considered. In the case of adaptive arrays and MIMO links, when the number of active neighbor links increases beyond K (= number of elements), the throughput starts to saturate (transmissions in the contention region bounded by K). Further, when the load is light and the number of active links in a contention region is less than K, then the throughput in the contention region and hence performance is bounded by the number of active neighboring links for adaptive arrays, while it is still bounded by K in the case of MIMO links. In the case of switched beam antennas, more than the number of active neighboring links, it is the orientation of communication of the neighboring links that matter. Hence, more than K transmissions can possibly be scheduled in the same region thereby showing a slight increasing trend in performance with increasing load. But for a given K, the transmissions in a contention region can be viewed to be bounded by the number of active neighboring links. However, the reduction in performance due to multi-path is much more significant. Also, they are not capable of performing flexible interference suppression unlike adaptive arrays and MIMO links.

Observation 4: Increasing the load increases the number of active neighboring links; but performance is bounded by the number of links that can be accommodated in a contention region, which in turn is determined by the number of elements.



a) Varying Elements: Rate Increase b) Varying Load: Rate Increase

Figure 27: General Performance Trade-offs (Rate Increase)

6.5.4 Range Extension vs. Rate Increase

From the results presented thus far in Figures 26 and 27, it can be clearly seen, that the performance improvements obtained from using the gains towards rate increase are far more than those obtained from using the gains for range extension. This is possibly because, when the gains are used for range extension, with a decrease in the number of hops resulting from increased range, the spatial reuse also decreases. However, when the gains are used for rate increase, the spatial reuse is retained while the capacity is increased.

6.6 Unified Routing Overview

Recall from the design of unified MAC framework, that the capabilities of MIMO links forms a super-set of those possible with other smart antenna technologies. This in turn helped exploit the optimization considerations identified for MAC protocol design with MIMO links, to other smart antenna technologies as well, thereby resulting in a unified MAC framework. Going by a similar principle, the routing optimization consideration identified in *MIR* can now be easily extended into a unified routing framework for use with smart antennas in general. While the unified MAC framework focused on the unified centralized algorithms (derived from the theoretical formulation) in addition to the unified distributed algorithms, the unified routing framework will primarily focus on the unified distributed algorithms, for two reasons: (i) The distributed algorithms are more vital from a practical deployment perspective; and (ii) The efficiency of routing protocols can be truly

estimated only under *network dynamics* like mobility, network partitions, etc., which are tough to model theoretically. Hence, a theoretical treatment of the unified routing problem becomes both intractable as well as non-beneficial for providing any insights into the design of practical routing protocols for deployments, unlike the unified MAC problem.

One of the key aspects of *MIR* routing protocol design that makes it easily extensible to a unified framework, is the abstraction of MIMO capabilities in the form of strategies relevant to routing protocol operations. Specifically, *MIR* exploits the following three strategies in its routing protocol operations, namely:

- Rate: Exploitation of smart antenna gain (spatial multiplexing gain in MIMO) for providing increased data rates on the communication link.
- Range: Exploitation of smart antenna gain (diversity gain in MIMO) for providing increased communication ranges on the communication link.
- Reliability: Exploitation of smart antenna gain (diversity gain in MIMO) for providing increased communication reliability (reduced BER) on the communication link.

The above three strategies are also possible in other smart antenna technologies like switched beam and adaptive arrays, although the specific physical layer mechanism contributing to the strategy will be different for the different smart antenna technologies. In MIMO links, spatial multiplexing contributes to the *rate* strategy, while diversity contributes to the *range* and *reliability* strategies. On the other hand, in adaptive arrays and switched beam antennas, it is the array gain resulting from adaptive beamforming and the directional gain resulting from fixed beamforming respectively that contribute to all the three strategies. The array and directional gains resulting from adaptive and fixed beamforming respectively, contribute to an increase in the mean SNR of the link, which in turn can be flexibly used for increased data rate, communication range or reliability on the communication link. Thus, the three strategies identified for exploitation in routing using MIMO links, apply to adaptive arrays and switched beam antennas as well.

6.7 Unified Design Rules for Routing

Given that the three routing strategies are supported by all the smart antenna technologies purely by exploiting different smart antenna gains, the design rules derived based on the three strategies for routing with MIMO links will now apply *generically* to smart antennas, namely:

- Spatial multiplexing gain in MIMO and increased modulation rates resulting from array and directional gains in adaptive arrays and switched beam respectively, manifest themselves as the *rate* strategy. The SNR gain from diversity, array and directional gains in MIMO, adaptive arrays and switched beam respectively manifest themselves as the *range* strategy for a fixed link error probability and as the *reliability* strategy for a fixed communication range. The contribution of the different gains towards strategies (increased rate, range or reliability) is more for MIMO links than for adaptive arrays and switched beam.

- *rate* strategy outperforms *reliability* and *range* strategies in the case of non-lossy links. However, it is possible for *reliability* strategy to outperform *rate* and *range* strategies in the case of lossy links even for moderate loss rates.

- Optimal strategy with respect to varying node density would be to use *range* strategy at lower network densities and *rate* strategy at higher network densities.

- *rate* strategy, performing the best under most mobility conditions, serves to be the optimal base strategy of operation. However, *range* strategy's ability to provide increased communication ranges should be used in tandem with *rate* strategy to alleviate its vulnerability to route failures.

- Appropriate strategy in the presence of link losses would be to use *rate* strategy as long as packet error rates are negligible; have mechanisms to detect persistent channel errors and switch to *reliability* strategy to increase reliability.

While the strategies and design rules are common to the different smart antenna technologies, the specific gains provided by the individual strategies will vary from one antenna

technology to another and are outlined in the performance evaluation study in Chapter 7. We now proceed to present the components of the Unified Routing framework for Smart antennas referred to as *URS*, which are presented in the context of a reactive source routing protocol and later extended to a reactive table-driven routing protocol.

6.8 MAC Layer Support

Irrespective of the strategy employed, the PHY layer receiver must be aware of the strategy being used by the transmitter. Only then will it be able to decode the packet using the appropriate decoding strategy. This information exchange is achieved with the help of MAC layer. The MAC transmitter always transmits the preamble of the packet using diversity (STBC) or beamforming (without adaptive modulation and using omni data rate) in which it conveys the strategy to be employed for the actual packet transmission. Since a route can be comprised of both rate (multiplexing or increased modulation rates) and range/reliable (diversity or array or directional gain) links, a receiver cannot determine which strategy to receive a packet on apriori. This is because, the receiver does not know which of its neighboring transmitters is going to transmit and also the strategy that is going to be used by the transmitter. To overcome this problem, the transmitters and receivers use range/reliability (diversity/array/directional gain at omni rate) as the invariant strategy for the preamble in order to account for both rate and range/reliable links. Further, the preamble would convey the actual strategy that will be used by the transmitter for the rest of the packet. In order for the receiver to be able to estimate the channel and use it in its receiver processing, a training sequence is added to the front of the preamble.

Another important role of the MAC layer is to determine the minimum range extension factor with which a packet can be received on a range link. This would help the routing protocol in choosing routes with maximum spatial reuse as we shall see later in this section. Recall that the function f that characterizes range extension factor is linear in the number of elements for MIMO and square root in the number of elements for adaptive arrays and switched beam[1]. Thus, when a packet is transmitted through DIV-RANGE using K

[1]The range extension factor is given by $K^{\frac{2}{p}}$, where p is the path loss exponent and is assumed to be 4 in

elements, it can reach upto K hops in MIMO and anywhere between $\sqrt{K}$ to K hops for adaptive arrays and switched beam depending on the path loss component. However, at the end of the packet, the node adds $K-1$ short preambles, each being transmitted-received at unity rate similar to the packet but using lesser number of elements ($[1, K-1]$) and hence lower diversity/array/directional gain. These preambles correspond to the different $K-1$ range extensions and carry the respective extension information in them. Only a node within the i^{th} ($\sqrt{i}^{th}$) hop will be able to decode the preamble that was transmitted-received using i elements with a diversity (array/directional) gain i^2 in MIMO (adaptive arrays and switched beam). By this method, each node keeps track of the minimum range extension [1,K] ([1,$\sqrt{K}$]) obtained with MIMO (adaptive arrays and switched beam) with which the packet was received from the upstream node and stamps it on the packet to be used by the routing layer.

6.9 Unified Routing Protocol Components

Route discovery and maintenance form the two main components of a reactive routing protocol, under which we consider the following five components, namely, route metric, route request, route response, route failure detection during mobility and channel degradation, and route maintenance. The route metric presented in MIR is generic enough to apply to all smart antenna technologies. Hence, we proceed to discuss the rest of the unified components.

URS uses a route discovery procedure whose two main components are (i) propagation of route request that keeps the route discovery time small, and (ii) route reply propagation and route selection that help obtain quality routes.

6.9.1 Route Request

Every node on receiving a route request packet, goes ahead and propagates the request on the maximum range possible using *range* strategy. Whenever a route request is transmitted using *range*, it is done on all elements (K), and can thus reach upto K ($\sqrt{K}$ to K) hops in

the rest of the discussions.

MIMO (adaptive arrays and switched beam). With support from the MAC layer through the use of short preambles, each node keeps track of the minimum range extension ([1,K] in MIMO and [1,$\sqrt{K}$] in adaptive arrays and switched beam) with which the request was received from the upstream node. Whenever a node forwards a request, in addition to adding itself to the source route, it also adds the extension factor of the link with its upstream node. The source route carried on the request packets will thus "initially" consist of predominantly range links. However, when the destination receives the request packet with a range factor of say i ($i\epsilon[1,K]$ for MIMO and $i\epsilon[1,\sqrt{K}]$ for adaptive arrays and switched beam), it waits for approximately ϕiT seconds (where T is packet transmission duration on an omni link, ϕ is a constant that accounts for the additional time taken to access the channel due to contention) to help the nodes in the last set of i hops including itself to bridge/patch their local upstream range links with rate links. The time spent by the destination in waiting coupled with route response propagation time is sufficient to helps the other upstream nodes to have their local upstream range links bridged (replaced) with rate links if permitted by the underlying topology. The bridging is achieved as follows. Each intermediate node after forwarding a route request keeps track of the reverse path to the source along with its metric. Later on, if it receives another request with a much better route back to the source, then it replaces the local route it has stored for the source with the better route.

6.9.2 Route Response

When a route reply finds its way back to the source using the information on the source route along with the extension factors to be used on the different links, it is checked by every intermediate node along the path. If the locally cached route to the source has a better route metric than that currently being used, then it replaces the portion of the route from itself to the source with the locally stored route. Thus, when the route reply finds its way back to the source, the nodes update the route in the packet with their locally cached route that is made up of minimal range links as determined by the metric. Hence, when the source finally gets back the route reply, the route contained in it consists of all rate

links if such a quality route exists or with the minimal impact from range links otherwise. When a flow is considered in isolation, the worst case delay incurred by URS in getting the request delivered to the destination is $T \cdot (l + K - 1)$ $(T \cdot (l + \sqrt{K} - 1))$ for MIMO (adaptive arrays and switched beam) neglecting the short preambles, with $T \cdot (l - 1)$ constituted by the propagation of request until the last set of i hops and TK $(T\sqrt{K})$ constituted by the worst case waiting time at the destination to allow for local bridging in MIMO (adaptive arrays and switched beam) with $l = \frac{h}{K}$ $(l = \frac{h}{\sqrt{K}})$. This is reasonable compared with the best case delay of $T\frac{h}{K}$ $(T\frac{h}{1+\frac{\log K^2}{\rho}}$, ρ is the mean omni SNR) seconds in MIMO (adaptive arrays and switched beam). The route reply propagation would however incur the same latency in both cases.

URS is thus able to obtain quality routes with the minimal impact from range links at the cost of a small additional delay. The main benefit of this mechanism is that, it does not require the routing protocol to be aware of the network density in order to be able to change strategy in determining routes but can obtain the quality routes provided by the underlying topology for any given density in a transparent fashion. Also, by helping obtain routes with high rate (predominantly rate links), it provides more potential to alleviate link breakages due to mobility and channel effects by switching to diversity techniques, which are explained subsequently.

While the sparse nature of the network is a steady-state feature of the network, mobility and channel errors are network dynamics that need to be handled by the routing protocol through its route maintenance component to avoid throughput degradation.

6.9.3 Route Failure Detection during Mobility and Channel Degradation

The challenge is to detect that a link is going to break due to mobility or persistent channel degradation. The MAC in URS addresses this challenge by switching from *rate* to *range/reliability* after four trials of the RTS packet (RTS has a maximum retry limit of seven in IEEE 802.11b). However, depending on the nature of the loss (mobility or channel degradation), the gain in SNR (diversity/array/directional) automatically provides

increased range for a far-by receiver, or increased reliability for a close-by receiver to appropriately recover from the loss. Hence, it is not necessary for us to differentiate between mobility and channel error losses which is a very useful feature. When switching to diversity, the number of elements used towards *range/reliability* is initially two, since in most cases the range extension or increased reliability resulting from a diversity/array/directional gain order of four is easily sufficient to sustain the link from breaking even at high speeds and high packet error rates. This keeps the reduction in rate and spatial reuse (if increased communication range is exploited) due to diversity/array/directional gain to a minimum and exploits the remaining elements to increase the rate of transmission through multiplexing or increased modulation rates. It also keeps the overhead due to short preambles very small. If the link is already operating in *range/reliability*, then an increase in number of elements exploited for *range/reliability* by one will still serve the purpose.

If the increased communication range or reliability is able to get the packet across, then there are two possible cases: (i) receiver was actually moving away and/or the channel quality was bad; or (ii) neither the receiver was moving away nor the channel quality was bad. The second case is a false alarm (negative) but can be handled easily. The change of strategy is known to the receiver through the normal preamble that is always transmitted using diversity/array/directional gain. Once the receiver receives the RTS, it sends back the CTS using the same strategy used by the transmitter. However, the transmitter on receiving back the CTS packet can determine the false alarm (assuming symmetric channel conditions) if it can successfully decode the short preamble that was transmitted using a lesser (previously used) diversity/array/directional gain order than that used by it on its RTS. On the other hand, if the transmitter cannot decode short preambles of lower diversity/array/directional gain orders, then a "proactive" route error can be generated confidently. However, when the extended communication range or increased reliability is also not able to get the packet across, we again have two cases: (i) neither the receiver was moving nor the channel quality was bad but there is contention; or (ii) receiver was moving and/or channel quality was bad and there is contention. In both cases, we would appropriately infer congestion and would not generate a proactive route error.

6.9.4 Route Maintenance

Now, once the link failure due to mobility or persistent channel errors has been "proactively" detected, the routing protocol at the node is informed. URS at the detection node recognizes the proactive link failure and hence does not purge packets from the queue that are using the link. It generates a proactive route error to the source. The intermediate nodes delete routes from their caches (but do not purge packets) that contain the detected link as in a normal route error so that they do not respond back to the new proactive route discovery process with the stale route. Once the source receives the proactive route error, it initiates a new route request for the destination to obtain a better route with minimal impact from range links through the route discovery component. However, it does not purge the routes that contain the detected link until a new route is obtained or an official route error is received and uses the existing route for sending the packets currently in its buffer. The MAC at the node that detected the proactive route error, though it has switched from *rate* to *range/reliability* to enable communication on the link, it continues to keep track of the diversity/array/directional gain order that is supported by the link in conjunction with its receiver through the short preambles. Once it finds that the gain order used prior to switching can be supported again by the link within four transmissions in succession, it generates a "route error cancel" notification to the source to prevent the unnecessary route change. This would automatically take care of the case wherein the receiver moves out of the current communication range and then comes back into it within a very short duration, and also the case where the channel degradation is short-term. This is because when the receiver moves back in or the channel quality becomes good, the transmitter would receive the short preamble corresponding to a lower gain order.

When a new route is obtained, the source switches its route to the new route and purges the old routes that contain the detected link. The reaction to an official route error is the same as in a conventional source-driven reactive protocol such as DSR. Further, the granularity of the time taken for a link to break due to mobility or persistent channel errors under increased diversity/array/directional gain is large when compared to the route discovery latency. This is because the route discovery component in URS obtains a route in

a time close to that of the pure *rate* strategy. Hence, the proactive route is obtained before the detected link actually fails. Thus, proactive switching from *rate* to *range/reliability* helps reduce the number of route errors due to mobility and persistent channel errors that occur in time-correlated fading, thereby reducing the degradation in throughput, while also addressing all the challenges.

CHAPTER VII

DESIGN RULES FOR OPERATION OF SMART ANTENNAS IN AD-HOC NETWORKS

7.1 Overview

With the help of the unified MAC and routing frameworks and consequently the communication algorithms developed for the different antenna technologies, we now evaluate these different smart antenna technologies and their respective strategies to identify the optimal antenna technology and strategy of operation for a given network condition and application requirement.

7.2 Strategies

We consider the following four strategies with respect to each of the technologies.

7.2.1 RATE

For a given modulation scheme, the BER on the link is determined by its SNR. In this strategy, a gain in SNR (G) due to directional and array gains in switched beam and adaptive arrays respectively, is used to perform adaptive modulation, increasing the number of bits transmitted per symbol, while maintaining the BER at its original value. The increase in capacity can bounded as C_g, where

$$C_g \approx \log_2(1 + \rho G) \tag{60}$$

In MIMO links, the capacity increase results directly from spatial multiplexing and is given in Equation 6.

There is no change in transmission range, link reliability or power consumption.

7.2.2 RANGE

In *range*, the gain in SNR is used for increasing the range of communication r ($r \propto (\frac{1}{SNR})^{\frac{1}{p}}$, p = path loss exponent), while maintaining the same rate (except in MIMO when $K > 2$), degree of reliability and power consumption on the link. Increased range reduces hop length and hence the multi-hop burden, but at the same time also decreases spatial reuse (number of simultaneous transmissions that can exist without causing interference). In switched beam and adaptive arrays, the directional/array gains (say G) are used to provide a range extension factor of R_f given by,

$$R_f = (G)^{\frac{1}{p}} \tag{61}$$

In MIMO, the diversity gain (diversity order = $d = K^2$) can be exploited for range extension. However, since diversity gain relates to reduction in SNR variance, the slope of the BER-SNR curve changes for the link with extended range, making it difficult to translate the diversity gain (unlike directional/array gain) to a range extension factor. So we have evaluated the range extension factor that can be obtained from diversity gain through MATLAB experiments (bit-level simulations) and found it to be a linear function of elements ($R_f \approx K$), as a reasonable approximation.

7.2.3 REL

In *rel*, the gain in average SNR due to directional/array gains in switched beam and adaptive arrays, and the reduction in SNR variance in MIMO due to diversity gain, is retained to directly increase the link reliability and hence reduce the packet loss probability on the link (Equation 7). However, there is no change in the rate (except in the diversity mode in MIMO when $K > 2$), transmission range and power consumption on the link. For switched beam and adaptive arrays with a diversity order of one, the BER decreases by a factor (p_f) of,

$$p_f = \frac{1}{G} \tag{62}$$

while in MIMO, it decreases by a factor of,

$$p_f = \frac{1}{\rho^{d-1}} \tag{63}$$

7.2.4 POW

In *pow*, the gain in SNR is exploited to perform power control and reduce the transmit power P_t on the link, such that the link reliability remains the same. The transmission range and rate (dependent on the STBC code in case of MIMO) remain the same. In switched beam and adaptive arrays, the directional and arrays gains are exploited to reduce the transmit power by a factor, Pt_f,

$$Pt_f = \frac{1}{G} \tag{64}$$

In MIMO, the diversity gain is exploited for power control. If P_{t1} is the transmit power before applying the diversity gain, then the transmit power P_{t2} with diversity gain and power control (using Equation 7) is given by,

$$P_{t2} = (P_{t1})^{\frac{1}{d}} \tag{65}$$

7.3 Simulation Model, Metrics and Algorithms

We first present the antenna, channel and network models used in the simulation study, followed by the metrics of comparison. The algorithms used by the different technologies are presented in the following section. We use the *ns2* network simulator for our evaluations by incorporating the necessary modules into it.

7.3.1 Antenna Model

While the beam pattern generated by the omni-directional, adaptive arrays and MIMO links are either omni-directional or dynamically tunable, the beam pattern of switched beam is fixed and must hence be modeled. The model is similar to the one used in [44] and incorporates front, side and back lobes. Further, front-side and front-rear ratios are assumed to be equal. The main lobe, front-side and front-back gains for switched beams with varying elements (beams) are assumed as provided in [44]. Scalloping loss is also modeled.

7.3.2 Channel Model

In addition to the free space, two-ray and shadowing path loss models considered in the *ns2* simulator, we incorporate the notion of Rayleigh fading into the channel model. Since the simulator works in the granularity of packets, we account for the packet loss probability arising from Rayleigh fading through a new collision model. The collision model captures the probability of packet errors for various configurations (locations, different antenna technologies and strategies used) of desired transmitters and interferers in the presence of Rayleigh fading. This is in turn derived from the BER statistics obtained from bit-level Matlab simulations of detailed physical layer modelling of the different antenna technologies and their strategies along with the required receiver processing techniques (zero forcing, interference cancellation, etc.) in the presence of Rayleigh fading. Thus, different tables of realistic packet drop probabilities for various combinations of antenna technologies, strategies and environment considered are generated and any given configuration of desired transmitter-receiver and interferers is indexed into the appropriate table to obtain the corresponding packet drop probability for the communication.

We have assumed typical parameter values corresponding to the 802.11 standard namely, frequency of operation to be 2.4 Ghz, transmit power to be 20 dBm, fade margin to be 10 dB, packet size to be 1000 Bytes, maximum node speed to be 20 m/s and the channel data rate to be 2 Mbps. The SNR threshold on the link is maintained at 10 dB with fade margins ranging from 0 dB till 10 dB. We have also assumed fast fading and that its not correlated, which is typical at moderate-high speeds.

7.3.3 Network and Traffic Model

We consider a network of 100 nodes, randomly and uniformly distributed in a rectangular grid, and communicating using time division duplexing. The size of the network is varied to vary the node density, ranging from 400m by 400m (average node degree of 20), to 1000m by 1000m (average node degree of 3). Node degree is used as a measure to indicate the node density. The transmission range used by the nodes is set to 100m. The load in the network is varied by varying the number of flows from as low as 1 to as high as 50, where

every node either acts as a source or a destination. The degree of multipath scattering considered in the network varies from 0 degrees (LOS) to 180 degrees (rich urban setting). The impact of node speed (upto 20 m/s) is incorporated in multipath fading through means of packet losses upto 30%. The number of elements possessed by each node determines the available *DOFs*; and is varied between 1 and 12 elements. The sources and destinations are randomly chosen in the network, resulting in a random traffic pattern. Each simulation is run for about 400 secs.

7.3.4 Metrics

7.3.4.1 Throughput

Throughput is measured as the number of bits successfully delivered from the source to the destination per unit time. We measure throughput/flow in our evaluations. However, in comparisons with *range* in the presence of low node densities and low loads alone, aggregate flow throughput (throughput capacity) is considered to account for the difference in the number of flows existing in the network in the different strategies.

7.3.4.2 Throughput/Energy

Here, we measure the number of bits that can be successfully delivered to the destination per unit of Joule consumed. The different components determining the energy consumed per slot transmission are the communication energy (circuit-driving power P_c and transmit power P_t) and computational energy (mainly signal processing operations).

While the operational complexity of the different technologies might be of concern from the perspective of energy and cost, it turns out that the communication energy accounts for most of the energy consumed per transmission-reception. Hence, the complexity of the different antenna technologies in terms of the signal processing required does not significantly impact the power consumption.

7.3.5 Protocols and Algorithms

Since the focus of the study is to obtain fundamental tradeoffs in the operation of the different antenna technologies, the impact of distributed inefficiencies of the protocols on

```
Routing :
1  Estimate Scattering
2  If (tech == "sw"||tech == "adap")
3    Update dir/array gain
4   If (strat == "range")
5    Find_Txrange(tech, gain)
6   Run shortest path routing
7   Generate Flow Contention Graph, G = (V, E)

Scheduling :
8  ∀ iϵV, res_i = K, alln_i = 0, service_i = 0
9  While (slot ≤ 100,000)
10  ∀ i, jϵV & (i, j)ϵE,
11       Obtain channel matrix H_s with coefficients, h^s_ij
         Obtain H_sr with Rayleigh fading coefficients, h^sr_ij
     LINK:
12  ∀ iϵV, If (Is_pkt(i))
13       Register (i, tech, strat, K_tx)
14       R → R ∪ i
15       If (tech == "sw"||tech == "adap")
16            res_i = K_rx  //Update res_i based on scattering

     SCHEDULER:
17  Find J ⊆ R, such that, ∀ jϵJ, alln_j = 0
18  & Check_self (j, K_tx, H_s) & Check_neighbor (j, K_tx, H_s)
19  While (J ≠ ∅)
20       Find k = arg[min(service(J))]
21       Update_SINR (strat, gain, F,h^s_kk,K_tx)
22       alln_k = alln_k + K_tx · R(strat, gain, K_tx)
23       res_k = res_k − K_tx
24       ∀ jϵN(k)
25            Update_SINR (strat, gain, F,h^s_kj, K_tx)
26            res_j = res_j − w_kj · K_tx  //w_kj = f(h^s_kj)
27       energy_k = energy_k + E(strat, gain, SNR_k, d)
28       Re-estimate J
29  ∀ iϵV & alln_i > 0
30       Update_SINR (strat, gain, F, H_sr, K_tx)
31       If Unif (0,1) > PER_Table (SINR,Config)
32            serv_k = serv_k + K_tx R(strat, gain, K_tx)
33  slot++, alln_i = 0, res_i = K

Check_self (i,K_tx,H)
34  If (res_i ≥ K_tx &
    Temp_SINR (strat, gain, F, h_ii, K_tx) ≥ SINR_th)
35       return 1;
36  else return 0;

Check_neighbor (i,K_tx,H)
37  ∀ jϵN(i) & alln_j > 0
38       If (res_j < w_ij · K_tx ||
         Temp_SINR (strat, gain, F, h_ij, K_tx) < SINR_th)
39       return 0;
40  return 1;
```

Figure 28: Pseudo Code for Centralized Algorithm

performance is eliminated by considering centralized protocols for evaluation. While the consideration of distributed protocols for the different antenna technologies is not within the scope of this chapter, the *relative* performance trends derived will continue to hold in the case of distributed protocols as well.

All the sources are assumed to be back-logged for the entire simulation duration. Since the different antenna technologies are considered in all possible strategies of operation for evaluation, we cannot consider the possibility of an antenna technology adapting between multiple strategies in a network and consequently do not use the distributed unified routing components of UFS for the evaluation. Hence, a shortest path routing algorithm (Djikstra's algorithm) is used to determine the routes for the flows based on the transmission range used by the nodes. The medium access control functionality is achieved by the presence of a centralized scheduler which performs max-min node fairness. To better understand the operation of the centralized scheduler, it helps to review some basic terminology: A

flow contention graph ($G = (V, E)$) represents the interference existing between the different links in the underlying network. Hence, the vertices (V) in this graph represent the communication links in the network topology and an edge (ϵE) between two vertices indicates that the two links interfere with each other when operating at the same time (assuming a single DOF). When determining if a link interferes with another, we assume bi-directional communication over the links, which is the case in most of the modern MAC protocols in ad-hoc networks. The weight of the edges is indicative of the amount of interference caused. A necessary condition for a **contention region** is one, where every link in the region contends with every other link in the same region. It can also be identified by determining the maximal cliques (complete subgraphs of maximal cardinality) in the flow contention graph.

The pseudo code for the centralized algorithm is presented in Figure 28. Once the routing protocol determines the routes (lines 1-7), the sources start pumping in traffic into the network. At the beginning of the slot, the channel coefficients with shadowing (H_s), and with both shadowing and Rayleigh fading (H_{sr}) are generated between every pair of communicating and interfering nodes (lines 10-11). If a node has a packet to transmit, it registers with the centralized scheduler (lines 13-14). The centralized scheduler determines the next-hop node and hence the link requesting for service. It also records the technology, the strategy, and the number of *DOFs* (resources, K_{tx}) with which the link would communicate. It also determines the impact of multipath scattering on the technology and if needed limits the gain and determines the effective number of resources K_{rx} that can be used by the link in its communication (lines 15-16). It then determines the set of links (J, lines 17 and 18) that can potentially transmit in the slot. A link belongs to the set if it has sufficient resources (DOFs) of the total effective available (K_{rx}) to go ahead with the transmission/reception and maintain the required SINR (based on the path loss, shadowing,

fade margin F and gain from the antennas[1]), after suppressing interference in all its contention regions due to links that have already been scheduled (lines 34-36). The scheduler also checks to see if the already scheduled links in the concerned link's contention regions will not have their SINR's degraded below their required threshold due to the scheduling of this link (lines 37-40). If both the checks are positive, the link is added to the set J. The link with the lowest service is then chosen from the set J to be scheduled with K_{tx} DOFs (line 20). The scheduled link as well as its neighboring links have their available resources then updated for the current slot to reflect the newly scheduled transmission (lines 21-26). The energy consumed by the scheduled link is also updated based on the strategy employed (line 27). Once all possible links have been scheduled based on their service, available resources and SINR, the impact of multipath fading on the success of the communication is then incorporated. Based on the Rayleigh channels generated between the source, and the destination of the link as well as other interfering nodes, the new SINR on every scheduled link is calculated (line 30). The receiver then uses this SINR along with the configuration (location) of desired transmitter and other interferers to index into the packet drop probabilities table generated by the physical layer (see Section 7.3.2) to determine the success of this transmission (line 31). If successful, the service obtained (bits/slot) by the link is updated based on the strategy employed (line 32). The scheduling then moves to the next slot.

The *rate* strategies apply equations 6 or 60 to determine the additional bits that can be transmitted using their appropriate gains, when a link is scheduled successfully (function R). The impact of *range* strategies is indirectly felt through the presence of lesser number of links and increased size of contention regions due to larger transmission range. The *rel* strategies use the gain for increased link SINR to counteract fading loss, reducing the probability of packet loss in the presence of fading as governed by equations 62 or 63 (functions $Update\_SNR$ and $Temp\_SNR$). Finally, the *pow* strategies apply their respective

[1]Note that, since we have considered fast fading, it is not possible to determine if a transmission will be successful using channel estimates from the previous time slot. So the decision is made based on the antenna gain, fade margin, path loss and shadowing models initially and the effect of fast fading is applied to the packet reception once the transmission is scheduled.

gains to reduce the energy consumption per slot as governed by equations 64 or 65 (function E).

Omni-directional antennas possess *a single DOF for both transmission and reception* ($K_{tx} = K_{rx} = 1$) and are not associated with any gain. Switched-beam antennas use *a single DOF for both transmission and reception* ($K_{tx} = K_{rx} = 1$), but by virtue of directing transmissions, they avoid causing and incurring interference in directions other than their own beam direction. This is taken into account in the generation of the flow contention graph. They incur scalloping loss which is accounted for in the number of bits allocated to a switched-beam link when it gets scheduled successfully. The impact of side and back lobes is also taken into account in determining if the required SINR can be sustained on the link as well as on the active links in its contention regions. Adaptive arrays use *a single DOF for transmission but all available resources for reception and flexible interference suppression* using nulling ($K_{tx} = 1, K_{rx} = K$). MIMO links are capable of using *all DOFs for both transmission as well as reception*. However, only in spatial multiplexing and hence in *rate* the used resources directly translate to multiple independent data streams ($K_{tx} = K_{rx} = K$). In diversity and hence in its *range*, *rel*, and *pow* strategies, the transmitted streams are dependent and hence do not translate to multiple data streams ($K_{tx} = K_{rx} = 1$). However, in these strategies the gain is exploited in a different manner than in *rate*.

7.4 Comparison of Strategies

Recall that each of the antenna technologies (switched beam, adaptive, MIMO) can be used with one of the four strategies (*rate*, *range*, *rel*, *pow*). While we have evaluated and compared each of the technologies operating in each one of these strategies (12 combinations in all excluding omni-directional antennas), we adopt a two-level discussion on the studies conducted for clarity.

First, we evaluate the different strategies in each antenna technology and identify the strategy that delivers the best performance with respect to a metric of interest for a given set of network parameters. Then using these insights, we evaluate the different technologies,

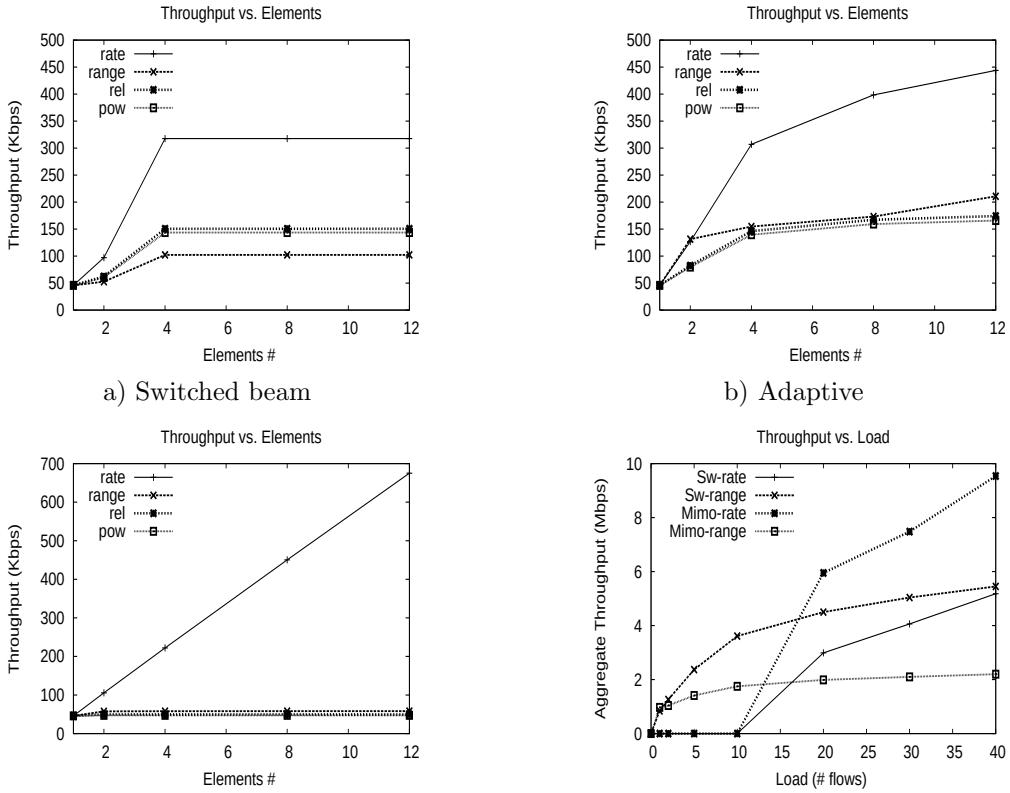
each employing its best strategy that delivers the best performance for the given set of network parameters, and draw inferences on their relative performance.

The key components that impact throughput and throughput/energy are: number of active links/contention region (α), number of independent contention regions (m), and number of resources (K_{rx}) available in each contention region. While α depends on node density, load (number of flows) and hops/flow (h, determined by transmission range); m depends on network size and transmission range. The dependence of network capacity (N_c) on the network parameters can now be captured by, $N_c \propto min\{\alpha, K\} \cdot m$. For MIMO, the resources used in a contention region will always be K and are not limited by α or K_{rx}, unlike in switched beam and adaptive where every link uses a single DOF for transmission (and also for reception in switched beam). The throughput per link T_l, which directly impacts the throughput per flow T_f, is now captured by $T_l \propto \frac{N_c}{load \cdot h}$.

The default values for the parameters used in the results (when not varied) are: a node degree of 12 for density that ensures that the network is connected, a fading loss of 5% which is common in wireless ad-hoc networks, a load of 50 flows ensuring that every node is either a source or destination, array size of 4 elements that ensures easy deployment, and a small scattering angle of 25 degrees to isolate the impact of scattering on other parameters. These default values hold for both the comparisons of strategies as well as technologies in the subsequent section, unless specified otherwise.

7.4.1 Throughput

The throughput (T) results for the different strategies are presented in Figure 29. *rate* performs the best amongst the four strategies under most conditions due to its ability to utilize the available gain from elements to directly increase throughput. In *range*, the decrease in spatial reuse reduces the throughput gain that can be obtained from the decrease in hop length. While *rel* is expected to be a good strategy at high loss rates, this happens only when losses are extremely high ($> 40\%$) and when small number of elements is used. This is because when losses are moderate or low, the reduction in throughput in *rate* is not significant enough compared to its advantages. Further, most of the protection against



a) Switched beam b) Adaptive

c) MIMO d) Exceptions

Figure 29: Throughput - Strategies

losses is leveraged at smaller number of elements itself in *rel* and hence increasing elements further does not contribute to any significant additional gain. Finally, power control does not exploit the available elements to increase throughput and hence performs the worst. So the general trend[2] in throughput performance is $\{rate > rel > pow > range\}$ as can be seen in Figures 29 (a), (b) and (c), where a scattering angle of 90 degrees was considered.

However, the trend is violated under the following conditions: When density is low and hence the network is not connected, then *range* can help more flows exist in the network and hence provide better aggregate throughput. Also, when load and hence the number of flows is already low, the decrease in spatial reuse due to increased range does not impact much. Though the number of flows existing in *rate* is not as many as in *range*, its still possible for the directional/array/multiplexing gain available in *rate* for those existing flows

[2]We use the operator $>$ in $A > B$ to indicate that strategy (technology) A's performance is better than or similar to that of strategy (technology) B.

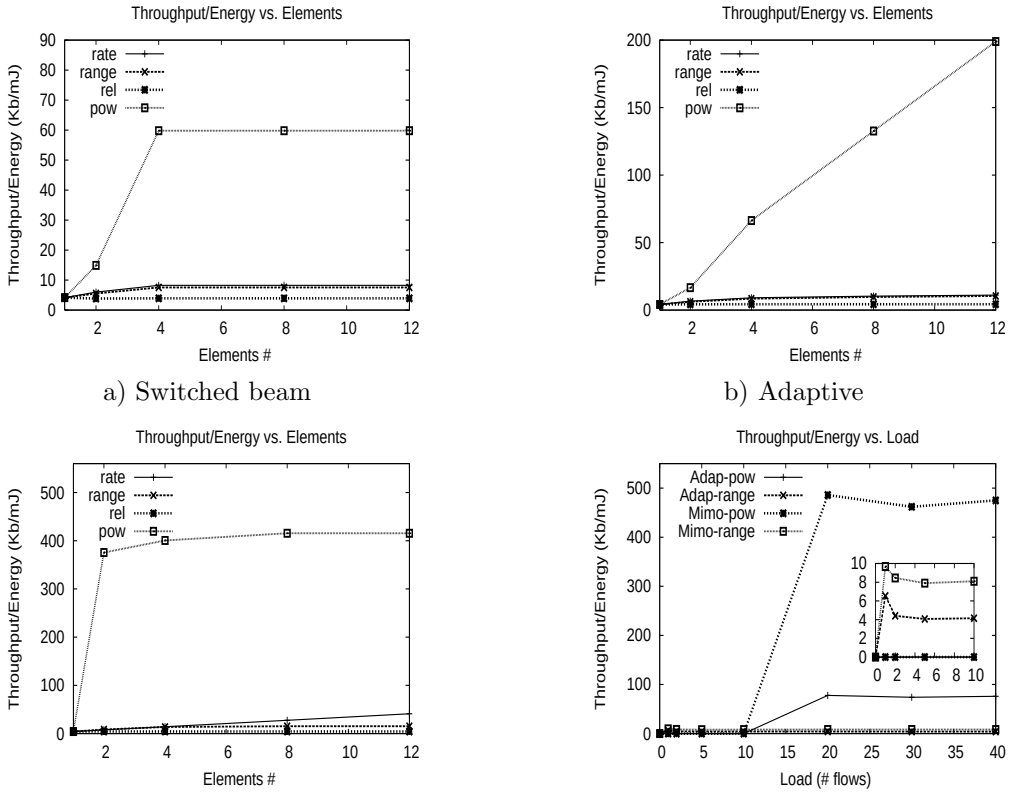
to outperform the aggregate throughput in *range*. Hence, the specific conditions vary with respect to the technology considered (Figure 29(d)). For switched beam, *range* performs the best at low density; and moderate density with low loads. For adaptive arrays, the region is low density and low load; low density, high load but small-medium number of elements; and moderate density, low load with small-moderate number of elements. The reason for small-moderate number of elements in adaptive is because, unlike in switched beam where the directional gain in *rate* is limited severely by scattering and scalloping loss, the array gain in adaptive's *rate* increases significantly with elements to outperform *range*. For MIMO, *range* is best only at low density and low load since the large range extension resulting from diversity gain significantly reduces spatial reuse. This can be seen from the aggregate throughput in Figure 29(d), where a node degree of 3 was considered.

For T, ***rate*** *is the best strategy for all technologies under most network conditions, except at low densities where* ***range*** *is the best strategy.*

7.4.2 Throughput/Energy

In considering throughput/energy (TE) it becomes necessary to consider the relation between circuit driving power P_c and transmit power P_t.This is because, if $P_t \gg P_c$, then the *pow* strategy will help significantly improve TE since it directly exploits the available smart antenna gain to reduce P_t. However, if $P_t \sim< P_c$ (P_t less than or comparable to P_c), then no matter how large the gain is, power cannot be reduced beyond P_c in *pow* since P_c places a lower bound on the power and hence the energy consumed. Hence, the amount of reduction in $(P_t + P_c)$ will provide diminishing returns with larger number of elements. Thus, the effectiveness of *pow*, which is on P_t, is significantly reduced in this case, thereby affecting the trend.

When $P_t \gg P_c$: The TE results for this case are presented Figure 30. *pow* performs the best under most conditions since power control directly helps exploit the gain in reducing P_t unlike the other strategies. The other strategies do not optimize energy and hence their TE depends on how well they can optimize their throughput. Thus, we have the following



a) Switched beam b) Adaptive

c) MIMO d) Exceptions

Figure 30: Throughput/Energy - Strategies

trend for TE performance: $\{pow > rate > range > rel\}$ (Figures 30 (a), (b) and (c); scattering angle = 90 degrees). However, at low loads and low densities, *range* performs the best by virtue of it being able to allow all flows to exist in the network (miniplot in Figure 30 (d); node degree = 3).

When $P_t \sim< P_c$: The ability of *pow* to reduce energy consumption is significantly reduced. In addition, it also does not possess the ability to optimize throughput. Hence, the advantage of *rate* and *range* overshadow *pow*'s energy reduction capability to result in the following trend $\{rate > range > pow > rel\}$. Once again, there exist some exceptions to this trend. *range* performs the best at low loads and low densities for all technologies. Further, in MIMO, *pow* performs the best at low elements, low-moderate loads and high densities unlike in switched beam and adaptive arrays since the reduction in P_t is large even at small elements due to diversity gain and hence the net power P is reduced to almost P_c even at small elements (larger elements in diversity only provide diminishing returns).

Further, *range* does not have the advantage of improving connectivity at high densities. Hence, *pow* outperforms both *range* and *rate* in this region in MIMO.

*For TE, **pow** is the best strategy when $P_t \gg P_c$, and **rate** is the best strategy when $P_t \sim< P_c$, for all technologies under most network conditions. At low densities, **range** is the best strategy in all cases.*

7.4.3 Inferences

We make the following inferences regarding the performance ordering of the different strategies.

For moderate-high network densities,

$$\begin{aligned} T &: rate > rel > pow > range \\ TE(P_t \gg P_c) &: pow > rate > range > rel \\ TE(P_t \sim< P_c) &: rate > range > pow > rel \end{aligned}$$

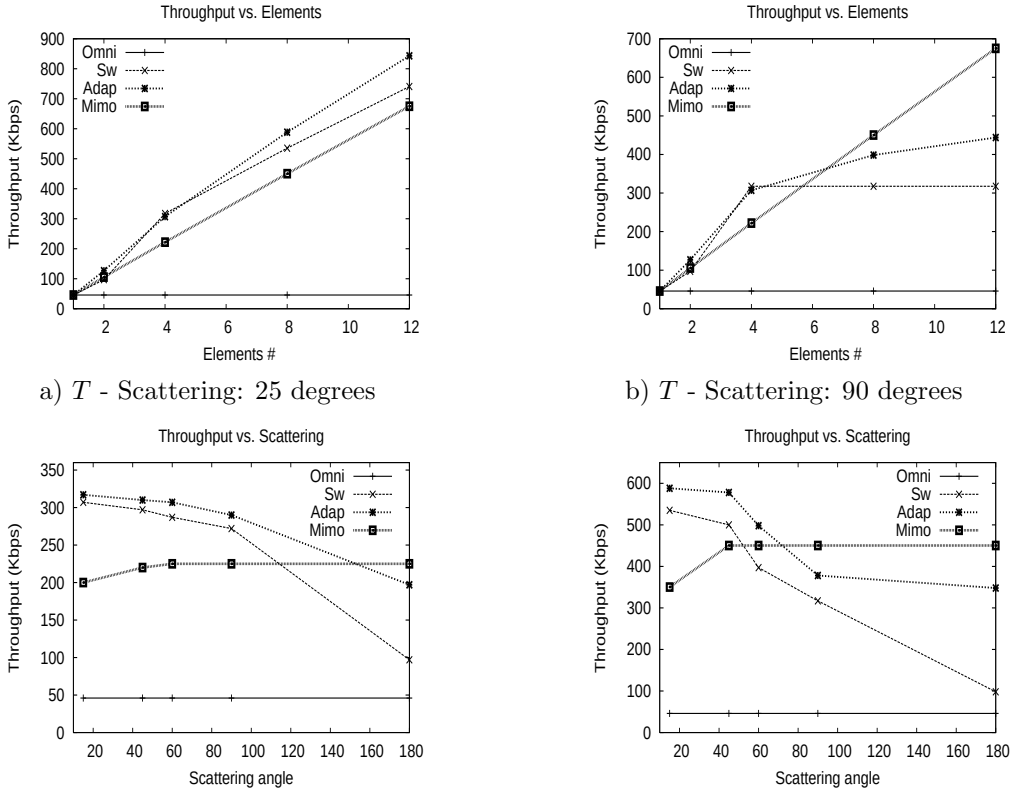
For low network densities,

$$\begin{aligned} T &: range > rate > rel > pow \\ TE(P_t \gg P_c) &: range > pow > rate > rel \\ TE(P_t \sim< P_c) &: range > rate > pow > rel \end{aligned}$$

Thus, we find that the optimal strategy of operation varies not only with respect to the network conditions but also with respect to the performance objective. To the best of our knowledge, this is the first time all possible strategies possible with smart antenna technologies have been evaluated and their performance ordering has been obtained. Further, we have also considered different possible network conditions as well as different performance objectives such as throughput and throughput/energy (taking into account both communication and circuit power). This would help a network designer choose the best strategy of operation based on network conditions as well as the performance objective desired. Further, even if it is not possible for the network designer to operate his network using the best strategy (say, not being able to perform power control and hence use the *pow* strategy

in connected networks with $P_t \gg P_c$), the ordering can help him determine the next best strategy to operate on (*rate* strategy).

In the rest of our comparison of the different antenna technologies, we consider *rate* for T, *pow* for TE when $P_t \gg P_c$, and *rate* for TE when $P_t \sim< P_c$, as the default strategies unless otherwise specified.



a) T - Scattering: 25 degrees

b) T - Scattering: 90 degrees

c) T - Elements: 4

d) T - Elements: 8

Figure 31: Scattering and Elements

7.5 Comparison of Technologies

We evaluate the different antenna technologies with respect to the five network parameters of density, elements, load, scattering and fading loss. However, not all parameters impact each other and hence its not necessary to evaluate the inter-play between all five parameters simultaneously. Recall that the key components that impact T and TE are: number of active links/contention region (α), number of resources available in each contention region (K), and number of independent contention regions (m). α is influenced by load and density;

K is influenced by elements and scattering (in case of switched beam and adaptive); and m is influenced by density. Fading loss directly reduces T and TE and does not inter-play with any of the other parameters. Now, *since each of the components are atmost impacted by two parameters, evaluating the inter-play between every pair of parameters is sufficient.* However, note that α and K together determine the effective number of resources that can be used in any contention region. Hence, the study of the inter-play between parameters impacting both these components is also necessary. While we have evaluated the technologies with inter-play between the parameters for various combinations, we present results and discussions only for the important subset of the combinations that can be used to derive generic inferences and design rules, due to lack of space. Also, we restrict the discussions on inferences and implications to important regions in the inter-play between the parameters considered with respect to both the metrics of T and TE.

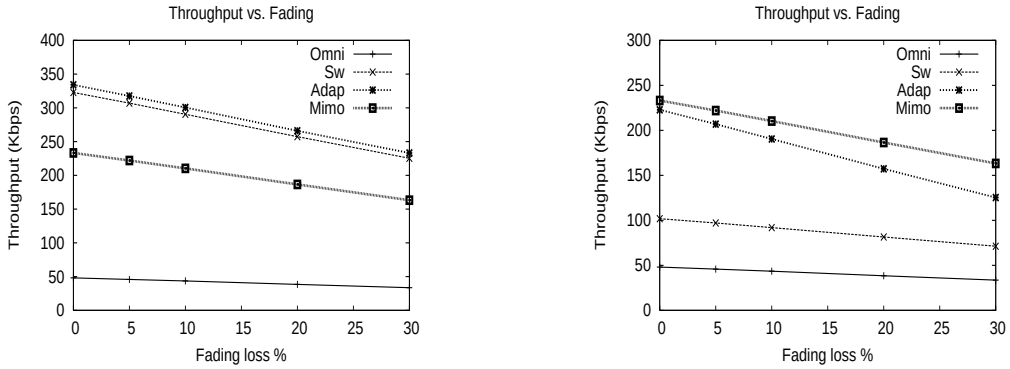
7.5.1 Throughput

7.5.1.1 Scattering and Elements

T decreases with an increase in scattering in switched beam and adaptive arrays due to the loss of energy in undesired directions, unlike in MIMO (Figures 31(c) and (d)). As number of elements increases, the available resources per contention region increases and hence improves T. Scattering limits the amount of gain that can be leveraged from the available elements in switched beam and adaptive (Figures 31(a) and (b)) and is hence more influential.

For increasing scattering angles, the trend in performance is $\{MIMO > adaptive > switched\}$ with the improvement being more at larger number of elements (Figures 31(c)-(d)). But when scattering angles are low, MIMO suffers the most due to the lack of multi-path scattering which is in fact essential for spatial multiplexing, resulting in $\{adaptive > switched > MIMO\}$. The improvement from switched beam to adaptive arrays is much more than the improvement from adaptive arrays to MIMO since scattering has a very significant (negative) impact on switched beam, a slight impact on adaptive arrays and almost no impact on MIMO. Further, the improvement over switched beam is more at large scattering angles and larger number of elements.

At large scattering angles, all loads and all densities, MIMO performs the best. At low-moderate scattering angles, high loads and high densities, adaptive performs the best.



a) T - Scattering: 45 degrees b) T - Scattering 180 degrees

Figure 32: Scattering and Fading

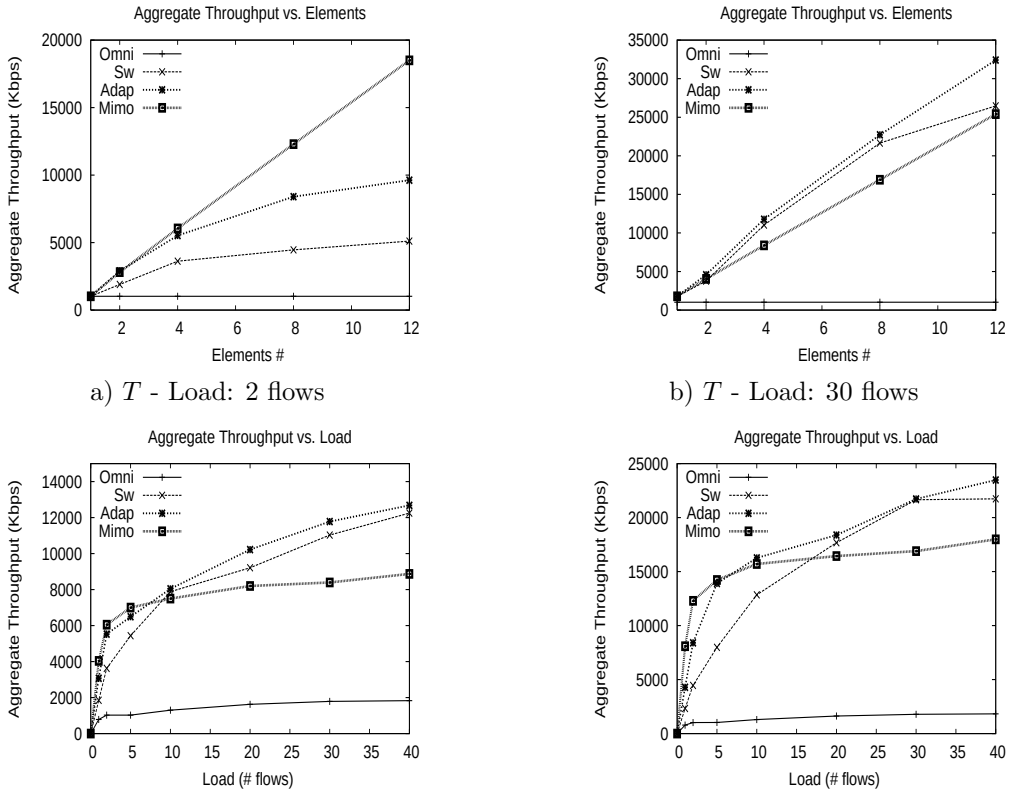
7.5.1.2 Scattering and Fading

Throughput decreases with both multipath scattering and fading losses. While fading degrades performance in all technologies, the impact of scattering is relatively more in the case of switched beam (Figures 32(a) and (b)).

None of the technologies employing *rate* have protection against fading and hence suffer a degradation in throughput. However, the degradation in throughput is not significant enough to shift the strategy to *rel*. In fact, as long as the fading is not highly correlated, semi-reliable MAC layers (eg. IEEE 802.11) will be able to recover from most of these losses, suggesting that the available gain should be wisely leveraged through *rate*. In this comparison, since the number of elements is moderate (four), when scattering angles are low, switched beam and adaptive outperform MIMO (Figure 32(a)). However, at larger scattering angles MIMO outperforms switched beam and adaptive (Figure 32(b)).

Fading impacts all rate strategies alike. The choice of best technology in the presence of fading, is determined by the other network parameters considered.

Fading does not have any influential effect on the other parameters in terms of affecting the trends. Also, since scattering limits the gain (in switched beam and adaptive), it has



a) T - Load: 2 flows

b) T - Load: 30 flows

c) T - Elements: 4

d) T - Elements: 8

Figure 33: Load and Elements

the same effect as reducing the effective available elements (resources) at each node. Hence, we do not present combinations of other parameters with fading and scattering in the rest of the discussions in this section, since their trends can be easily extrapolated from the individual results.

7.5.1.3 Load and Elements

Throughput increases with increasing load and number of elements (Figure 33). While both elements and load seem to be equally dominant in switched beam and adaptive (load increases the number of flows and elements increase the throughput obtained by each flow), elements is the more dominant component in MIMO since even at low loads, a MIMO link can use up all available resources in a contention region unlike in adaptive and switched beam.

As load increases, T increases, as more and more network resources get utilized by the

addition of flows. The trend in performance is $\{MIMO > adaptive > switched\}$ for low loads (Figures 33(a), (c) and (d)), where switched beam and adaptive are not able to use up all available elements ($\alpha < K$). The gain is especially more at larger elements due to high under-utilization in switched beam and adaptive (Figure 33(a)). But when loads are high ($\alpha \geq K$), the resources are not under-utilized in switched beam and adaptive and hence the directional/array gain helps switched beam and adaptive outperform MIMO in the presence of low scattering resulting in $\{adaptive > switched > MIMO\}$ (Figures 33(b), (c) and (d)).

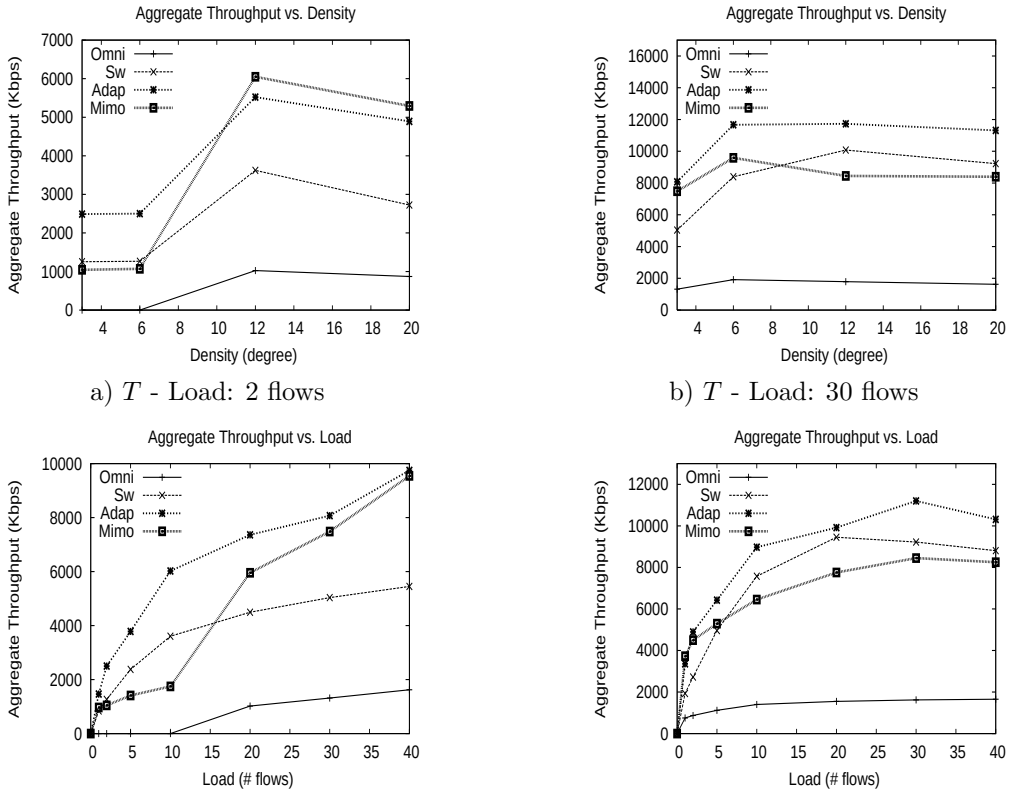
MIMO performs the best at low loads and high densities even in the presence of low scattering, with the gain being more at larger elements, due to the under-utilization of resources in its counterparts.

Since load and density, both impact the number of active links per contention region (α), their influence on the performance trends are similar. Hence, the combined variation of density and elements is not presented here.

7.5.1.4 Load and Density

As density increases (by decreasing network size), T increases initially due to increased number of connected flows but then starts to decrease once the network is connected wherein the impact of the decrease in spatial reuse is felt (Figures 34(a) and (b)). Further, if the elements are not sufficient to accommodate all the active links in the contention region, which in turn increases with density, then T will start to decrease. As load increases, T increases, and then starts to saturate when the available resources are exhausted (Figure 34(d)). Load is a more influential factor since it directly increases the number of flows that can utilize the resources in the network, especially at low densities (Figure 34(c)) and hence directly impacts the number of active links/contention region in the network.

MIMO performs the best for low loads and high densities, when the number of active neighboring links is less than the available elements, thereby resulting in under-utilization of resources in adaptive and switched beam (Figures 34(a) and (d)). For the rest of the cases, adaptive performs the best. As identified in Section 7.4, *range* replaces *rate*



a) T - Load: 2 flows

b) T - Load: 30 flows

c) T - Node degree: 3

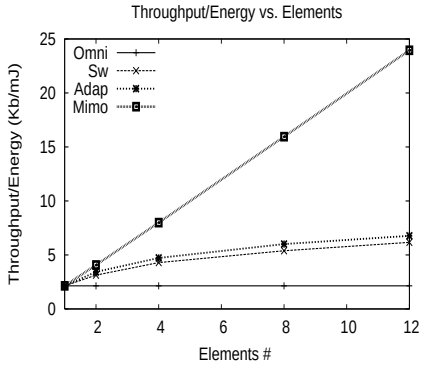
d) TE - Node degree 20

Figure 34: Load and Density

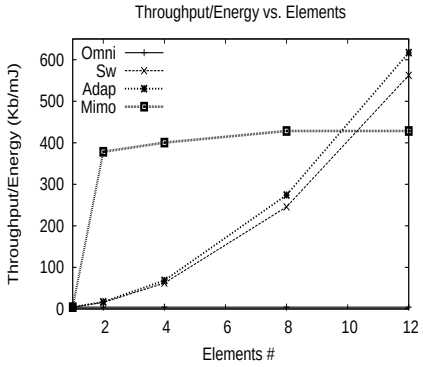
strategy for specific network conditions, namely, at low densities, and low load at moderate densities for switched beam and adaptive; and at low densities with low load for MIMO. We now have the following trends when using *range* under specific conditions: $\{adaptive > switched > MIMO\}$ at low densities with low load (Figures 34(a) and (c)); $\{adaptive > MIMO > switched\}$ at low densities with high load (Figures 34(b) and (c)); and $\{adaptive > MIMO > switched\}$ at moderate densities with low load (Figure 34(a)). Thus, MIMO performs the best only at high densities with low loads. The reason for poor performance of MIMO at low densities with low load is that: when we move to low densities (with low load for MIMO) we need to shift to *range* strategy which provides better T. However, *while the same directional/array gain is used in switched beam/adaptive for range extension, the diversity scheme needs to be used in MIMO, whose large range extension significantly reduces spatial reuse.* Further, diversity has code rates < 1 for $K > 2$ and uses up all degrees of freedom, whereby only one active link can transmit in any contention

region. This degrades MIMO's performance worse than adaptive and switched beam. Further, adaptive (*range*) outperforms MIMO's best strategy (*rate*) at low densities with high load since the range extension from array gain helps find routes for more flows and also reduces the hops/flow (which helps significantly at large loads) without significantly reducing spatial reuse. Note that, we have considered low scattering here. If however, scattering is large then MIMO will perform the best.

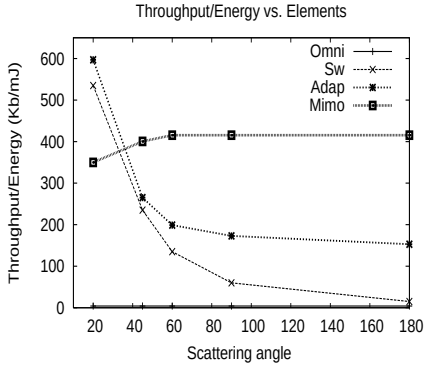
At low scatterings, low densities and all loads, adaptive performs the best, while MIMO performs the worst owing to its ***range*** *strategy resulting in a large reduction in spatial reuse.*



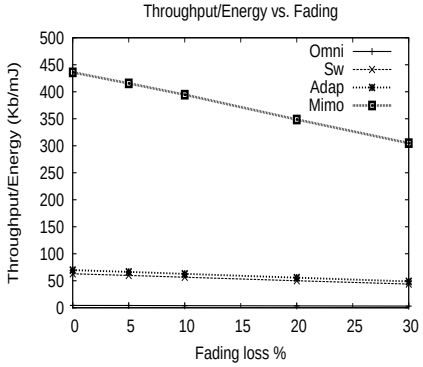
a) TE - Load: 10 flows, $P_t \sim< P_c$



b) TE - Load: 10 flows, $P_t \gg P_c$



c) TE - Elements: 12, $P_t \gg P_c$

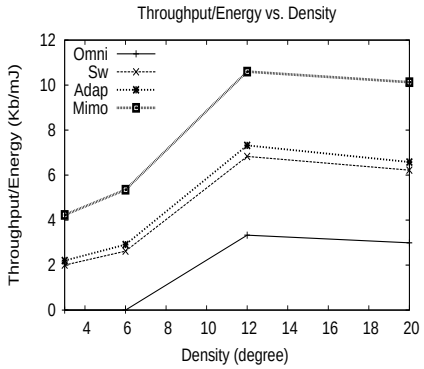


d) TE - Scattering: 90 degrees, $P_t \gg P_c$

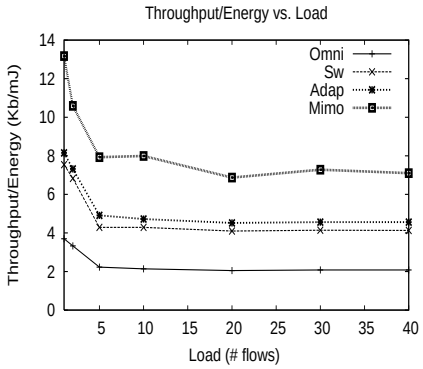
Figure 35: Throughput/Energy (1)

7.5.2 Throughput/Energy

The throughput/energy (TE) results are presented in Figures 35 and 36. TE shows an increasing trend with elements due to increased rate from directional/array/multiplexing gain



a) TE - Load: 2 flows, $P_t \sim< P_c$



b) TE - Node degree 12, $P_t \sim< P_c$

Figure 36: Throughput/Energy (2)

(when $P_t \sim< P_c$; Figure 35(a)) or reduced power consumption from directional/array/diversity gain (when $P_t >> P_c$; Figure 35(b)). TE decreases with an increase in scattering angles in switched beam and adaptive arrays due to reduced antenna gains unlike in MIMO (Figure 35(c)). Increased fading losses directly degrade T and consequently TE (Figure 35(d)). TE decreases with increasing load and tends to saturate at large loads (Figure 36(b)). In fact, both T and energy decrease with increasing load, but the decrease in T is relatively more. This is because the energy associated with a flow is proportional to the product of its throughput and hop length; and as the load increases, there is an increase in the average hop length initially, which in turn tends to saturate with larger number of flows (higher load). Finally, TE shows an increasing trend with density (reduced network size) since hops/flow decreases at high densities, thereby reducing the energy consumption (Figure 36(a)). At very high densities, the impact of decrease in spatial reuse is more than the decrease in hops/flow, thereby degrading TE slightly.

7.5.2.1 When $P_t \gg P_c$:

pow serves as the strategy in all technologies. MIMO performs the best in most conditions due to its diversity gain contributing to large reduction in power, with the trend being $\{MIMO > adaptive > switched\}$ (Figure 35(b)). However, at small scattering angles and large number of elements, MIMO is outperformed resulting in
$\{adaptive > switched > MIMO\}$ (Figures 35(b) and (c)). This is because, *the diversity*

gain and hence power reduction diminishes with increasing elements unlike the array gain in switched and adaptive arrays as can be seen in Equations 64 and 65. At low densities and low loads, *range* is always the strategy employed by all the technologies. In these conditions, while switched beam and adaptive outperform MIMO in T due to the large reduction in spatial reuse (due to increased range from diversity) in MIMO (Figure 34(a)), MIMO performs the best with respect to TE due to the large reduction in hop length resulting from diversity gain (low density region in Figure 36(a)).

For TE, when $P_t \gg P_c$, MIMO performs the best with a large gain for most network conditions (including low densities, unlike in T) owing to its diversity gain. At low scatterings and large elements, adaptive performs the best.

7.5.2.2 When $P_t \sim< P_c$:

rate serves as the strategy in all technologies. MIMO performs the best in most conditions with the trend being $\{MIMO > adaptive > switched\}$ (Figure 35(a)). The main reason is that while switched beam (adaptive) can exploit directional (array) gain, and also potentially enable multiple parallel transmissions in their contention regions due to their K available resources (interference suppression gain), every transmission requires one unit of energy. However, in MIMO, K equivalent transmissions can take place on a link at the cost of one unit of energy due to the spatial multiplexing gain, resulting in power-efficient resource usage. This helps MIMO scale well in TE with respect to elements unlike switched beam and adaptive. Since multiplexing is used, MIMO does not suffer from diminishing returns from diversity gain and hence outperforms switched beam and adaptive even at low scatterings and large number of elements (compare Figures 35(a) and (b)).

For TE, when $P_t \sim< P_c$, MIMO performs the best in all conditions, showing good scalability (with elements) and gain due to the power-efficient resource usage in multiplexing.

7.5.3 Inferences and Implications

We now summarize the inferences with respect to the performance of the different smart antenna technologies.

- As expected, we observe that smart antennas provide significant benefits compared to omni-directional antennas with respect to both T and TE in multi-hop wireless networks. We also observed MIMO to perform the best in the presence of significant scattering, while adaptive and switched beam tend to perform the best when scattering is low or moderate. Further, MIMO's diversity technique provides the best protection against multipath fading losses.

- One of the key inferences is with respect to scalability. MIMO provides the best scalability with respect to increasing elements (resources), due to its ability to use up all available resources efficiently unlike in switched beam and adaptive arrays where a single DOF is used for desired communication, with the remaining resources being used for interference suppression. Hence, the ability to use up all resources and hence scale, depends on the number of active neighbors available to a node in the case of switched beam and adaptive arrays. Thus, as long as number of neighbors ($\theta(\log n)$) is small compared to the available resources ($K < \theta(\log n)$ in adaptive and $K^2 < \theta(\log n)$ in switched beam), there will be under-utilization, thereby limiting scalability. Hence, it is important for network designers to deploy smart antenna arrays with appropriate technology and number of elements based on the network topology and traffic pattern envisioned.

  Another important observation is with respect to deploying large number of elements (and hence large diversity order) for diversity in MIMO. In the case of *rel*, large diversity order can reduce the link BER to arbitrarily small values. However, most applications only require that packet error rates satisfy a certain threshold, especially in the presence of FEC mechanisms and semi-reliable MAC layers such as IEEE 802.11 that employ re-transmissions. Hence, a large diversity gain that comes at the cost of rate, is obviously un-necessary in such situations. Even worse, when diversity is exploited for increased communication range as in *range*, a large range extension than that required (say for connectivity) would significantly reduce the spatial reuse in multi-hop networks, thereby degrading throughput performance. Hence, it becomes necessary

for network designers to devote only as many number of elements for diversity as required for the purpose (increased reliability, range, etc.) and use the remaining for spatial multiplexing, thereby employing a combination of both strategies on the link.

- One would normally expect the relative performance between the technologies to follow $\{MIMO > adaptive > switched\}$. However, we observe that the optimal technology and strategy (Section 7.4) depends largely on (i) network conditions and (ii) performance objective considered, thereby proving the importance of the conducted study. For eg., we find MIMO to suffer significantly in T at low densities where the use of large range extension from diversity significantly reduces spatial reuse. However, we find the same *range* strategy of MIMO to perform the best in TE, due to the large reduction in hop length, which matters the most for TE. We also find the adaptive arrays and switched beam to outperform MIMO in TE ($P_t >> P_c$) at large elements where the diversity gain tends to saturate unlike the array and directional gains.

- Similarly, considering the performance between adaptive arrays and switched beam, we find that adaptive arrays perform significantly better than switched beam at *large scattering angles and large elements*. Hence, unless scattering and elements are large, the gain of adaptive over switched beam may not justify the complexity and communication overhead incurred in adapting its beam-pattern and hence its deployment by network designers.

Note that we have summarized only some of the key inferences above, with the remaining being presented in Sections 7.4 and 7.5. Thus, we have helped identify the best strategy and technology of operation for various combinations of network conditions and for different performance objectives. We believe this would help network designers in the deployment of smart antennas in multi-hop wireless networks (eg. mesh networks), as well as help researchers design better protocols tailored for smart antennas. We also validate some of the inferences observed theoretically in the next section.

Table 1: Throughput and Throughput/Energy Bounds

Strategy	Omni-directional	Switched Beam	Adaptive Array	MIMO
Rate	$\lambda \leq \frac{16AW}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW \cdot K^2(1+\frac{logK^2}{log\rho})}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW \cdot K(1+\frac{logK^2}{log\rho})}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW \cdot K}{\pi\Delta^2 r^2 nh}$
Range	$\lambda \leq \frac{16AW}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW \cdot K^{2-\frac{2}{\gamma}}}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW \cdot K^{1-\frac{2}{\gamma}}}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW}{\pi\Delta^2 r^2 nhK}$
Pow	$\lambda \leq \frac{16AW}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW \cdot K^2}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW \cdot K}{\pi\Delta^2 r^2 nh}$	$\lambda \leq \frac{16AW}{\pi\Delta^2 r^2 nh}$
Rate	$\lambda_E = \frac{W}{h(P_c+P_t)}$	$\lambda_E = \frac{W \cdot (1+\frac{logK^2}{log\rho})}{h(P_c+P_t)}$	$\lambda_E = \frac{W \cdot (1+\frac{logK^2}{log\rho})}{h(P_c+P_t)}$	$\lambda_E = \frac{W \cdot K}{h(P_c+P_t)}$
Range	$\lambda_E = \frac{W}{h(P_c+P_t)}$	$\lambda_E = \frac{W \cdot K^{\frac{2}{\gamma}}}{h(P_c+P_t)}$	$\lambda_E = \frac{W \cdot K^{\frac{2}{\gamma}}}{h(P_c+P_t)}$	$\lambda_E = \frac{W \cdot K}{h(P_c+P_t)}$
Pow	$\lambda_E = \frac{W}{h(P_c+P_t)}$	$\lambda_E = \frac{W}{h(P_c+\frac{P_t}{K^2})}$	$\lambda_E = \frac{W}{h(P_c+\frac{P_t}{K^2})}$	$\lambda_E = \frac{W}{h(P_c+(P_t)^{\frac{1}{K^2}})}$

7.6 Theoretical Analysis

We provide a theoretical basis to validate some of the inferences drawn in Sections 7.4 and 7.5. Note that the goal is not to provide a fundamental theoretical analysis for smart antennas and hence we extend the existing analysis and model used in [20] for our purpose. Let us consider a random network operating on a single channel as a disc with area $A\ m^2$ with n nodes independently and uniformly distributed. Assume the nodes to possess omni-directional antennas initially. Each node is capable of generating $\lambda(n)$ bits/sec of traffic to a random destination. Let us consider a generic single path routing protocol. Further let the bandwidth of the wireless channel be W bits/sec. From the seminal work in [20], we know that when edge effects are incorporated, the following inequality holds in a planar disk,

$$n\lambda(n)h \leq \frac{16AW}{\pi\Delta^2 r^2} \tag{66}$$

where h represents the average hop count and r the transmission/reception range. Δ represents the guard zone parameter which roughly determines the region silenced by a transmission. The term on the left side in inequality 75, $n\lambda(n)h$ determines the total amount of traffic (bits/sec) that needs to be carried by the entire network due to multi-hop burden. The term on the right side determines the total number of simultaneous transmissions in a channel (or spatial reuse), which in the presence of edge effects can be bounded by $\frac{16A}{\pi\Delta^2 r^2}$, where A is the area of the network. Hence, for the generated traffic to be carried by the network we need to ensure that the inequality in equation 75 holds good.

The corresponding energy required by a flow of h hops per slot with a packet size of S bits is $\frac{\lambda(n)}{S}he_t$ and can be bounded as,

$$E \leq \frac{16A}{\pi\Delta^2 r^2 nS} \cdot e_t, \tag{67}$$

where e_t is the energy incurred in a transmission.

Now, the number of bits that can be transmitted by a flow for every Joule of energy (referred to as throughput/energy) can be obtained from λ and the power corresponding to E as,

$$\lambda_E = \frac{W}{h(P_c + P_t)} \tag{68}$$

Thus, the key components that impact the throughput are (i) spatial reuse, (ii) hop length of flows, (iii) channel capacity, and (iv) number of nodes; while the key components impacting throughput/energy are (i) channel capacity, (ii) hop length, and (iii) total power consumed per transmission-reception. Even when smart antennas are considered, they do not change the multi-hop nature of the network and hence the scaling laws and bounds. However, they do impact spatial reuse and channel capacity to provide constant order improvements that are dependent on their respective gains. Further, since the components impacting throughput (identified above) apply to average case throughputs as well, the net improvement (gain) factors provided by smart antennas in the context of throughput upper bounds will hold for the average case throughputs as well. Hence, we consider the net gain factors in the throughput upper bound expressions and throughput/energy expressions to validate the relative performance trends of the different antenna technologies observed in

Sections 7.4 and 7.5. We assume the network is connected and also does not suffer from fading for the initial analysis. Hence, *rel* strategy is not considered initially. Later on, we show how fading can be incorporated into the analysis. Further, we assume that each node in the network is equipped with a MEA with K elements, the average SNR required on a link is ρ, the path loss component is γ and the size of a packet is S bits. The expressions for the throughput (λ) and throughput/energy (λ_E) for the different strategies in the different technologies are provided in Table 1, wherein the technologies have been considered for the specific environments they were designed for.

7.6.1 Switched beam

The directional beams reduce the interference region and hence increase the spatial reuse and consequently the throughput during a transmission-reception by a factor of K^2 [77]. This is valid for all the strategies in switched beam. In addition, the directional gain is used to increase the capacity of transmission by a factor of $1 + \frac{logK^2}{log\rho}$ as governed by equation 60 in *rate*. The throughput/energy also increases by the same factor in *rate* due to the increase in the channel capacity. In *range*, the increased transmission range reduces the hop length by a factor of $K^{\frac{2}{\gamma}}$, but also decreases spatial reuse by a factor of $K^{\frac{4}{\gamma}}$. This coupled with the increase in spatial reuse due to directionality results in a net increase in throughput by a factor of $K^{2-\frac{2}{\gamma}}$. The throughput/energy increases by a factor of $K^{\frac{2}{\gamma}}$, due to the reduction in hop length. In *pow*, the throughput increases by a factor of K^2 due to directionality and throughput/energy increases due to a decrease in transmit power by a factor of K^2. However, the extent of increase in throughput/energy depends on how significant P_t is when compared to P_c.

Note that switched beam antennas are designed for LOS environments and hence the gains and expressions presented are for the low or no scattering case.

7.6.2 Adaptive Arrays

Adaptive arrays deliver high spatial reuse as in switched beam antennas in LOS environments. Even in multipath scattering, adaptive arrays can adapt their beam patterns to

help maximize the SNR of the link. They also suppress interference effectively. In multipath environments, this however, also places a limit on the number of links that can be active at any instant in a contention region (spatial reuse), which is bounded by the number of elements available (K) at any node. The impact on the rest of the components influencing throughput and throughput/energy remain the same as in switched beam and the corresponding expressions are presented in Table 1.

7.6.3 MIMO

MIMO employs spatial multiplexing in *rate* to increase the link capacity by a factor of K in the presence of rich multipath, which increases throughput directly by a factor of K. The interesting fact is that the throughput/energy also increases by the same factor, since a MIMO link is capable of operating on K data streams simultaneously in spatial multiplexing at the cost of one unit of energy while switched beam and adaptive incur K units of energy for the same number of resources. In *range*, MIMO obtains a gain of K in throughput due to hop length reduction, but suffers a factor of K^2 decrease in spatial reuse, resulting in a net degradation factor of K. However, in throughput/energy, the hop length reduction provides a gain of K. Finally, in *pow*, there is no increase in throughput, but the transmit power decreases as $P_t^{\frac{1}{K^2}}$. Note that MIMO links require rich multipath to help the different data streams fade independently at the receiver and hence the gains and expressions are valid for multipath scattering environments.

7.6.4 Comparison of Strategies

From the respective net gain factors (in the expressions) presented in Table 1, it can be observed that with respect to throughput, irrespective of the technology, the relative trend in improvement is $\{rate > pow > range\}$. This is because *range* incurs more degradation due to reduced spatial reuse than the gain due to reduction in flow length. Also, in MIMO, *range* provides a worse gain factor than omni due to the significant reduction in spatial reuse resulting from the large range extension obtained from diversity (although it helps in throughput/energy). With respect to throughput/energy irrespective of the technology, it can be seen that *rate* has more potential to provide better gains *range*. However, their gains

are similar in MIMO since the linear increase in link capacity in *rate* is compensated by the linear reduction in hop length in *range*. The position of *pow* in the relative performance order, essentially depends on how large P_t and hence the reduction in transmit power is, when compared to P_c. If $P_t \gg P_c$, then $\{pow > rate > range\}$. If $P_t \sim< P_c$, then it is possible for $\{pow > range\}$ in switched beam and adaptive depending on the ratio of P_t to P_c, but it might not be possible for $\{pow > rate\}$. On the other hand, in MIMO, depending on the ratio, it is possible for $\{pow > rate \approx range\}$ for smaller values of K owing to the significant diversity gain, reducing net power to the limit of P_c. However, if $P_t << P_c$, then the performance trend is $\{rate \approx range > pow\}$ for MIMO and $\{rate > range > pow\}$ for switched beam and adaptive arrays.

When fading is considered, it becomes necessary to consider the *rel* strategy. Taking into account the speed of the node and the required fade margin, one can estimate approximately the corresponding average bit (p_b) and hence packet loss probability. The throughput and throughput/energy for *rate*, *range* and *pow* will all be scaled by the factor of $(1 - p_b)^S$. The *rel* strategy will have the following throughput and throughput/energy expressions for the different technologies:

Switched beam:

$$\lambda \leq \frac{16AW}{\pi\Delta^2 r^2 nh} \cdot K^2 (1 - \frac{p_b}{K^2})^S \tag{69}$$

$$\lambda_E = \frac{W}{h(P_c + P_t)} \cdot (1 - \frac{p_b}{K^2})^S \tag{70}$$

Adaptive:

$$\lambda \leq \frac{16AW}{\pi\Delta^2 r^2 nh} \cdot K (1 - \frac{p_b}{K^2})^S \tag{71}$$

$$\lambda_E = \frac{W}{h(P_c + P_t)} \cdot (1 - \frac{p_b}{K^2})^S \tag{72}$$

MIMO:

$$\lambda \leq \frac{16AW}{\pi\Delta^2 r^2 nh} \cdot (1 - \frac{p_b}{\rho^{K^2-1}})^S \tag{73}$$

$$\lambda_E = \frac{W}{h(P_c + P_t)} \cdot (1 - \frac{p_b}{\rho^{K^2-1}})^S \tag{74}$$

Hence, it can be inferred that when fading is considered, the relative trend in throughput gains would change as $\{rate > rel > pow > range\}$. However, *rel* would provide the worst

gains when throughput/energy is considered. Further, the analysis assumes a network with sufficient load and density (connected). When the density is low and the network is disconnected, *range* is the only option to maintain connectivity and hence communication in the network. This has been observed in the simulations.

7.6.5 Comparison of Technologies

The bounds presented in the table do not account for scattering in switched beam and adaptive arrays. This is easily accomplished by modifying the resources available and hence the directional/array gain as in equations 29 and 34. Hence, it can be inferred from the table that switched beam and adaptive provide better gains than MIMO in terms of throughput at low scattering angles (LOS), where the directional/array gains dominate. However, their gains tend to degrade at higher scattering angles where MIMO is not impacted. Further, the gains of switched beam will degrade more than adaptive in the presence of scattering. In terms of throughput/energy, MIMO clearly provides significant gains when compared to its counterparts. In cases of low density, when the technologies are forced to use *range*, it can be observed from the table that MIMO will be outperformed in terms of gain in throughput but will still perform the best in terms of gain in throughput/energy . Further, the analysis assumes sufficient load, in the absence of which the importance of MIMO's efficient resource usage in spatial multiplexing will be significantly highlighted. These impacts have been observed in the simulations.

CHAPTER VIII

THROUGHPUT CAPACITY ANALYSIS OF HETEROGENEOUS SMART ANTENNA NETWORKS

The previous chapters dealt with the protocol, algorithmic and theoretical aspects of exploiting smart antennas in ad-hoc networks, the nature of the networks considered was homogeneous, wherein all devices in the network possess the same kind of smart antenna technology. In the rest of this thesis, we shall focus on a more practical and challenging aspect of smart antenna networks namely, heterogeneous smart antenna networks, where different devices in the network could potentially employ different antenna technologies and build efficient communication strategies and protocols for these networks.

8.1 Overview

Smart antennas, due to their unique signal processing capabilities, are considered to hold promise for use in wireless ad-hoc networks. While the properties of smart antennas have been well understood at the physical layer, their relevance to the higher layers of the protocol stack is still being explored. This has motivated several researchers to design MAC [10, 4, 27, 65] and routing protocols [48, 9] that exploit the capabilities of the underlying smart antenna technology. A common underlying assumption of all such work is that all nodes in the ad-hoc network have antenna technology with the same degree of sophistication. Specifically, while smart antennas range from simple directional or switched beam antennas to sophisticated multiple element arrays capable of performing adaptive beam forming and multiple-input-multiple-output (MIMO) transmissions, the above works uniformly assume exactly one type of antenna technology in use by all nodes in the network, and address the relevant protocol in that context. For example, [10, 4, 27] all address the case of all nodes using switched beam technology, while [65] addresses the use of MIMO links by all nodes in the network.

While such homogeneous networks form an essential first step towards understanding the true performance gains that can be delivered by each of the specific antenna technologies, there are several reasons why it is important to consider a heterogeneous wireless antenna network, where different nodes can possibly operate with different antenna technologies: (i) *Economic Feasibility:* Since most of the existing wireless networks operate with omni-directional antennas, revamping entire networks as a whole and equipping all the nodes with a smart antenna technology may not be economically feasible. However, incremental deployments are highly desirable, consequently resulting in heterogeneous antenna networks. (ii) *Mesh Networks:* In mesh networks, one has control over the deployment of atleast some nodes in the network, which can serve as relay points for traffic for other nodes. Such nodes are limited in number for easier network management and can be assumed to be stationary. Hence, it is possible to conceive such "special" nodes to be vested with smart antennas capabilities to improve the overall network performance. Other examples would include digital battlefields and zer-configuration community networks.

In this context, the focus of this chapter is to consider multi-hop wireless networks where the nodes are equipped with varied antenna technologies, and answer the following question: *What performance benefits can one theoretically expect from such networks, and how one would go about realizing the benefits practically using appropriately designed protocols?* In the process of answering the above question, as a vital first step, we first identify the link layer capabilities when nodes with heterogeneous antenna technologies interact. We then make the following contributions:

- We study what the ideal performance bounds are for heterogeneous wireless antenna networks. Perhaps more interestingly, through the bounds, we provide insights into what percentage of the nodes in an ad-hoc network have to be "smart" in order to achieve a certain level of network performance.

- We design medium access control (MAC) and routing algorithms that will help realize the identified performance benefits by maximum exploitation of heterogeneous smart antenna capabilities available in the network.

In this chapter, we present the throughput capacity analysis for heterogeneous smart antenna networks and analyze the impact of increasing degree of antenna heterogeneity on the throughput capacity of the network. In the subsequent chapter, we design heterogeneity-aware communication (MAC and routing) protocols based on the performance insights gained from the throughput capacity study.

8.2 Heterogeneous Link Capabilities

In this section, we identify the different link layer combinations that are possible between nodes with different antenna capabilities and their associated gains and benefits. For a detailed exposure on the different smart antenna capabilities and gains, please refer to Chapter 6.

In order to be able to analyze the performance of multi-hop wireless networks with heterogeneous antenna technologies and design appropriate protocols, one must first understand the different link level capabilities that can be leveraged when nodes with similar and different antenna technologies interact. This in turn forms the subject of this section, where we identify what specific gains can be achieved from a capacity stand-point. However, the specific manner in which they can be achieved will be discussed in Section 10.4. In the ensuing discussions, we refer to nodes with different antenna technologies as O (omni-directional), D (directional), A (adaptive arrays) and M (MIMO) nodes; and to links as x-y links, where $x, y \epsilon \{O, D, A, M\}$, with x serving as the transmitter and y serving as the receiver. The links are compared based on the following three parameters of SNR gain, spatial reuse and diversity order. **SNR gain** refers to the increase in the link SNR resulting from the antenna technology employed on the link. **Spatial reuse** refers to the number of simultaneous transmissions made possible by the interference suppression capability of the antenna technology, in addition to those existing in multi-hop wireless networks due to limited range transmissions. **Diversity order** refers to the degree of protection available on the link against channel errors and can be related to link BER using Equation 7.

We consider two kinds of environments in this work, namely, line of sight (LOS) and multipath scattering (no LOS or NLOS). In LOS environments, we assume that a strong line

of sight signal component exists between the transmitter and receiver. Further, we consider only omni-directional and switched beam nodes in LOS environments. MIMO nodes rely on rich multipath scattering to help the transmitted data streams fade independently and hence do not contribute any gain in LOS. On the other hand, adaptive arrays provide the same performance as that of switched-beam antennas in LOS since no adaptation is required. Hence, switched beam and adaptive arrays are commonly referred to as "directional" nodes in LOS environments in the rest of our discussions. In NLOS environments, we assume that there exists no strong line of sight signal component between any transmitter-receiver pair and consider omni-directional, adaptive arrays and MIMO nodes in these environments. Switched beam antennas suffer significantly in the presence of multipath scattering and their gain (performance) tends to unity (omni-directional) at large scattering angles, which is the assumption in NLOS environments. Hence, switched beam antennas are not considered in NLOS environments. The respective gains for the different link combinations involving smart nodes are listed in Table 2 and are explained subsequently.

8.2.0.1 LOS environments

The different link combinations possible in the LOS environments are O-O, O-D, D-O and D-D links.

The O-O link corresponds to the normal link in omni-directional wireless networks. It provides unity gain, no additional spatial reuse and a diversity order of unity.

The directional node in the O-D and D-O links provides a directional gain of K by virtue of being able to direct (receive) energy in a particular direction. In addition, it also eliminates the interference caused (received) in the remaining $K-1$ beam directions, thereby providing a reuse factor of K [77]. However, since all the antenna elements are used in directing energy between the transmitter and receiver, no diversity gain can be exploited and hence the diversity order is unity. The D-D links provide a directional gain of K^2 owing to both the transmitter and receiver focussing energy towards each other. In addition, the K factor reduction in interference at both the transmitter and receiver results in a K^2 net reuse factor [77]. However, the diversity order is still unity.

Table 2: Heterogeneous Link Layer Capabilities

Combinations	Gain	Spatial Reuse	Diversity Order
O-O	1	1	1
O-D	K	K	1
D-D	K^2	K^2	1
A-A	K^2	K	1
O-A	K	K	1
A-O	K	1	1
M-M	$C_f = K$	1	1
O-M	1	1	K
M-O	1	1	1
A-M	K	1	K
M-A	K	K	1

8.2.0.2 NLOS environments

The different link combinations in the NLOS environments are O-O, A-A, O-A, A-O, M-M, O-M, M-O, A-M and M-A.

The A-A link provides an array gain of K^2 due to its ability to adapt its radiation pattern at both the transmitter and receiver in order to maximize the SINR on the link in the presence of multipath scattering. Further, it also has the capability to form nulls towards $K-1$ interferers in addition to receiving its desired signal, thereby allowing them to operate in parallel and hence providing a reuse factor of K. Since all the resources are used up in this effort, no diversity gain can be expected, resulting in a diversity order of unity. But when there are less than $K-1$ interferers, then not all the K degrees of freedom are used up for reception and nulling, and hence the remaining DOFs contribute to diversity gain at the receiver. The O-A and A-O links provide an array gain of K, but a reuse factor of K is available only for the link O-A, where A is the receiver. This is because the node must be able to adapt its antenna pattern to form nulls towards interferers in order to leverage spatial reuse. If the receiver is A, then it can estimate interference and form nulls in its reception pattern towards interferers, but if the transmitter is A, then it must determine which other receivers it would potentially interfere with and accordingly form nulls towards them in its transmission pattern. For the transmitter to estimate the

potential interference it is bound to cause is more difficult than the receiver estimating the interference faced. Hence, for practical reasons we do not consider spatial reuse for the A-O link where only the transmitter is an adaptive array node. Again, any possible diversity gain resulting from unused DOFs can be exploited only when A is the receiver, unless the transmitter performs some form of transmit diversity. However, we assume sufficient number of interferers (simultaneous transmitters) to any node and hence all DOFs at a node to be used up for interference suppression. Thus, diversity gain is not considered in adaptive arrays.

The M-M link performs spatial multiplexing to result in a linear capacity increase factor of K. However, with all resources being used up for multiplexing at both the transmitter and receiver, there is no additional spatial reuse and diversity. Since spatial multiplexing requires that both ends of the link be M, no linear capacity increase can be expected in the O-M and M-O links as in the case of spatial multiplexing. Also, no spatial reuse is possible. However, when M is the receiver (link O-M), the MEAs at M can be used to receive multiple copies of the transmitted signal, thereby providing a diversity gain[1] and hence a diversity order of K. Finally, the A-M and M-A links provide an array gain of K due to the adaptive array node, a spatial reuse of K when A is the receiver (link M-A) and a diversity order of K when M is the receiver (link A-M).

While diversity gain improves the reliability of a link, it does not contribute to a significant increase in the steady-state capacity (at low fading losses). Further, since our objective is to exploit the available gains to increase the net capacity, we focus mainly on the SNR and spatial reuse gains provided by a link in the rest of the thesis. Thus, the net gain contributed by a link towards throughput capacity (also referred to as *link gain*) is considered to be given by the product of the capacity increase and spatial reuse increase factors contributed by the link. The spatial reuse factors can be obtained directly from the tables. The capacity increase factors can be obtained from Equation 32 using the SNR gains from the tables, except for a M-M link where it is directly given by K. Thus, the D-D link provides

[1]Diversity gain can contribute to a logarithmic increase in capacity in the presence of channel fading losses.

the maximum link gain in LOS environments ($K^2 \log K^2$), while the A-A link provides the maximum link gain ($K \log K^2$) in NLOS environments, but at the cost of requiring feedback of channel state information at transmitter for its array gain. The O-A and M-A links are capable of providing a better link gain than the M-M link only when the adaptive node is the receiver. Note that, we have considered an open-loop version of MIMO (eg. BLAST [16]) for the M-M links, which does not require receiver feedback unlike the A-A links and is hence less complex. However, if closed-loop MIMO is considered, then the M-M links can also deliver significant gains as the A-A links, although at the cost of increased complexity. Furthermore, it must be noted that, while directional and adaptive arrays deliver a higher throughput capacity gain over MIMO due to their spatial reuse factors of K^2 and K respectively, they do so at the expense of K^2 and K transmissions requiring K^2 and K units of transmission energy respectively, unlike MIMO, which delivers a throughput gain of K for a single transmission and hence single unit of transmission energy. Thus, MIMO would perform the best when throughput per energy is considered as the metric.

8.3 Overview of Capacity Analysis

Having understood the heterogeneous link layer capabilities, the next obvious question that arises is: *Given the heterogeneous nature of links in the network, what performance bounds and gains can one expect from these heterogeneous antenna networks?* We attempt to answer this question in this section through both theoretical analysis and simulations. We study both random and arbitrary* multi-hop wireless networks that are of practical interest. *By random networks, we refer to networks where the node placement as well as the traffic pattern is random, while arbitrary*[2] *networks refer to those with random traffic pattern but involving controlled deployment (location placement) of a fraction of the nodes in the network.* There are several practical reasons and applications for considering arbitrary* networks such as mesh networks, sensor networks, digital battlefields, etc. We first present the model used for the theoretical analysis, followed by a detailed analysis

[2] ** is used to differentiate these networks from networks with both arbitrary node placement and arbitrary traffic pattern.*

of the performance bounds for heterogeneous antenna networks. The relative performance gains provided by the heterogeneous antenna technologies in both random and arbitrary* networks are then evaluated analytically as well as through simulations to draw insights into their performance.

8.4 Scaling Laws for Homogeneous Antenna Networks

The basic model assumed for this analysis is similar to the one used in [20]. Note that the focus of this section is not to provide a fundamentally new capacity analysis for smart antennas, but instead to mainly derive insights into the relative performance benefits of heterogeneous smart antenna networks over homogeneous omni-directional networks. Hence, we adapt and extend the basic analysis in [20] for heterogeneous smart antenna networks. Let us consider a random network operating on a single channel with n nodes independently and uniformly distributed on the surface of a sphere S^2 of unit square area. Assume the nodes to possess omni-directional antennas initially. Each node is capable of generating $\lambda(n)$ bits/sec of traffic to a random destination. Let us consider a generic shortest path routing protocol. Further let the bandwidth of the wireless channel be W bits/sec. We now extend the analysis provided for upper and lower bounds in [20] for omni-directional antenna networks to the case of heterogeneous antenna networks.

The first question that arises when exploring the performance bounds of heterogeneous antenna networks is that *Can the gains provided by the smart antenna technologies overcome the current scaling laws ($\lambda(n) = O(\frac{1}{\sqrt{nlogn}})$) of multi-hop wireless networks?* This question has been answered in the context of directional antennas [41, 77] where it has been shown that the presence of directional antennas at all nodes in the network, cannot alter the scaling laws. We now extend this message to show that smart antennas in general cannot alter the current scaling laws of multi-hop wireless networks; although they can provide a constant order improvement over omni-directional antennas. Further, the amount of improvement depends on the specific smart antenna technology considered. From [20], we know that the following inequality holds for random networks,

$$n\lambda(n)h \leq \frac{4W}{c\pi\Delta^2 r^2} \tag{75}$$

where c is a deterministic constant not dependent on W, n or Δ; h represents the average hop count and r the transmission/reception range; $h = \frac{\bar{L}}{r}$, where $\bar{L}$ represent the average path distance. Δ represents the guard zone parameter which along with r determines the region silenced by a transmission. The term on the left side in inequality 75, $n\lambda(n)h$ determines the total amount of traffic (bits/sec) that needs to be carried by the entire network due to its multi-hop nature. The term on the right side determines the total number of simultaneous transmissions in a channel (or spatial reuse), which can be bounded by $\frac{4}{c\pi\Delta^2 r^2}$. Hence, for the generated traffic to be carried by the network, we need to ensure that the inequality in equation 75 holds good.

The key components that impact per-node throughput capacity are (i) spatial reuse, (ii) hop length, and (iii) channel capacity. However, the factors responsible for the $O(\frac{1}{\sqrt{nlogn}})$ scaling of per-node throughput are spatial reuse and hop length (multi-hop burden). Smart antennas, when used for capacity increase (increased transmission rate), do not alter the hop length[3]. However, they do increase spatial reuse and the channel capacity. But, these improvements are functions that scale with the number of antenna elements employed (K). However, since the number of elements at any node is fixed and cannot scale with the number of nodes in the network for practical reasons, the improvements provide only a constant scaling factor to the throughput capacity. Thus, smart antennas cannot alter the scaling law itself. Further, the net throughput improvement provided by smart antennas is given by the product of their spatial reuse (S_{xx}) and capacity increase (C_{xx}) gains, and hence in turn by their link gain. Thus, the throughput capacity in the presence of smart antennas is given by,

$$\lambda(n) \leq \frac{4W}{c\pi\bar{L}\Delta^2 rn} S_{xx} C_{xx}, \quad where\ x\epsilon\{O, D, A, M\} \tag{76}$$

In LOS environments, switched beam (directional) antennas provide a link gain that is bounded by $K^2 \cdot C_f$, where $C_f = \{1 + \frac{log_2 K^2}{log_2 \rho}\}$, and hence the throughput capacity is

[3]We assume open-loop systems. In closed-loop systems, one can use the antenna gain to provide range extension as well.

bounded by,

$$\lambda(n) \leq \frac{4W}{c\pi\bar{L}\Delta^2 rn} \cdot K^2\{1 + \frac{log_2 K^2}{log_2 \rho}\} \tag{77}$$

In NLOS environments, adaptive arrays provide a link gain that is bounded by $K^2 \cdot \{1 + \frac{log_2 K^2}{log_2 \rho}\}$, resulting in a throughput capacity given by,

$$\lambda(n) \leq \frac{4W}{c\pi\bar{L}\Delta^2 rn} \cdot K\{1 + \frac{log_2 K^2}{log_2 \rho}\} \tag{78}$$

Finally, MIMO links provide a direct linear increase in the channel capacity in NLOS environments, to provide a throughput capacity given by,

$$\lambda(n) \leq \frac{4W}{c\pi\bar{L}\Delta^2 rn} \cdot K \tag{79}$$

However, there is no increase in spatial reuse. But it must be noted that, while switched beam and adaptive arrays deliver a higher throughput capacity gain over MIMO due to their spatial reuse factors of K^2 and K respectively, they do so at the expense of K^2 and K units of transmission energy respectively, unlike MIMO, which delivers a throughput gain of K at a single unit of transmission energy. Thus, when throughput per energy is considered as the metric instead of throughput, switched beam and adaptive arrays will lose out on their corresponding spatial reuse factors, resulting in a gain of just C_f that is much lower than the gain of K in MIMO.

Given that homogeneous smart antenna networks cannot alter the scaling laws, it follows that heterogeneous antenna networks that involve only a fraction of the nodes to be smart, will also not be able to alter the scaling laws of ad-hoc wireless networks. So it now becomes important to see how much of a constant order improvement can these heterogeneous antenna networks actually provide. Specifically, *(i) For random networks, given a fraction of the nodes in the network to be smart, how much of an improvement can one expect when compared to a pure omni-directional antenna network? and (ii) For arbitrary* networks, when controlled deployment is allowed, how should the placement of the smart nodes be and how much of an improvement can they provide?* We answer these specific questions in the rest of this section.

8.5 Throughput Capacity of Heterogeneous Antenna Networks

We consider the upper and lower bounds in random and arbitrary* networks involving both LOS and NLOS environments. A link between a transmitter x and a receiver y is represented as $l(x, y)$ with the transmitter given by $T(l)$ and the receiver by $R(l)$. We refer to the capacity increase resulting from the SNR gain on a link $x - y$ as C_{xy} and the increased spatial reuse factor by S_{xy}. Note that, while $C_{xy} = C_{yx}$, $S_{xy} \neq S_{yx}$ in general. The SNR and spatial reuse gains of the different links can be obtained from the tables in Section 8.2.0.2, while the capacity increase factor for a given SNR gain can be obtained from Equation 32. Let the set N refer to the types of nodes present in the network, which is a subset of $\{o, d, a, m\}$ depending on the environment considered.

8.5.1 Random Networks

8.5.1.1 Upper Bound

Consider the upper bound in the presence of all omni-directional nodes [20],

$$\lambda(n) \leq \frac{4W}{c\pi\bar{L}\Delta^2 rn} \tag{80}$$

There can be atmost $(\frac{4}{c\pi\Delta^2 r^2})$ simultaneous transmissions in the network. Now, the presence of smart nodes in the network has the potential to increase the gain for each of these transmissions, thereby contributing directly to the throughput capacity. In effect, the transmission on a link $x - y$ involving smart node(s) increases the number of simultaneous transmissions by the spatial reuse factor (S_{xy}), as well as the capacity of each of these transmissions by the capacity increase factor C_{xy}, resulting in a net gain of $S_{xy}C_{xy}$, which is referred to as the *link gain* (G_{xy}). However, since the traffic and placement of the smart nodes is random, one needs to find the expectation of the link gain $(\bar{G})$ over the different link combinations that are possible. Hence, the throughput capacity in the presence of smart nodes is now given by,

$$\lambda(n) \leq \frac{4W\bar{G}}{c\pi\bar{L}\Delta^2 rn} \tag{81}$$

$$where,\ \bar{G} = \sum_{x,y\epsilon N} Prob\{T(l) = x\ \&\&\ R(l) = y\}G_{xy};\ and\ G_{xy} = C_{xy}S_{xy} \tag{82}$$

LOS Environments: As reasoned earlier, we need to consider only omni-directional and directional nodes in these environments. Hence $N = \{o, d\}$. Let p_d denote the fraction of directional nodes and p_o denote the fraction of omni-directional nodes ($p_o = 1 - p_d$) in the network. The improvement in capacity in the presence of $p_d n$ directional nodes is given by the expectation of the link gain ($\bar{G}$) over the link combinations of O-O, O-D, D-O and D-D, as

$$\begin{aligned} \bar{G} &= p_d^2 C_{dd} S_{dd} + p_d(1-p_d) C_{od}(S_{od} + S_{do}) + (1-p_d)^2 C_{oo} S_{oo} \\ &= p_d^2 C_{dd} K^2 + 2p_d(1-p_d) C_{od} K + (1-p_d)^2 C_{oo} \end{aligned} \quad (83)$$

NLOS Environments: We consider omni-directional, adaptive array and MIMO nodes in these environments. Hence $N = \{o, a, m\}$. Let p_a denote the fraction of adaptive arrays nodes in the network, p_m denote the fraction of MIMO nodes and p_o denote the fraction of omni-directional nodes ($p_o = 1 - p_a - p_m$). The expected link gain ($\bar{G}$) and hence the improvement in throughput capacity is given by,

$$\bar{G} = p_m^2 C_{mm} + p_a^2 K C_{aa} + p_o^2 C_{oo} + p_m p_a C_{ma}(K+1) + 2p_m p_o C_{mo} + p_a p_o C_{ao}(K+1) \quad (84)$$

8.5.1.2 Constructive Lower Bound

We adopt the construction provided in [20] for omni-directional antennas and extend it to the case of heterogeneous antenna networks. Briefly, the construction for omni-directional antennas involves the generation of a Voronoi spatial tessellation on the surface of a sphere S^2 and is achieved by partitioning a plane with l points into l convex polygons such that every polygon contains exactly one point. Further, any point in a given polygon is closer to its central point than to the central point of any other polygon. The readers are referred to [20] for a detailed discussion on the properties of such tessellations.

The construction basically, consists of several cells. The number of interfering neighbors to a cell is bounded by a deterministic constant c_1 which grows no faster than linearly in $(1+\Delta)^2$, where Δ is the guard zone used by the MAC protocol [20]. Thus, every cell gets a throughput capacity of $\frac{W}{1+c_1}$ bits/sec. Every node chooses a random location and tries to communicate with the node closest to that location. The routing strategy is to make

the route approximate the straight line between the source and the destination. Thus, the cells that are intersected by the routes (straight lines) serve as the intermediate relays for the packets. The transmission range is such that a node in one cell can talk to any node in any of the adjacent cells. Further, the number of routes (lines) served by each cell has been shown to be bounded by $c_2\sqrt{n \log n}$ with a high probability [20]. Also, each of the route carries $\lambda(n)$ bits/sec of traffic. Thus, for a cell to be able to serve all the routes passing through it, we need to ensure

$$c_2\lambda(n)\sqrt{n \log n} \leq \frac{W}{1 + c_1} \tag{85}$$

Now, when smart nodes are introduced into the network, they increase the channel capacity (W) when involved in a transmission, by the capacity increase factor contributed by their link. They also reduce the number of interfering neighbors (c_1) by the spatial reuse factor associated with their link. Thus, the net improvement obtained in the throughput capacity is once again given the expected value of the product of the two factors, resulting in the expected value of the link gain ($\bar{G}$). Further, observing the linear growth of c_1 in $(1 + \Delta)^2$, we obtain that for a constant c_3 not dependent on n, Δ, W, or K the following throughput is achievable with a high probability:

$$\lambda(n) = \frac{c_3 W \bar{G}}{(1 + \Delta)^2\sqrt{n \log n}} \tag{86}$$

Though some of the nodes in the network possess smart antennas, their location as well as the traffic pattern considered are still random. Hence, the improvement contributed by the smart nodes in LOS and NLOS environments to the throughput capacity for this construction ($\bar{G}$) remains the same as the corresponding gains obtained in the case of upper bound.

8.5.2 Arbitrary* Networks

In arbitrary* networks, we assume that one has control over the location placement of a fixed fraction of nodes in the network, while the traffic pattern is still random. While the location placement would not impact the network performance if all the nodes in the network were omni-directional nodes, this is definitely not the case with smart nodes. Hence,

before studying the performance bounds of these networks, we first consider the question: *What is the optimal placement of a set of **smart** nodes given their capabilities, in order to maximize the aggregate network throughput?* The optimal solution to the location placement problem depends on the nature of the smart nodes available as well as their number. In the case of directional and adaptive nodes in LOS and NLOS respectively, a potential solution could be one that maximizes the number of directional/adaptive nodes that every route passes through without increasing the multi-hops burden when compared to shortest path routing. However, in the case of MIMO nodes in NLOS, a potential solution could be one that maximizes the number of M-M links and not just the number of MIMO nodes in every route. Note that determining the optimal location placement of smart nodes given their antenna/link gains is a hard problem in its own right and hence we do not attempt to solve it in this work.

8.5.2.1 Upper Bound

As in the case of random networks, we need to determine the expected link gain $\bar{G}$ that is contributed by the smart nodes to every transmission. However, the location placement of some or all of the smart nodes is now within our control and this must be accounted for in the bounds.

LOS Environments: The optimal routing strategy and placement of directional nodes is one that ensures (i) the intermediate relays on any route are always directional nodes, and (ii) when an omni-directional or directional node is a source or a destination, the other end of the link is always a directional node. While this may not be achievable depending on the number of smart nodes whose location placement is within our control, it would still serve as an upper bound. If h is assumed to be the average hop length of a flow, the upper bound is now given by,

$$\lambda(n) \leq \frac{cW\bar{G}}{\Delta^2 rn} \tag{87}$$

$$where,\ \bar{G} \leq \frac{2}{h}[(p_o)G_{od} + p_d G_{dd}] + \frac{h-2}{h}G_{dd} \tag{88}$$

The solution ensures that every hop of a flow is a D-D link of maximum gain unless the source and/or the destination is an omni-directional node in which case the first and/or last hops would alone be O-D links. Thus, the bound on the expected link gain is determined by conditioning on whether the link considered is the first/last hop or is an intermediate hop as in Equation 88, along with the maximum expected link gain possible in each of the cases.

NLOS Environments: In NLOS environments, A-A link delivers the maximum gain. The capacity is upper bounded by a routing strategy and location placement that has the following properties: (i) all intermediate relay nodes for all flows are adaptive nodes, (ii) when an omni-directional, adaptive or MIMO node serves as a source or destination, the other end of the link is always an adaptive node. Thus, the expected link gain in the upper bound is conditioned as before on whether the link considered is a first/last hop or an intermediate hop, along with their respective maximum gains, and is given by,

$$\bar{G} \leq \frac{2}{h}\{p_m \frac{G_{ma} + G_{am}}{2} + p_a G_{aa} + (p_o)\frac{G_{oa} + G_{ao}}{2}\} + \frac{h-2}{h}\{G_{aa}\} \tag{89}$$

8.5.2.2 Constructive Lower Bound

The construction we provide for arbitrary* networks is an extension to the one used for random networks. We consider a Voronoi spatial tessellation on S^2, with the fraction of nodes whose location can be controlled as being equal to the number of smart nodes available (which is a variable). The generator (central point) of each cell is designated as the relay for the cell and is used to relay all traffic through that cell. Smart nodes are located at these central points and in turn serve as relays. Thus, every source and destination of a h-hop flow are atmost one hop away from a relay node and will hence traverse atleast $h-2$ hops through relay nodes. Hence, the net gain contributed to the transmission in each cell and hence to the throughput capacity, must be conditioned on whether the hop (provided its a smart link) and hence the transmission is on an intermediate hop (consisting of relay nodes on either side of the link) or a first/last hop. Further, since our goal is to provide a construction for an achievable throughput, one must consider different cases of whether the

number of smart nodes available is lesser, equal to, or greater than the number of Voronoi cells (V, required number of relays) in the construction. Let P_x denote the number of smart nodes of type x ($x\epsilon\{d, a, m\}$) as a fraction of the number of Voronoi cells (V),i.e. $P_x = \frac{p_x n}{V}$ and P'_x represents its complement.

LOS Environments: Given $p_d n$ directional nodes, we start assigning these smart nodes to the relay points at the center of each of the cells. However, the cells are chosen at random during the assignment of smart nodes considering the random traffic pattern and the possibility that $p_d n < V$. We have two cases, depending on whether the number of directional nodes is more than or equal, or less than V. Similar to the case of random networks, the following throughput capacity is now feasible with a high probability,

$$\lambda(n) = \frac{c_3 W \bar{G}}{(1+\Delta)^2 \sqrt{n \log n}} \tag{90}$$

However, the expected link gain obtained for every transmission $\bar{G}$ varies from one case to another as follows.

(i) When $p_d n \geq V$, we have,

$$\bar{G} = \frac{2}{h}(1 - \frac{V}{n})\{\frac{(p_d n - V)G_{dd}}{n - V} + \frac{(p_o)nG_{od}}{n - V}\} + \{\frac{h-2}{h} + \frac{2}{h}\frac{V}{n}\}G_{dd} \tag{91}$$

In this case, all the relay nodes are directional. Hence, $h - 2$ hops of every flow will consist of D-D links. The first and last hops are either D-D or O-D links. However, if the source and destination are also relay nodes (with probability $\frac{V}{n}$), then the first and last hops will only be D-D links, as can be seen in Equation 91.

(ii) When $p_d n < V$, we have,

$$\bar{G} = \{\frac{h-2}{h} + \frac{2}{h}\frac{V}{n}\}\{P_d^2 G_{dd} + 2P_d P'_d G_{od} + P'^2_d\} + \frac{2}{h}(1 - \frac{V}{n})\{P_d G_{od} + P'_d\} \tag{92}$$

Here, the relay nodes consist of both directional and omni-directional nodes, while the non-relays are all omni-directional nodes. Hence, all the h hops of a flow can consist of D-D, O-D and O-O links, except if the source and destination are non-relay nodes, in which case the first and last hops can only form O-D and O-O links. This can be seen in Equation 92.

NLOS Environments: In NLOS environments, since links with an adaptive node(s) deliver high gains, we start by assigning an adaptive node to the center of every cell. If the number of adaptive nodes is less than the number of Voronoi cells, then after exhausting the adaptive nodes, we start assigning MIMO nodes to the center of the cells. The choice of the cells in this assignment is once again made at random. We have three possible cases here.

(i) When $p_a n \geq V$, the expected link gain for every transmission is given by,

$$\bar{G} = \{\frac{h-2}{h} + \frac{2}{h}\frac{V}{n}\}G_{aa} + \frac{2}{h}(1-\frac{V}{n})\{\frac{(p_a n - V)G_{aa}}{n-V} + \frac{p_m n}{n-V}\frac{(G_{ma}+G_{am})}{2} + \frac{n(p_o)}{n-V}\frac{G_{ao}+G_{oa}}{2}\}$$

In this case, h-2 hops of every flow will have A-A links with the highest gain. The first and last hops will consist of A-A, A-O or A-M links, except when the source and destination are relay nodes, in which case they form only A-A links.

(ii)When $p_a n < V \leq (p_a + p_m)n$, the expected link gain is given by,

$$\begin{aligned}\bar{G} &= \{\frac{h-2}{h} + \frac{2}{h}\frac{V}{n}\}\{P_a^2 G_{aa} + 2P_a P_a' \frac{G_{am}+G_{ma}}{2} + P_a'^2 G_{mm})\} + \frac{2}{h}(1-\frac{V}{n})\{P_a \cdot \\ &\quad [\frac{((p_a+p_m)n-V)}{n-V}\frac{G_{ma}+G_{am}}{2} + \frac{(p_o)n}{n-V}\frac{G_{ao}+G_{oa}}{2}] + P_a'[\frac{((p_a+p_m)n-V)}{n-V}G_{mm} + \frac{p_o}{n-V}G_{mo}]\}\end{aligned}$$

Here, all h hops of every flow can form A-A, A-M, M-A or M-M links, except when the source and destination are non-relay nodes, in which case, the first and last hops can form M-M, M-A, O-M or O-A links.

(iii) Finally, $(p_a + p_m)n < V$, the expected link gain is given as,

$$\begin{aligned}\bar{G} = \{\frac{h-2}{h} + \frac{2}{h}\frac{V}{n}\}\{P_m^2 G_{mm} + P_a P_m (G_{am}+G_{ma}) + 2P_m(P_m+P_a)' G_{mo} + (P_a)^2 G_{aa} + P_a(P_a+P_m)'(G_{oa}+G_{ao}) \\ + (P_m+P_a)'^2\} + \frac{2}{h}(1-\frac{V}{n})\{P_m G_{mo} + P_a\frac{G_{ao}+G_{oa}}{2} + (P_a+P_m)'\}\end{aligned}$$

The h hops of every flow can have any link combination possible with O, A and M nodes, except when the source and destination are non-relay nodes, in which case they can form only O-A, O-M, or O-O links.

8.5.3 Inferences

We now evaluate the gains provided by random and arbitrary* networks with heterogeneous antenna technologies over that of a network with all omni-directional nodes, through theory

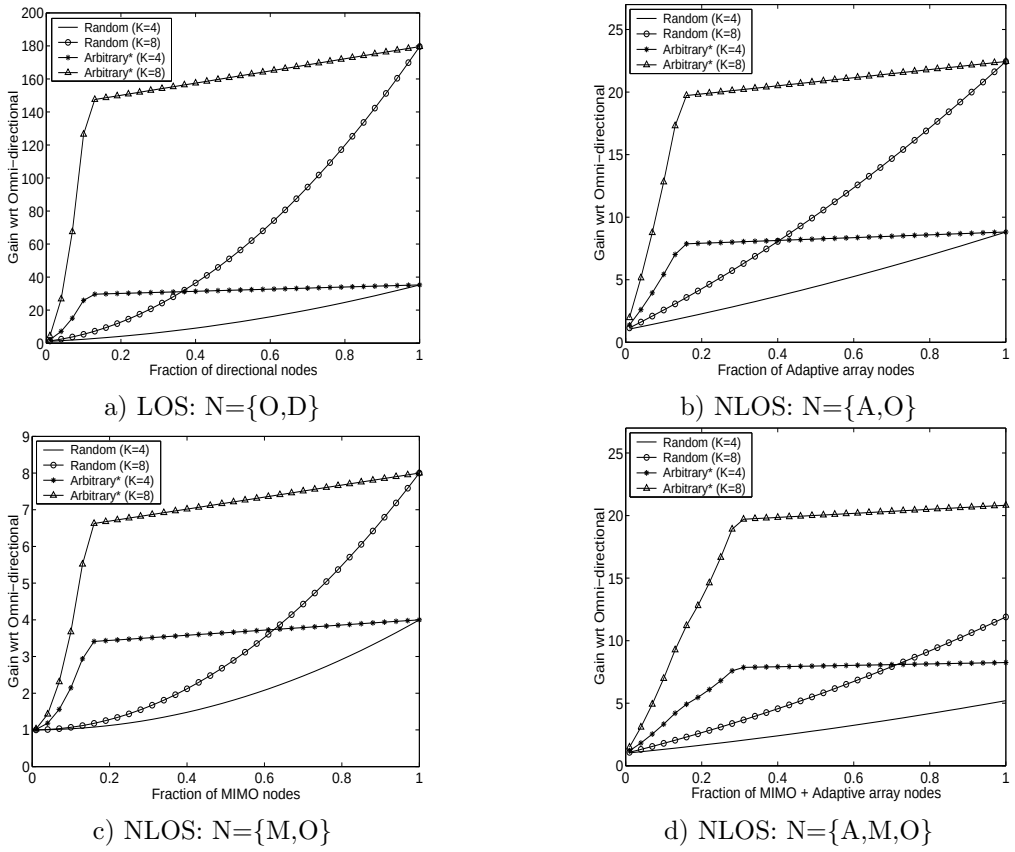
as well as simulations.

For theoretical evaluation, we use the constructions outlined in Section 8.3 for random and arbitrary* networks. Though these constructions may not be the optimal (in throughput), they do serve as reasonable constructions for the two classes of networks. Further, since the constructions and routing strategies for achievable throughput in both these networks were the same, the only difference being the type of relay nodes used in arbitrary* networks, this also helps us use their respective expected link gains for evaluation and hence drawing insights into their performance. The parameters in the construction used for evaluation are as follows: $n = 1000$, area enclosed by a Voronoi cell $= \frac{\log n}{n}$. The number of Voronoi cells for this construction comes up to about 145 cells ($p \approx 0.15$).

For simulations, we use the *ns2* [22] network simulator. The network consists of 100 nodes in a 1000m by 1000m grid. The transmission range is assumed to be 100m. Every node serves as a backlogged source for a flow and chooses a random destination to result in random traffic. For random networks, all the nodes are randomly distributed in the network. Shortest path routing is used for the traffic. A persistence-based contention MAC protocol is used. The two-ray and shadowing channel models available in *ns2* are supplemented with Rayleigh fading to account for mulipath environments. Details on the MAC protocl and physical layer models are presented in Section 10.5. In arbitrary* networks, a relay node is placed at every 100m in both the x and y direction. Hence, there exist a total of 36 relay nodes, which corresponds to a p of 0.36 for all the relay nodes to be smart. Apart from the relay nodes, the remaining nodes are distributed randomly in the network. A shortest path grid routing is used, where the intermediate nodes in a route are formed by the relay nodes. The MAC is the same as that used for random networks. Gain in throughput is used as the metric of comparison. As we can see, the networks considered for theoretical and simulation set-ups are not identical in parameters but are similar in their characteristics. Hence, while we cannot compare them in terms of their absolute results, we can still compare the performance trends exhibited by them.

The theoretical and simulation results for the different networks in different environments are presented as a function of the fraction of smart nodes in the network in Figures

37 and 38 respectively. It can be clearly seen that the *performance trends from simulations corroborate the expected theoretical trends very well.* In addition, the following inferences can be made:



a) LOS: N={O,D}

b) NLOS: N={A,O}

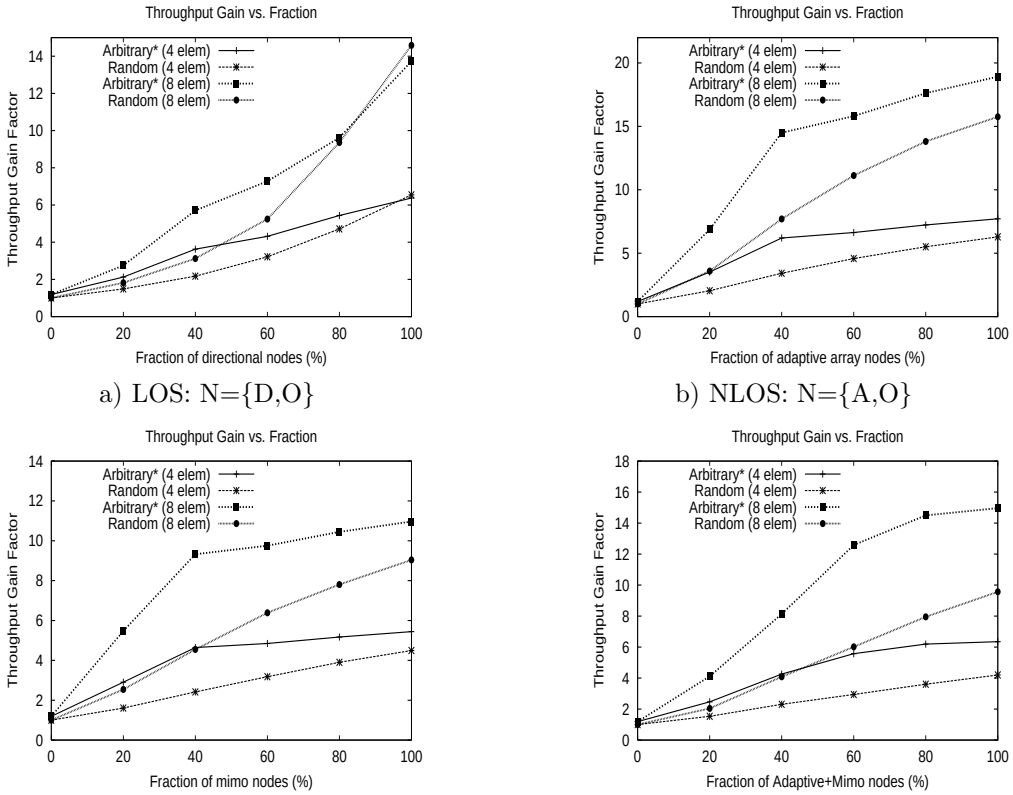
c) NLOS: N={M,O}

d) NLOS: N={A,M,O}

Figure 37: Theoretical Gains

(i) It can be clearly seen that heterogeneous smart antenna networks provide significant performance improvement over conventional homogeneous omni-directional networks and the gain increases with increasing fraction of smart nodes (Figures 37 and 38). While arbitrary* networks do provide more gains than random networks, even for random networks, just about 30% of smart nodes can provide a gain by a factor of 2 over a purely omni-directional random network in both LOS and NLOS environments, which is promising in itself.

(ii) The performance trend in random networks varies with the type of smart nodes. For adaptive arrays in NLOS, the gain with respect to increasing fraction of smart nodes is almost linear since every adaptive array node is capable of providing its share of antenna



a) LOS: N={D,O}

b) NLOS: N={A,O}

c) NLOS: N={M,O}

d) NLOS: N={A,M,O}

Figure 38: Simulation Results

gain irrespective of the nature of other node in the link (Figure 37(b) and 38(b)). In the case of directional (LOS) and MIMO (NLOS), the gain is slightly less than linear for small p but increases rapidly for large p (Figures 37(a),(c) and 38(a),(c)). In LOS, more directional nodes are needed to leverage the K^2 spatial reuse possible from favorable geometry of the nodes, while in MIMO (NLOS), links with both the ends being MIMO-enabled are needed to exploit the spatial multiplexing gain.

(iii) Though the gains for the random and arbitrary* are the same at $p = 0$ and $p = 1$, the gain for arbitrary* is significantly high when $0 < p < 1$. The performance trend has two distinct regions (Figures 37 and 38). A large portion of the gain comes in the initial region where the relay nodes are increasingly populated with smart antennas, thereby providing increasing gain to the flows passing through the relay nodes. The gain starts to saturate once all the relay nodes are smart (around $p = 0.15$ for theory, and $p = 0.36$ for simulations) because *all* the routes start going through links with the highest gain for a very significant

proportion of their path thereafter, providing marginal improvement. Further, the gain of arbitrary* networks over random networks is more with increasing elements, due to their ability to better exploit the available smart antenna capabilities. This indicates the importance of the location placement of smart nodes in heterogeneous antenna networks[4].

(iv) While the theoretical gains are especially high in directional antennas (LOS) than in other antenna technologies (NLOS) owing to the directional spatial reuse factor of K^2 contributed by the D-D links (Figure 37(a)), this is not the case in simulations. In fact, in simulations, random networks outperform arbitrary* networks at larger fractions of directional nodes (Figure 38(a)). Note that, the key contributor to the high gain in directional antennas is the directional spatial reuse factor of K^2. However, for this to be exploited one needs to have atleast K^2 active neighbors who can exploit it. Moreover, their direction of communication must also favor spatial reuse. But in arbitrary* networks, only the relay nodes carry traffic in the network. Thus, while the number of active neighbors depends on the node degree in random networks, it depends on the maximum number of neighboring relays to any relay (relay degree) in arbitrary* networks, which is significantly lower than the node degree in random networks. For a 4-element array, the maximum directional reuse factor possible is 16. In our simulations, the average node degree is twelve in random networks, while the relay degree is only four in arbitrary* networks. Thus, while arbitrary* networks provide high gain when the number of elements is small compared to relay degree, they perform worse than random networks when the number of elements and directional nodes in the network are large compared to relay degree.

(v) The theoretical gains of the arbitrary* and random networks converge at $p = 1$ for all cases but for the combination of MIMO, adaptive and omni nodes in NLOS environments (Figure 37(d)). This is because, arbitrary* networks exploit adaptive nodes better by forming and using more of A-A links in routing than random networks, wherein the formation of an A-A, A-O or A-M link, all happen with equal probability. However, in simulations, the performances of random and arbitrary* networks do not converge when $p = 1$ for any of the cases in NLOS environments. The reason is as follows: In arbitrary* networks, only

[4]The gains obtained in our results are for a simplistic, non-optimal placement of smart nodes.

the relay nodes contend for access to the channel in any time slot. Since the relay degree is much lower than the node degree in random networks, the amount of contention and hence the wastage of bandwidth resources is more in random networks. This is significant at high loads that has been considered for the simulation results. However, when lesser number of flows (eg. 30) are considered, the performances have been observed to converge at $p = 1$.

Thus, arbitrary* networks show potential for significant gains in most conditions even for a moderate fraction of smart nodes in the network (Figure 38), which is very encouraging to network providers especially in applications such as mesh networks, etc. However, care must be taken when the relay degree is less than the directional spatial reuse factor in LOS environments (Figure 38(a)). While controllable smart node placement does bring in significant performance benefits, one cannot expect this flexibility in ad-hoc networks, especially where the nodes could potentially be mobile. Hence, in the rest of this thesis, we focus on random heterogeneous smart antenna networks. But note that, even the random networks deliver a roughly linear improvement in throughput with respect to the fraction of smart nodes added (maximum being at $p = 1$) (Figure 38), thereby providing a great incentive for their consideration.

However, note that the simulation results and hence the predicted gains of random HSAN's from Figure 38 have been obtained by considering naive MAC and routing protocols that are not aware of the heterogeneity present in the environment. In the subsequent chapter, we thus consider the design of heterogeneity-aware MAC and routing protocols for random HSAN's that will help exploit the available heterogeneous smart antenna capabilities in the network in an efficient manner. Further, we do not intend to provide full-fledged MAC and routing solutions that address all possible issues at the MAC and routing layer. But instead, we provide first-cut solutions to MAC and routing protocols that exploit and execute the underlying heterogeneous link gains in their protocol operations. We show that even these first-cut solutions can help exploit the available smart antenna capabilities to provide significant performance benefits, thereby stimulating further research in this area.

CHAPTER IX

HETEROGENEITY-AWARE MAC AND ROUTING PROTOCOLS

9.1 Overview

In this chapter, we present heterogeneity-aware MAC and routing protocols that exploit the available heterogeneous capabilities in the extent in their operations. We first present the heterogeneity-aware MAC protocol (HSMA) followed by the routing protocol (HSR). Both the protocols are then evaluated in isolation and also in conjunction to study the performance benefits possible from such heterogeneity-aware communication protocols.

9.2 HSMA MAC Protocol

We refer to the proposed MAC protocol as the *Heterogeneous Smart antenna based Medium Access control protocol* (HSMA). The goal of the HSMA protocol is to efficiently leverage the capabilities of smart antennas available in the network while adhering to a desired fairness model. The general framework for designing an efficient MAC protocol for heterogeneous smart antenna networks consists of the following three steps: (i) The first step is to enable the MAC protocol to help two nodes realize the complete smart antenna capabilities of the link that can be formed between them. In other words, given a communicating link between two nodes, the MAC protocol should help the link realize the maximum gain that can be leveraged out of the smart link. (ii) The next step is to design a contention resolution mechanism that conforms to a fairness model of interest. (iii) Finally, distributed mechanisms must be incorporated in the contention resolution mechanism to realize any possible optimization considerations that will maximize the performance objective.

9.2.1 Realizing Link Layer Capabilities

In order for a transmitter and receiver to realize the true gain of the smart link between them, they need to exchange information and be aware of each others capabilities. For this reason, and for other reasons specific to multi-hop wireless ad-hoc networks such as hidden terminal problem, etc., we choose to use the RTS-CTS-DATA-ACK handshake mechanism for every packet transmission. The RTS and CTS packets will carry the smart antenna capabilities (including the array size) of the transmitter and receiver respectively, as well as their transmission and reception strategies and also a training sequence for possible channel/direction estimation if desired by the nodes. Once the transmitter informs the receiver of its capabilities in its RTS, the receiver taking into account its own capabilities, decides the optimal strategy of link operation (for DATA transmission and ACK reception by transmitter) and conveys this decision to the transmitter in its CTS. The appropriate strategy is then used by the transmitter for its DATA transmission and ACK reception. Since some of the nodes in the network are going to be omni-directional, all RTS and CTS are transmitted omni-directionally. Note that, since every smart node employs a MEA, it is capable of forming an omni-directional pattern. We now present the specifics of the exchange in the MAC protocol for realizing the different link combinations. Note that only the SNR gains will be realized through these exchanges. The spatial reuse component of the gain needs to be leveraged through specific local coordination mechanisms in addition to the basic contention resolution mechanism.

9.2.1.1 LOS Environments

The MAC protocol needs to realize the directional gain available in O-D and D-D links. Every directional node runs a simple *direction of arrival* (DOA) estimation algorithm in its receiver signal processing unit. The training sequences in the RTS and CTS are used by the transmitter and receiver to determine the direction of stronger signal strength in the transmission and hence the direction towards each other. Since MEAs are used, the fixed set of weights corresponding to every beam direction is pre-computed. Once the direction between the transmitter and receiver is determined, it is mapped to the appropriate set

of weights for focusing energy during DATA and ACK transmissions (SNR gain of K^2). Further, even if only one of the nodes is directional, it can still use its DOA algorithm to determine the beam direction towards its opponent and use it to beamform energy during its communication and hence realize its individual directional gain (K).

9.2.1.2 NLOS Environments

If the receiver node of a link is an adaptive array node (x-A link), then the transmitter (O or M) needs to send a training sequence in the RTS. This is used to estimate the channel coefficients between the transmitter and receiver and hence determine the set of coefficients to weight the signals received from the antenna elements that will maximize the SINR on the link (array gain, K). An adaptive array receiver can use a linear minimum mean squared error (LMMSE) receiver processing for this purpose. If transmitter is an adaptive node (A-x link), then irrespective of the antenna technology used by the receiver, the receiver needs to estimate the channel coefficients and feed this information back to the transmitter in its CTS. This information is used by the transmitter to determine the appropriate set of weights and hence the beam pattern that will maximize the SINR on the link (transmitter beamforming gain, K). If however, both the transmitter and receiver are adaptive (A-A link), then maximum beamforming gain (K^2) is obtained from the link.

If the transmitter and receiver are both MIMO enabled (M-M link), then this information is exchanged during RTS and CTS communication. No channel estimation and hence training sequences are required and the link can operate in open-loop with the transmitter using BLAST for transmitting K independent data streams and the receiver using successive interference cancellation (SIC) receiver processing for decoding the K independent streams, resulting in a capacity increase factor of K. If however, only the transmitter is a MIMO node (M-x link), then the MIMO node behaves like an omni node in its transmissions, but has to send a training sequence in its RTS when involved in an M-A link for the adaptive receiver to estimate the channel and beamform. If however, only the receiver is MIMO enabled (x-M), then the MIMO node needs to estimate channel coefficients and feed them back to the transmitter in an A-M link for the adaptive transmitter to beamform.

However, this is not necessary in an O-M link. Further, the MIMO receiver when used with an omni or adaptive array transmitter (A-M and O-M links) can also use its multiple (K) elements to receive multiple copies of the transmitted signal to provide a diversity gain (diversity order = K) that will in turn improve link reliability significantly. Note that since a pair of nodes communicating for the first time would not be aware of each other's antenna capabilities, training sequences are sent by default on the RTS and CTS packets. However, this can be made optional for subsequent transmissions between the same pair of nodes. The frequency of sending such training sequences can be restricted to once every L packets, where L is a system parameter that can be set based on the channel fading characteristics and network dynamics (mobility).

We consider open-loop (no feedback of channel state to transmitter) operation for almost all the heterogeneous link combinations for the purpose of keeping the processing time low at the receiver so that it does not interfere with the protocol operations. Only for the case where the transmitter is an adaptive array node, we consider closed-loop to allow the transmitter to adaptively beamform. Even in this case, the receiver side channel estimation algorithms are standard and can be performed efficiently. The issue arises only in the overhead of CSI on the CTS packet, to reduce which the feedback can be sent once every L packets between the same pair of nodes as in the case of training sequences.

9.2.2 Contention Resolution Algorithm

While max-min fairness has been shown to be unsuitable for ad-hoc networks from an efficiency perspective [43], proportional fairness proves to be a good compromise between efficiency and fairness and a suitable candidate for use in ad-hoc networks [43, 37, 72]. Hence, the fairness model that we adopt for our MAC protocol is proportional fairness. Further, since some links in our network contribute more to the channel capacity than others, we adopt a weighted proportional fairness model to maximize the aggregate network capacity utilized. We now proceed to explain these choices in detail.

9.2.2.1 Overview

Consider a utility function $U(s)$ for every (MAC level) flow that indicates the "satisfaction" received by the flow for a service s received. Now if the utility functions are continuous, differentiable, increasing and strictly concave over the range $s \geq 0$, then it has been shown [37] that the problem of fair channel allocation equivalently reduces to maximizing the aggregate utility of the network. Specifically, it has been shown that if the channel allocation rate a_i for each flow i changes according to the equation,

$$\dot{a_i} = \alpha - \frac{\beta p_i}{U'(a_i)} \tag{93}$$

then the network utility and the flow utility functions converge to their optimal value, and the system converges to the optimal point. Further, the nature of the fairness model depends on the utility function chosen. Recently, the area of network utility maximization has been used for designing random access MAC protocols for ad-hoc networks in [72] and [31] for the case of log-utility (proportional fairness) and general classes of concave utility functions respectively.

Proportional fairness has been shown to be a good compromise between efficiency and fairness and a suitable candidate for use in ad-hoc networks [37, 43], thereby forming our choice[1]. Further, since some links in our network contribute more to the channel capacity than others, we adopt a weighted proportional fairness model to maximize the aggregate network capacity where the smart links access the channel with a higher probability than the non-smart ones. Note that, trying to provide equal bandwidth fairness to the different links irrespective of their contribution to the network capacity, will not help us exploit the capabilities of the available smart nodes efficiently. If the smart links were to exploit the available antenna gain for reliability or transmit power reduction then this would not affect the packet transmission duration which would remain to be the same for all the links. However, since the smart links use the gain for rate increase, this reduces the channel usage time for a packet (fixed size) transmission by a factor that is proportional to the antenna gain. Hence, a higher priority of the smart links in accessing the channel (by a

[1]One can also adopt other fairness models and our mechanisms are not specific to the fairness model.

factor equal to their antenna gain) is needed to compensate for their reduced duration of usage of channel bandwidth, consequently resulting in a notion of fairness with respect to the channel access time. This is similar to the approach followed in [49] where channel utilization was maximized by allowing links with good channels to operate at higher rate while at the same time ensuring that all the links obtained a fair share of *channel access time* by making the good links transmit multiple back-back packets at a high rate.

For weighted proportional fairness, the aggregate utility becomes $\sum_i U(a_i) = \sum_i w_i \log(a_i)$ where w_i in pricing terminology is the price paid by user i to the network, which in our case corresponds to the gain contributed by the link to the network. Thus, the rate adaptation to be followed by each flow reduces to [37],

$$\dot{a_i} = \alpha w_i - \beta p_i a_i \tag{94}$$

In equation 114, α and β are system parameters corresponding to a utility constant and a penalty constant respectively. p_i is the contention loss probability experienced by the flow; w_i is the weight assigned to the flow that is proportional to its link gain (product of capacity and spatial reuse gains) contribution, and $\dot{a_i}$ is the rate of change of allocation. A node can belong to multiple contention regions and can potentially experience different loss probabilities in each of them. The net loss probability for a node is thus given by the sum of the loss probabilities in each of the contention regions that it belongs to. However, for any contention region *in isolation*, two considerations are vital: (i) the sum of the allocation rates in any contention region (region where all flows contend with each other) must not exceed one, and (ii) each flow in a contention region must experience the same loss probability. This helps the system converge to the stable throughput maximizing point [37] and also help result in: (i) approximating the allocation rate (a_i) with a persistence probability with which each flow decides whether to contend for a slot or not, and (ii) each flow uses the same maximum backoff value to result in a fair contention loss probability in the contention region. Hence, a_i is hereafter referred to as persistence probability in the rest of our discussions. Having the rate adaptation in the persistence domain instead of in the backoff domain also simplifies the implementation aspect of the protocol to a great

extent.

9.2.2.2 Channel Access

We extend the basic distributed persistence algorithm for omni-directional antenna networks in [37] to heterogeneous antenna networks, supplementing it with all the necessary mechanisms required to leverage the antenna gains. The state maintained at each node consists of the persistence probability, the system constants α and β and the flow's (link's) weight w_i. There are four possible states in which a node can be, namely *Contend*, *No_Contend*, *Acquire*, and *Co_Sched*. Every node i having a packet to transmit, first decides to contend for the channel with a probability of a_i. If the node succeeds, it moves from the *No_Contend* state to the *Contend* state, where it chooses a waiting time uniformly distributed from the interval (0,B). The node then waits for the backoff period (in slots), after which it tries to access the channel to see if the channel is busy. If the node finds the channel to be busy (a transmission has already begun), it gives up the slot and decrements its persistence by β. Similarly, if the channel is idle but if the node detects a collision, it decrements its persistence by β. On the other hand, if the node finds the channel to be idle, and does not experience any collision then it moves to the *Acquire* state where it transmits. At the end of the slot, all the nodes having a packet to transmit in the next slot increase their persistence by αw_i. Thus, the links that contribute more gain (eg. A-A, D-D links, etc.) will increment their persistence values more than those with a relatively lesser gain and hence contend with higher persistence to compensate for the reduced channel usage at higher transmission rates. The values for α and β are empirically set to 0.07 and 0.3 in our simulations.

While the above algorithm describes the contention resolution mechanism that exploits the SNR gain of the links, mechanisms are needed to enable the directional and adaptive nodes exploit their spatial reuse capabilities. This is accomplished by the *co-ordinated scheduling* mechanism wherein, when a directional/adaptive node finds the channel to be idle and acquires access to it, it also coordinates the schedule of other directional/adaptive links in the same time slot to exploit their spatial reuse capabilities. The co-ordinating node thus moves from the *Acquire* state to the *Co_Sched* state.

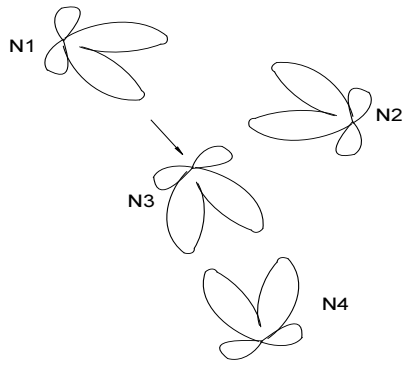


Figure 39: Deafness Example

9.2.2.3 Co-ordinated Scheduling

Exploiting spatial reuse in directional and adaptive arrays through purely localized and distributed mechanisms is a challenging task, which when not executed properly could degrade the performance of the system worse than that of one without spatial reuse. For example in adaptive arrays, consider an ongoing transmission on a link between two adaptive nodes, N1 and N2 as in Figure 39. It is possible for another adaptive link between N3-N4 to transmit at the same time by forming nulls towards the nodes N1 and N2. While this would not cause interference to the existing transmission between N1-N2, however, when the link N1-N2 is done with its transmission, it is not aware of the transmission that had begun between N3 and N4. Now, if N1 wishes to communicate with N3 then it would not obtain a response from N3 since N3 has formed a null towards N1. N1 would assume a collision and unnecessarily decrement its persistence counter. Even worse, if only N3 (transmitter) is within the transmission range of N1 and N2 but N4 (receiver) is outside the transmission range although within the carrier sensing range. Then N4 will not have the channel coefficient to form nulls towards N1 and N2. Hence, a transmission by N1 towards N3 later on could result in a collision at N4. The end result is a lot of wastage of time slots and hence bandwidth resources. The analogy to this problem in directional antennas is what is popularly referred to as the *deafness* problem [10]. The problem is exactly the same as in adaptive arrays except that the nulls correspond to the inactive beam directions in directional nodes.

Essentially, if the algorithm is aggressive in exploiting spatial reuse, then it exacerbates the deafness problem, resulting in a degradation in throughput. The problem is all the more difficult in heterogeneous antenna networks, since some of the nodes are omni-directional. So we adopt a conservative approach to exploiting spatial reuse. When an adaptive or directional node overhears a transmission in its channel, it is possible for it to independently decide to transmit based on if it has sufficient resources (in the case of adaptive arrays) to form nulls with respect to existing transmissions, or if its direction of communication does not interfere with those of the existing transmissions (in the case of directional). Even if the node decides that it is possible to transmit, the algorithm does not allow the node to transmit in this case since this would result in the deafness problem which in turn has a much more severe impact on throughput. All control packets carry a training sequence using which every node in the one-hop neighborhood can determine the DOA (in LOS) and channel coefficients (in NLOS) between the transmitter and itself. Further, links where both nodes are directional or adaptive are capable of effectively exploiting spatial reuse than those with a single smart node. Thus, every D-D or A-A link keeps track of other potential D-D or A-A links in its one-hop neighborhood and their corresponding DOA information (directional) or channel coefficients (adaptive) in a table. Hence, when an adaptive/directional link finds the channel to be idle, it schedules itself, but also checks its table to see if other adaptive/directional links in its one-hop neighborhood can be scheduled in the same slot without causing interference to its transmission given its degrees of freedom. If other such links can be scheduled, then the link appends a "co-ordinate schedule" directive to its RTS and CTS along with the list of links and the priority with which they should be scheduled in the same slot. The priority is used to avoid unnecessary collisions between the links in the list. Every link that gets a directive, is capable of transmitting only if there is no other transmission in its channel except for the one that gave the co-ordinated schedule directive, to avoid the deafness problem. In addition, in the case of directional links, it also checks to see if it has DOA information for other links with a higher priority in the coordinated schedule, and if so, whether its communication would not interfere with their direction of communication. In the case of adaptive links, it checks its table for

channel coefficients to links with higher priority in the schedule and if present, estimates if it will have sufficient resources after trying to null them out as interference. If the above tests are positive, the link decides to obey the directive and schedules itself. Thus, spatial reuse is exploited primarily through co-ordinated scheduling rather than nodes trying to independently schedule multiple simultaneous transmissions, thereby also alleviating the deafness problem. Unlike the co-ordinating link, all the other co-ordinated links that did not obtain *direct* access to the channel, continue to decrement their persistence counter for that time slot.

Note that the goal of HSMA is not to provide a full-fledged MAC solution but to identify the key elements in a heterogeneity-aware MAC protocol, namely (i) realization of heterogeneous link layer capabilities, and (ii) incorporation of heterogeneous link gains in contention resolution to compensate for the reduced channel usage resulting from exploitation of smart antenna gain for increased transmission rates. Thus, one could use solutions proposed in other MAC protocols for homogeneous smart antenna technologies [10, 27, 65, 35], etc. in tandem with the elements identified in HSMA to further improve the performance of heterogeneity-aware MAC protocols. Further, the incorporation of heterogeneous link gains in contention resolution is not specific to the persistence domain adopted in HSMA, but can also be accommodated in the backoff domain followed by CSMA/CA-variant protocols by appropriate adaptation of *contention window* parameter based on the heterogeneous link gain.

9.3 HSR Routing Protocol

We now present a routing protocol for the target environment called *Heterogeneous Smart Antenna based Routing* (HSR). Since reactive routing protocols are preferred to proactive routing protocols in mobile ad-hoc networks, where network dynamics are common and energy resources are a premium, we design HSR as a reactive source routing protocol along the lines of the dynamic source routing (DSR) protocol. However, HSR is antenna technology aware and hence leverages the heterogeneous antenna capabilities available in the network in its routing mechanisms to improve network performance. We now explain the

role of route discovery and route maintenance, the two main components of any reactive routing protocol, in HSR.

9.3.1 Route Discovery

9.3.1.1 Route Metric

The goal of the route discovery component is to obtain "quality" routes by carefully exploiting the capabilities of smart antenna nodes in the network. We define "quality" of a route as its ability to allow for maximum spatial reuse in the *network* while at the same time incurring low multi-hop relaying burden and providing high rate for the *flow* itself. We consider a three-tuple route metric $M(R)$ (=$(M_1(R),M_2(R),M_3(R))$) for a route R as the combination of network $(M_1(R),M_2(R))$ and flow $(M_3(R))$ metrics. $M_1(R)$ captures the total area inhibited (along with duration of inhibition) when a packet travels *along the path* from source to destination. However, since the same transmission range is used by all links in the network, $M_1(R)$ effectively reduces to the total time a packet physically occupies the channel when it travels from source to destination along a path. $M_2(R)$ captures the total number of additional transmissions (spatial reuse contributed by directional and adaptive nodes) that can be enabled in the network when a packet travels along a path. Thus, while $M_1(R)$ tries to capture the degree of spatial reuse inhibited in the network by using larger transmission ranges, $M_2(R)$ captures the additional spatial reuse contributed by the smart nodes in the route to the network. $M_3(R)$ is considered to be the minimum of the rates used by the links in the path, thereby indicating the direct benefit to a flow. Using $M(R)$, routes are chosen lexicographically, based on a low-valued first metric, high-valued second metric and high-valued third metric, in that order of priority.

$$\begin{aligned} M_1(R) &= \sum_{i=1}^{h} \frac{Area\ inhibited\ by\ i^{th}\ hop\ Tx}{Tx\ rate\ at\ i^{th}\ hop} = \sum_{i=1}^{h} \frac{f_i^2}{C_i} \\ M_2(R) &= \sum_{i=1}^{h} S_i \qquad M_3(R) = min\{C_i\} \end{aligned} \tag{95}$$

where f_i is the transmission range used at the i^{th} hop of a h hop route, normalized to the range of an omni-directional transmission. Since all transmissions use the same omni-directional range, $(f_i = 1,\ \forall i)$. C_i and S_i correspond to the capacities and spatial reuse

available on hop i respectively. If a hop i is composed of nodes x and y, then $C_i = C_{xy}$ and $S_i = S_{xy}$. $M(R)$ biases towards the network metrics since the goal is to maximize the aggregate network throughput. However, note that minimizing $M_1(R)$ would in turn result in choosing routes with lesser hop length and larger hop capacities, thereby indirectly also helping increase the individual flow's throughput. Thus, $M_1(R)$ is given more priority than $M_2(R)$ in choosing a route. Different link layer combinations of smart antennas yield different gains. Thus, the metric tends to favor routes with lesser hop length and larger gains on the hops constituting the route. Hence, a route with all A-A links or all M-M links would be chosen in preference to one with O-A or A-M links (for similar hop lengths).

9.3.1.2 Route Request and Response

When a source has a packet to send to a destination to whom it does not possess a route, it initiates a route request. The route request is propagated as a flood in the network. Every node receiving the request, checks to see if it is the intended destination. If not, it appends its identity as well as its capability (antenna capabilities and number of elements possessed) to the request packet and forwards the packet as a broadcast. Once the destination receives the packet, it has the list of all intermediate nodes between the source to itself and hence the route. In addition, it also has the information of all the antenna capabilities of the intermediate nodes. The route thus obtained is reversed to send a unicast route reply back to the source indicating the discovered route. Further, the information regarding the antenna capabilities of all the constituent members of the route is also piggy-backed onto the route reply.

Once the source receives the route reply, it obtains a route to the destination. In addition, it can determine the link layer capabilities of the intermediate hops in the path based on the piggy-backed information. The source then determines the quality of the discovered route based on the metric in Equation 95. If the source discovers multiple routes to the destination, then it evaluates the routes based on the 3-tuple metric to determine the best route to send its packets.

While the route metric helps choose routes with large gains for the data packet transmissions, an interesting optimization would be to see if one can exploit the smart antenna gains to perform faster route discoveries. This would mean exploiting the antenna gains for faster propagation of route requests and responses. Since route requests mark the beginning of a flow and are broadcast packets without an intended next hop, nodes that forward requests do not have prior information on how to use their antenna capabilities. Further, since the goal is to ensure that all neighbors receive the route request (to ensure best routes are chosen), it becomes *necessary to consider omni-directional nodes and hence forward the route request as an omni-directional transmission.* Thus, we choose to use omni-directional transmissions for the propagation of route requests.

However, for the unicast transmission of route reply back to the source, we exploit the gains along the path. Since the route reply packet carries the antenna capabilities of every node in the path, when a node x receives the route reply, it looks up the route to find the next forwarding node y. In addition, it also looks up the antenna capabilities of y and decides on the link combination that can be executed between them. x and y then exchange their capabilities through RTS and CTS packets to mutually agree and execute a link combination possible between them with the largest link gain (rate). Thus, the route reply propagates quickly back to the source, thereby reducing significantly the route discovery latency.

9.3.2 Route Maintenance

The basic mechanism of route maintenance is similar to that in DSR, wherein the MAC layer is used to detect link failures (after several retransmissions of RTS and/or DATA packets) and report route errors back to the source. The source on receiving a route error, deletes the route from its cache and initiates a new route request for the destination. The larger the time a source spends without a route to the destination, the greater is the throughput degradation. Further, the intermediate nodes on the path also delete the route being used to the destination. In addition, they also purge the packets buffered in their queue, that were intended for the destination, thereby wasting the bandwidth resources already used by

those packets. Thus, route failures result in a significant throughput degradation. While the reason for link failures could be congestion or mobility losses, the latter is very pronounced in mobile environments which is common in ad-hoc networks. Hence, HSR's goal is to exploit the smart antenna capabilities to alleviate route failures resulting from mobility.

The underlying idea in HSR to alleviate mobility induced route failures, is to exploit the available antenna gain for range extension instead of for capacity (rate) during route maintenance. Since an SNR gain (increase in mean SNR) on the link can be directly used for range extension, this strategy is possible in adaptive and directional links due to their array and directional gains respectively. However, this is not possible with pure MIMO links where the increase in capacity comes directly through spatial multiplexing. Even with directional/adaptive links, it is possible to exploit range extension, only if the transmitter is a directional/adaptive node. This is because, the strategy requires the transmitter to reduce its rate of transmission by switching to a lower modulation rate, thereby increasing the mean SNR on the link and hence the transmission range. Once the transmitter is able to get through to the receiver, the receiver (if directional/adaptive) can also be instructed to reduce its rate for data reception to further increase the transmission range on the link. If however, only the receiver is directional/adaptive, then the transmitter has no means of informing the receiver of the required rate reduction in the first place. Further, the increase in range extension comes at the cost of sacrificed rate, and the reduction in rate required depends on the degree of mobility in the network. Nevertheless, the reduction in rate on the links to alleviate mobility-induced link failures, is a small price to pay compared to the throughput degradation suffered from route failures. This will be evident from the evaluation studies in Section 10.5.

When a receiver y moves out of range of a transmitter x, x would start experiencing RTS retransmissions. HSMA uses upto maximum seven RTS retransmissions (similar to IEEE 802.11) before declaring the link dead. Initially, all control packets are sent at the same rate as the DATA packet. If x is a directional (in LOS environments) or adaptive (in NLOS environments) node, then after experiencing four such consecutive RTS retransmissions, x shifts to a lower modulation rate that is next to the one currently being used. The reduction of

rate merely increases the decodeability (reliability and range) of the RTS packet without altering the interference pattern. Hence, if the reduction of rate (increase in range/reliability) is able to help the receiver successfully decode a packet, which was not possible otherwise, then with a high probability either the channel is bad or the receiver/transmitter is moving out of range rather than due to contention. However, x does not immediately conclude that y has potentially moved out of range[2], but instead, x switches back to the older (previous) rate for the next packet. If the next packet also requires a lower rate as the current packet, then x assures itself that the previous losses were not due to contention and confidently generates a proactive route error message to the source. If on the other hand, the lower rate is also not able to get the packet across, then x keeps reducing its rate till it reaches that of an omni-directional transmission and eventually generates a route error as in the normal case. The intermediate nodes that receive a proactive route error do not delete the route being used. The source on receiving a route error with the proactive flag set, initiates a new route request but continues to use the existing route till the new route is discovered. Further, the intermediate nodes do not respond to a route request (for a destination) with a cached route, for which a proactive route error has been issued. Thus, the time spent by the source without a new route is significantly reduced due to the proactive operation. Further, even a range extension factor of 1.5 to 2 (with respect to omni-directional range) that is possible with small number of elements (about 2-4), provides a margin of several couple of seconds to about ten seconds before the link actually breaks, considering transmission ranges of 100-200 m and speeds of even 10-20 m/s. This provides sufficient time for the source to determine an alternate route and also helps keep the reduction in rate on the links small. Further, since the transmitter is able to sustain the link from failure due to mobility for a sufficient time by extending its range, the wastage of bandwidth resources due to the purging of packets in the buffer of intermediate nodes using the route is significantly reduced. Thus, the careful exploitation of the antenna gain for range extension during route maintenance in HSR significantly alleviates the throughput degradation due

[2]The channel quality could have degraded instead of the node moving out of range, which still needs to be detected as a non-contention loss.

to mobility-induced route failures.

Designing a full-fledged routing protocol that exploits the available smart antenna capabilities in the network is a relatively new but a complete problem in itself and there have been very few works in this regard [48, 9]. Hence, HSR does not attempt to address all potential issues, which is out of the scope of this work, but instead serves as a first-cut solution with the following basic elements, namely (i) a heterogeneity-aware route discovery procedure with an appropriate metric to discover routes with large gains while also taking hop-length into consideration, and (ii) a route maintenance component that exploits the antenna gains available on the heterogeneous links to help in route patching during network dynamics such as node mobility. Improving upon the basic elements of HSR, as well as adding additional functionalities such as quicker route discoveries (exploiting the increased rates possible from heterogeneous links) to it, forms part of our future work.

9.4 Isolated Performance Evaluation

In this section, we evaluate the performance of the antenna technology aware MAC (HSMA) and routing (HSR) protocols proposed in Sections 10.4 and 9.3. We use *ns2* network simulator for all our simulations. The topologies consist of 100 nodes in a 1000m by 1000m grid. The transmission range used is 100m. The mobility scenarios are generated using the *setdest* tool and random waypoint mobility model is used for the purpose. The parameters used in the study are (i) fraction of smart nodes in the network, (ii) load (# flows) and (iii) mobility of nodes. The number of antenna elements considered at each of the nodes is either four or eight elements. Aggregate Throughput, measuring the number of packets successfully received by the destinations of all multi-hop flows in Kbits/sec, is the primary metric of interest. We also use normalized standard deviation (standard deviation normalized to mean) of the flows' throughputs to observe the fairness properties of the MAC protocol. The sources and destinations are chosen at random. UDP is used as the transport protocol. Every source generates traffic at a significantly high rate to remain back-logged for the entire simulation duration. The MAC and routing protocols used as baseline protocols for comparison are explained subsequently. The simulations are run for 100 secs and each of

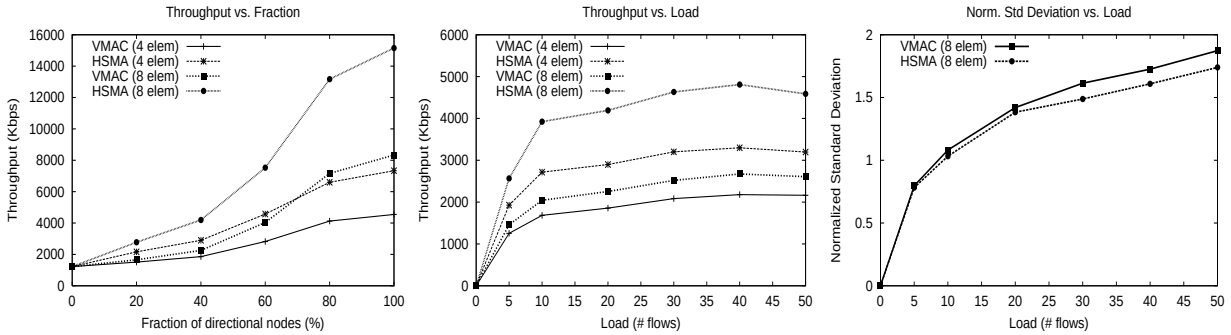
the data point in the graphs presented is averaged over 20 seeds.

9.4.1 Physical Layer Model

We consider two environments of study: LOS and NLOS environments. In NLOS environments, we consider an equal proportion of adaptive array and MIMO nodes. We use the two-ray propagation and shadowing physical models available in *ns2* for the LOS environment. However, for the NLOS environment where we consider multipath scattering, such a simple model would not suffice. To emulate the link characteristics present in multipath environments, we run Matlab simulations for the scenarios considered in the presence of Rayleigh fading with different types of smart antenna processing at the nodes. Thus, we obtain the link-level packet loss statistics based on the different SINR's for different (i) configurations (positions) of the (desired) transmitter-receiver pair and interferers, (ii) receiver processings (linear minimum mean square error and successive interference cancellation) used, and (iii) number of elements used in the MEA. We then incorporate these statistics into the physical layer model of *ns2* for our simulations. Every node uses a constant transmit power of 280 mW and the SINR threshold on the link is set to 10 dB. The frequency of operation is 2.4 GHz and the channel data rate is 2 Mbps. We evaluate the performance of the MAC and routing protocols in isolation before studying their combined advantages.

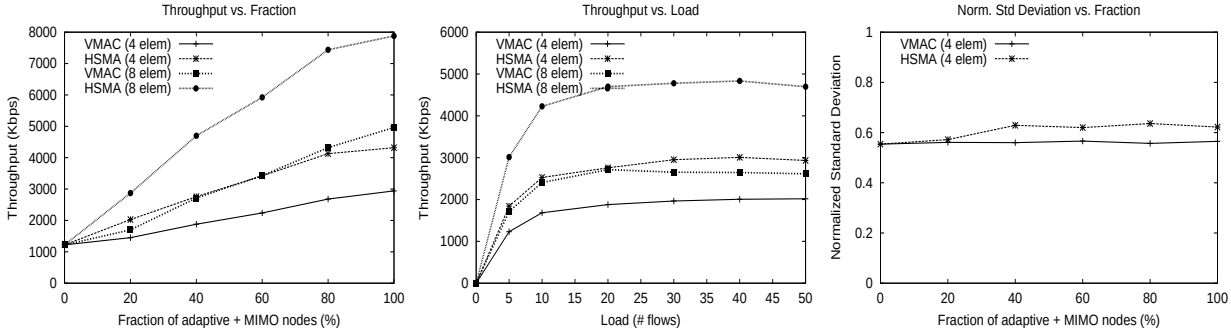
9.4.2 HSMA Performance

We refer to the baseline MAC protocol considered as *vanilla* MAC (VMAC). VMAC is a naive MAC protocol wherein a link can leverage its antenna capabilities only if either ends of the link (transmitter and receiver) possess the same antenna technology. If the transmitter and receiver are of different antenna capabilities, then VMAC cannot coordinate to leverage the best combination between them and would fall back to an omni-operation. Thus, O-O , D-D, A-A and M-M are the only combinations that can leverage the smart antenna gains with this MAC protocol. Further, VMAC does not have the capability to perform intelligent co-ordinated scheduling to effectively leverage the spatial reuse available with directional and adaptive array nodes. We consider a simple shortest path routing protocol for both the MAC protocols to isolate the performance of HSMA over VMAC. The throughput



a) Throughput: (Variable,20,0) b) Throughput: (40%,Variable,0) c) Normalized Std: (40%,Variable,0)

Figure 40: HSMA Performance: LOS Environments: (Fraction, Load, Speed)



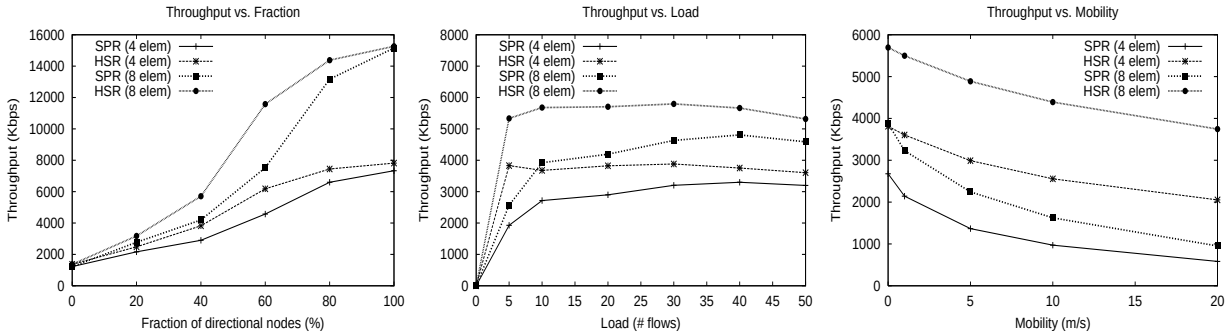
a) Throughput: (Variable,20,0) b) Throughput: (40%,Variable,0) c) Normalized Std: (Variable,20,0)

Figure 41: HSMA Performance: NLOS Environments: (Fraction, Load, Speed)

and fairness results for LOS and NLOS environments are presented in Figures 40 and 41 respectively.

The throughput results are presented as functions of fraction of smart nodes in the network and load in Figures 40(a) and (b) and 41(a) and (b). They indicate that HSMA provides significant gains of about 50-80% in LOS and about 30-50% in NLOS environments, due to its ability to better leverage the available antenna gains (capacity as well as spatial reuse) from a link. Hence, the improvement increases with increasing elements. Further, the gain is more in LOS due to the less severe environment, and hence the large spatial reuse available with directional antennas. In the case of LOS, the gain is large when the fraction of directional nodes p_d tends towards one (Figure 40(a)) since larger number of directional nodes are required to effectively leverage the K^2 spatial reuse theoretically possible.

While a distributed MAC protocol can ensure fairness between links contending in a region, it is not possible for it to ensure fairness between multi-hop flows in a distributed



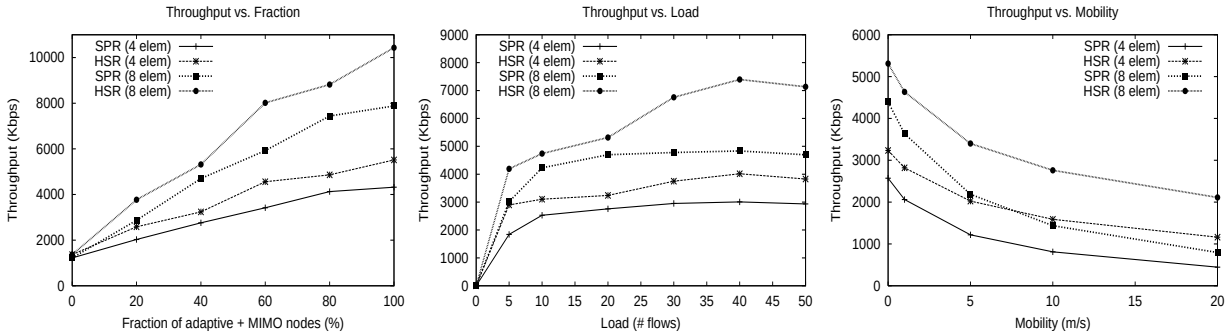
a) Throughput: (Variable,20,0) b) Throughput: (40%,Variable,0) c) Throughput: (40%,20,Variable)

Figure 42: HSR Performance: LOS Environments: (Fraction, Load, Speed)

environment. Since the goal of HSMA is to maximize the aggregate network utilization, we need to ensure that this is not achieved at the cost of significantly penalizing certain flows in the network. Hence, we use the normalized standard deviation as the unfairness index to compare the distribution of throughputs provided by HSMA against that of VMAC. For lack of space, we present only a subset of the fairness results in Figures 40(b) and 41(b). The degree of fairness provided by HSMA has been observed to be atmost 10% worse than that of VMAC. This indicates that HSMA does not penalize flows any more than VMAC, and hence its performance improvements are directly attributed to its ability to intelligently leverage the smart antenna capabilities available in the network.

9.4.3 HSR Routing Performance

We now evaluate the performance of the routing protocol HSR against that of a simple shortest path routing protocol that is not antenna technology aware and hence determines the same set of routes as in an omni-directional environment. In order to isolate the performance of the routing protocol, we consider HSMA as the MAC for both the routing protocols. Throughput performance of HSR and SPR are evaluated as functions of the fraction of smart nodes, load and mobility of nodes. The results are presented in Figures 42, and 43 for the LOS and NLOS environments respectively. HSR, by discovering routes with high antenna gains, is able to provide an improvement of about 30-50% in static environments. In the presence of mobility, the gain due to the route maintenance mechanism is significant measuring about three-four folds.



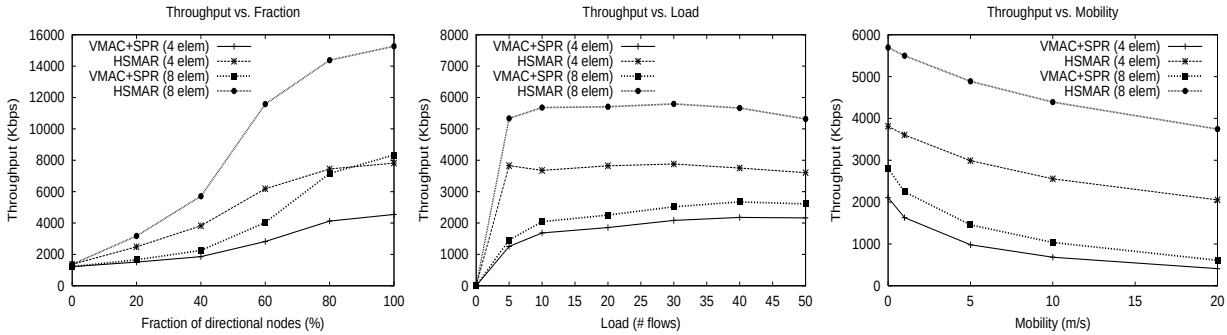
a) Throughput: (Variable,20,0) b) Throughput: (40%,Variable,0) c) Throughput: (40%,20,Variable)

Figure 43: HSR Performance: NLOS Environments: (Fraction, Load, Speed)

While the throughputs of SPR and HSR converge for LOS environments when all nodes in the network are smart (Figure 42(a)), this is not the case for NLOS environments (Figure 43(a)). This is because, in LOS environments, $p_d = 1$ corresponds to all nodes being directional, thereby requiring no intelligent route discovery. The gain of HSR would in this case be more for moderate fractions where it is able to obtain high gain routes for its flows unlike SPR. However, in NLOS environments, even when 100% of the nodes are smart, 50% are adaptive nodes while 50% are MIMO nodes. Since, the links involving adaptive nodes, contribute more than the links involving MIMO nodes, an intelligent selection of adaptive and MIMO nodes on the link makes a difference even if all the nodes in the network are smart.

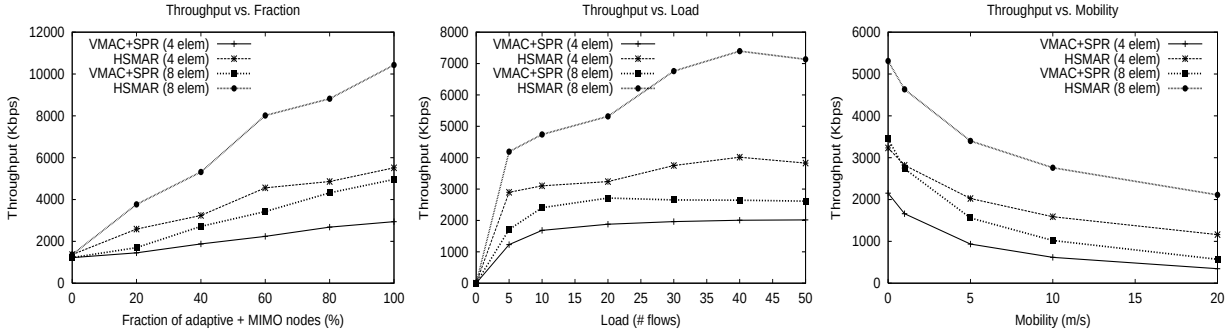
The throughput results, as a function of load (Figures 42(b) and 43(b)), indicate that most of the gain for HSR (and hence saturation) is achieved at lower loads in LOS itself, unlike in NLOS where it happens at higher loads. This can be attributed to the spatial reuse factor. In LOS, directional spatial reuse is effectively exploited only when there are large number of directional neighbors. However, the results are for 40% of nodes being directional and randomly distributed. Hence, this significantly limits the spatial reuse that can be exploited. However, in NLOS, since the adaptive nodes are capable of adapting their beam pattern, they can better leverage spatial reuse (upto a factor of K) and hence saturate at higher loads than in LOS.

The route maintenance component of HSR helps the throughput degrade gracefully with increasing node speeds unlike in SPR, thereby contributing to significant gains at



a) Throughput: (Variable,20,0) b) Throughput: (40%,Variable,0) c) Throughput: (40%,20,Variable)

Figure 44: Net Performance: LOS Environments: (Fraction, Load, Speed)



a) Throughput: (Variable,20,0) b) Throughput: (40%,Variable,0) c) Throughput: (40%,20,Variable)

Figure 45: Net Performance: NLOS Environments: (Fraction, Load, Speed)

large speeds (Figures 42(c) and 43(c)). Further, the gain is more in LOS than in NLOS, since most of the links in LOS (O-D,D-D) can help in proactive route failure detection and avoidance unlike in NLOS, where only links involving an adaptive node are capable of helping. Thus, it can be clearly seen that exploiting the antenna gains for range extension at the cost of rate to alleviate route failures, prevents significant throughput degradation.

9.5 Combined Performance Evaluation

Finally, we have also evaluated the net performance improvements provided by HSMA and HSR together (HSMAR). The results are presented for the same set of experiments in Figures 44 and 45. The baseline considered is a combination of the naive VMAC and SPR protocols to study the maximum benefits one can derive from heterogeneous antenna technology aware MAC and routing protocols. Results indicate that significant gains of over 100% can be achieved by considering antenna technology aware solutions. We just

summarize the key inferences here. (i) The major contributor to the improvement is the MAC protocol. This is because, the MAC layer is the one that is immediately impacted by a change in the physical layer technology and is hence also the most important layer that needs adaptation to be able to leverage the true potential of smart antennas. (ii) The significant advantage provided by the route maintenance component during node mobility, indicates that protocols must be tuned to adapt their strategy of operation (how the gains should be leveraged - rate increase, range extension, etc.) based on the perceived network conditions. (iii) Finally, while we observed in Section 8.3 that random networks can provide a promising improvement in throughput (theoretically) that is a linear function (in NLOS) of the fraction of smart nodes introduced, to be able to achieve such an improvement one must design an antenna technology aware MAC protocol at the very least.

CHAPTER X

COOPERATION IN HETEROGENEOUS SMART ANTENNA NETWORKS

10.1 Overview

In this chapter, we investigate and motivate the need for a simple form of node cooperation, also popularly referred to as *retransmit diversity*, in heterogeneous smart antenna networks (HSANs). We first show that while such a simple form of node cooperation cannot bring significant benefits to *homogeneous* omni-directional antenna networks, they can bring great gains (several folds improvement) to *heterogeneous* smart antenna networks, thereby motivating the need for their exploitation. In this work, the term "smart" refers to beamforming antennas, which would include switched beam antennas in *line of sight* (LOS) and fully adaptive array antennas in multipath scanttering environments (NLOS) respectively. All the antenna technologies considered, are assumed to employ multi-element arrays (MEAs). We consider two forms of practical HSAN's: (i) random networks - traffic as well as node placement are random; and (ii) arbitrary* networks - traffic is random but node placement is controllable atleast for some nodes. One of the main examples of arbitrary* networks is wireless mesh networks, which are becoming increasingly popular lately. In arbitrary* HSAN's, one can expect the controllable nodes (eg. mesh routers/relays) to be made up of smart nodes and function as the backbone routing infrastructure for all the traffic, thereby providing significant gains to all the flows. We investigate the performance gains that node cooperation can bring to both these networks and present several interesting properties and results in the process. Our contributions and results can be summarized as follows:

- We show that the gains from cooperation are always more for random networks than for arbitrary* networks, indicating that random networks have a much larger potential

for leveraging cooperation. This makes node cooperation a natural tool for exploitation in random networks. This also contributes to the following effect: While it is obvious that arbitrary* networks will provide much better performance than random networks owing to their control over node placement and routing, we show that leveraging cooperation provides a means to bridge this performance gap between the two networks without resorting to any autonomous control over node placement.

- For random networks, we show that the gain obtained from cooperation is a concave function in the number of smart nodes in the network. This implies that with increasing number of smart nodes, the (SNR/throughput) gain from cooperation[1] increases initially when the degree of heterogeneity increases, peaks and then starts to decrease when the network tends towards a homogeneous smart antenna network.

- While one would expect the gain from cooperation to be the same for a *homogeneous* omni network and a *homogeneous* smart antenna network, we show that it is possible for the homogeneous smart antenna network to achieve a lower cooperation gain than the homogeneous omni network. We identify that the increasing spatial sensitivity of transmissions due to increasing number of smart nodes is responsible for decreasing the potential for cooperation.

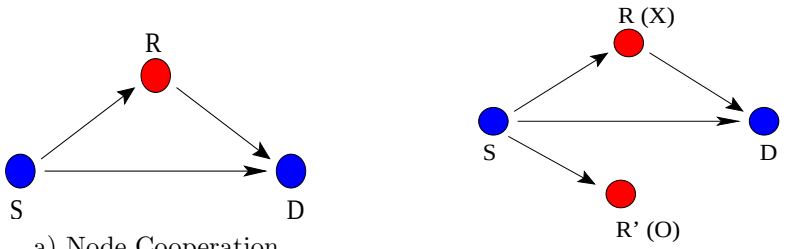
- We identify that the increasing spatial sensitivity of transmissions due to increasing number of smart nodes in the network, decreases the potential for cooperation. Consequently, this results in a tradeoff between exploiting smart antenna gain and cooperation gain and the relative importance of the two gains varies with the nature of fading (fast or block fading) network conditions. We then present a strategy that attempts to leverage both the cooperation and antenna gains appropriately based on fading conditions, without requiring the estimation/knowledge of fading statistics.

- Finally, we present a distributed, localized MAC protocol that addresses all challenges

[1]Improvement in SNR/throughput over a network with no cooperation.

arising in the implementation of the proposed strategy, thereby leveraging the available cooperation and antenna gains in the network to provide significant performance improvements.

10.2 Node Co-operation in HSAN's



a) Node Cooperation

b) Node Cooperation in HSAN

Figure 46: Illustrations

In this section, we introduce and study the impact of node cooperation in (random) HSAN's. By node cooperation, we refer to the process of neighboring node(s) of a transmitter assisting the receiver in successful reception of packets in the event of losses resulting from fading. Consider the simple topology in Figure 46(a) where transmitter/source S tries to send a packet to receiver/destination D. Let R be an intermediate node within communication range of both S and D. Assuming omni-directional transmissions, node R is capable of receiving the packet from node S when it transmits to D due to the *wireless broadcast advantage*. Let the transmission from S to D fail due to fading while R successfully receives the packet. It is possible for S to try retransmitting the packet. On the other hand, node R can assist in the retransmission. There are two advantages to node R re-transmitting instead of node S: (i) R may be closer to D and hence has a possibly lower path loss and higher link gain, (ii) if fading is time-correlated (existing over several consecutive packets) it is futile for node S to retransmit; however the channel between R and D is independent of that between S and D. Hence, it is possible for R to succeed with a higher probability than S.

What essentially happens here is that retransmit diversity is indirectly built into the system by way of exploiting neighboring nodes that possibly have a better channel gain in the event of fading. There are several advantages to increasing diversity in the system since

it can indirectly improve the performance (throughput) and energy efficiency (required error performance at lower SNR's) of the system.

10.2.1 Motivation

Note that, such a simple form of retransmit diversity may not necessarily provide performance improvement in homogeneous omni-directional networks if the relay does not have a better link gain to the destination than the source, and if the fading is fast and independent from one packet to another. This has motivated researchers to consider more sophisticated forms of cooperation diversity such as distributed space-time codes, virtual MIMO, etc. [30, 11, 40] in omni-directional networks. Such sophisticated approaches deliver good diversity gains at the cost of requiring synchronization, distributed code design, rate and/or power control amongst the cooperating nodes, which prevents their distributed implementation from being light-weight. While such sophisticated approaches are warranted in omni-antenna networks, we show that even the simple form of retransmit diversity presented in the example above can provide significant performance improvement and hence has incentives to be exploited in heterogeneous smart antenna networks.

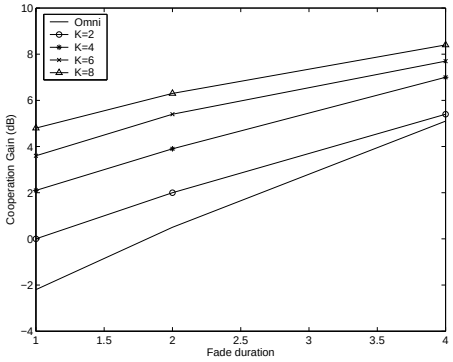


Figure 47: Cooperation Gain

The key elements in a HSAN are the smart nodes that contribute a SNR gain on their links and hence serve as sources of diversity gain that can be exploited by omni-directional nodes during conditions of fading. Whenever links involve omni-directional transmitters and have a smart node as their neighbor, exploiting the neighboring smart node for cooperation delivers significant gains than when an omni-directional neighbor is used for cooperation.

The choice of high gain smart links over the low gain omni links for retransmission also ensures gains during conditions of fast fading. In Figure 46(b), when the link S-D suffers from fading losses (S being an omni transmitter), the smart node R is exploited for cooperation rather than the omni node R'. Further, the cooperation mechanism exploits only one relay for cooperation, keeping the coordination mechanism simple and easily deployable. The gain in SNR from the cooperation mechanism (over non-cooperation) in a homogeneous omni antenna network and a heterogeneous smart antenna network with 20% smart nodes is presented in Figure 47 as a function of fade duration and number of elements. This clearly indicates the incentive for exploiting retransmit diversity in HSANs, wherein diversity gains are the result of exploiting the available smart nodes in the network better.

10.2.2 Theoretical Analysis

Having motivated the need for node cooperation in HSANs, we now present some theoretical analysis that will help us establish properties of the cooperation mechanism.

10.2.2.1 Network Model

We consider a mixture of smart and omni nodes in the network, the fraction of smart nodes being given by p_x. We consider two forms of networks: random and arbitrary*. In arbitrary* networks, a small set of nodes, whose location can be controlled, form a routing backbone infrastructure (as in mesh networks) and all flows route their data through the backbone. We assume that the number of smart nodes available in the network is sufficient such that all forwarders in the routing backbone are smart nodes.

We consider block (time-correlated) Rayleigh fading, where the fading (and hence channel coefficients) is Rayleigh from block to block but remains the same within a block. The length of the block could vary from a single packet to several packets (order of several tens of milliseconds [71]). We consider a practical version of loss recovery at the MAC layer where a data packet is retransmitted during fading losses for only a fixed number of times F (eg. four in IEEE 802.11 DCF). We also assume that any available antenna gain is used for reliability. Exploiting antenna gain for rate is considered later on in Section 10.3. The two basic versions of communication that we consider are:

No Cooperation (NC): The transmitter continues to (re)transmit the packet on fading loss using its normal strategy of operation without any change for a maximum of F trails. If the link involves a smart node then the smart antenna gain on the link would contribute to reliability.

Cooperation (C): The transmitter transmits using its normal strategy of operation. On experiencing a fading loss, if there is a neighbor within the communication pattern of both the transmitter and receiver, then that node can potentially receive the packet from the transmitter due to wireless broadcast advantage and hence relay the packet (on successful decoding) to the receiver. In any case, the number of retransmissions for the packet (including transmitter and relay) is bounded by F, after which the packet is dropped. In the absence of a relay, the operation is the same as that of non-cooperation.

10.2.2.2 Average Outage Probability

We use outage probability as a measure to compare the different schemes. Outage probability refers to how often (probability) does the BER (or equivalently SINR) experienced falls below a certain threshold. It is both a popular and practical measure for robustness to fading [40, 30] especially for block fading where it can directly be related to frame/packet error rate. It can be measured by determining the probability that the mutual information of communication (capacity) is less than the information rate. Since most applications require a specific outage probability to be satisfied, we also compare the mechanisms based on the SNR required to achieve a desired outage probability.

We consider an independent, quasi-static, frequency non-selective Rayleigh fading (complex channel coefficients being uncorrelated and circularly symmetric Gaussian random variables with zero mean and unit variance) along with free space path loss for our channel model. Let a channel realization between S and D be denoted by H_{SD}. Now, the mutual information for a given channel realization in the case of basic omni-directional communication is in general given by,

$$C(H_{SD}) = \log_2 (1 + SNR|H_{SD}|^2) \tag{96}$$

where SNR is the average received signal-noise ratio (based on path loss). If $\tilde{R}$ is the

information rate (bits/sec) applied to the system (normalized to bandwidth) then outage probability can be given by,

$$P_{out}(SNR) = Pr[C(H_{SD}) < \tilde{R}] = Pr[|H_{SD}|^2 < \frac{1}{SNR}(2^{\tilde{R}} - 1)] \tag{97}$$

By way of definition of $|H_{SD}|$, $|H_{SD}|^2$ follows an exponential distribution. Hence, on averaging the outage probability over all possible channel realizations, we get,

$$P_o(SNR) = 1 - e^{-\frac{1}{SNR}(2^{\tilde{R}}-1)} = 1 - e^{-\frac{T}{SNR}} \tag{98}$$

where $T = 2^{\tilde{R}} - 1$. We assume that the maximum number of trials for every packet is bounded by F.

Let f be the fade duration (number of consecutive fading packet intervals). The outage probability on a link is determined by the information/data rate of operation (R), the antenna gain (G) and reduction/increase of path loss (in case of relays alone) contributing to the received SNR. To simplify expressions, we define the following term for non-cooperation:

$$Q_{nc}(G) = 1 - e^{-\frac{T}{G \cdot SNR}}, \;\; G = \{1, K, K^2\} \tag{99}$$

10.2.2.3 No Cooperation - Random Networks

Since we have a combination of smart and omni nodes in the network, the outage probability on a link must be conditioned on the nature of the link. Also for a fade duration f with maximum number of trials being F, the outage probability on the i^{th} trial (t_i) will be dependent on the outage probabilities of the previous f trials. For eg., when $F = f = 3$, if the first trial fails, the remaining trials will also fail with probability one. Thus, the general form of outage probability can be given as,

$$P_o(SNR) = \sum_{(xy)} Pr\{xy\} P_{o,xy}(t_1) P_{o,xy}(t_2|t_1) ... P_{o,xy}(t_F|t_{F-1} ... t_{F-f}) \tag{100}$$

where $(xy) = \{oo, ox, xo, xx\}$; x being a smart node.

The outage probability for the non-cooperative scheme (P_{nc_R}) can now be given as,

$$P_{nc\_R} = p_o^2 Q_{nc}(1)^{\lceil \frac{F}{f} \rceil} + 2 p_o p_x Q_{nc}(K)^{\lceil \frac{F}{f} \rceil} + p_x^2 Q_{nc}(K^2)^{\lceil \frac{F}{f} \rceil} \tag{101}$$

The total SNR used in the scheme for transmission for a desired outage probability can be calculated from the outage probability expression by replacing $Q_{nc}^{\lceil \frac{F}{f} \rceil}(y)$ in Equation 101

with $S_{nc}(y) = \sum_{i=1}^{F} Q_{nc}^{\lceil \frac{i-1}{f} \rceil}(y)$ as,

$$SNR_{nc\_R} = p_o^2 S_{nc}(1) + 2p_o p_x S_{nc}(K) + p_x^2 S_{nc}(K^2) \tag{102}$$

where SNR_t is the SNR consumed in transmission on a single trial.

10.2.2.4 Cooperation - Random Networks

In the case of cooperation, the relay link will also be involved in the outage probability analysis and the possible path loss reduction/increase on the relay links with respect to source and destination must be incorporated. The change in path loss (d_g) normalized to that of the S-D link can be captured using the relative distances on the links as follows,

$$d_{g1} = \left(\frac{d(S,D)}{d(S,R)}\right)^{\gamma}; \quad d_{g2} = \left(\frac{d(S,D)}{d(R,D)}\right)^{\gamma} \tag{103}$$

where γ is the path loss exponent. Since cooperation involves both the S-R and R-D links in determining outage probability (atleast one of them fails), it can be captured using the following term,

$$Q_c(d_{g_1}, G_1, d_{g_2}, G_2) = 1 - e^{-\frac{T}{d_{g_1} G_1 SNR} - \frac{T}{d_{g_2} G_2 SNR}} \tag{104}$$

where d_{g_1} and d_{g_2} are expected path loss gains for the relay links assuming a uniform distribution of the relay location in the intersection area of the communication patterns of the source and destination; G corresponds antenna gain on the link. For simplicity, we represent the outage probability only using the antennas gains as $Q_c(G_1, G_2)$ hereafter.

Here, the outage probability must also be conditioned on the availability of a cooperating (relay) node as well as on its type (smart/omni). Let P_r denote the probability of finding a relay within the communication range of both the transmitter and receiver. For a smart transmitter, the relay must not just lie within the communication range but also within the spatially sensitive communication pattern of the transmitter (P_{rx} and P_{ro} for smart and omni transmitter respectively). The expected outage probability can now be given as,

$$\begin{aligned} P_{c\_R} = p_o^2\{P_{ro}Q_{nc}(1)Q_c^{\dagger}(1,1) + \bar{P_{ro}}Q_{nc}(1)^{\lceil \frac{F}{f} \rceil}\} + p_o p_x\{P_{ro}Q_{nc}(K)Q_c^{\dagger}(1,K) + \bar{P_{ro}}Q_{nc}(K)^{\lceil \frac{F}{f} \rceil}\} \\ + p_x p_o\{P_{rx}Q_{nc}(K)Q_c^{\dagger}(K,1) + \bar{P_{rx}}Q_{nc}(K)^{\lceil \frac{F}{f} \rceil}\} + p_x^2\{P_{rx}Q_{nc}(K^2)Q_c^{\dagger}(K,K) + \bar{P_{rx}}Q_{nc}(K^2)^{\lceil \frac{F}{f} \rceil}\} \\ Q_c^{\dagger}(x,y) = p_o Q_c(x,y)^{\lceil \frac{F-1}{f} \rceil} + p_x Q_c(K \cdot x, K \cdot y)^{\lceil \frac{F-1}{f} \rceil} \end{aligned} \tag{105}$$

The average SNR expended in transmission for a required outage probability can be obtained by replacing $Q_c^{\lceil \frac{F-1}{f} \rceil}(x,y)$ in Equation 105 with $S_c(x,y) = \sum_{i=1}^{F-1} Q_c^{\lceil \frac{i-1}{f} \rceil}(x,y)$ and $Q_{nc}^{\lceil \frac{F}{f} \rceil}(y)$ with $S_{nc}(y)$ and scaling the expression by SNR_t.

10.2.2.5 Arbitrary Networks*

The difference with respect to random networks is that the types of links are not just based on the nature of the node but also on the location of the node. Since the forwarders on the routing backbone are made of smart nodes, there exist three types of links: (i) backbone links which are all smart; (ii) source (first hop) links where the receiver is smart while the source may be smart/omni; and (iii) destination (last-hop) links where the sender is smart while the receiver may be smart/omni. Let V be the number of nodes in the backbone, $p_b = \frac{V}{n}$ represent the probability that a node belongs to the backbone, and $p_h = \frac{h-2}{h}$ represent the probability that a link is a backbone link. Also, p'_o and p'_x represent the probability of finding an omni and smart nodes outside the backbone, which are given as $\frac{p_x n - V}{n-V}$ and $\frac{p_o n}{n-V}$ respectively. Now, the outage probability for no cooperation and cooperation schemes can be given as,

$$P_{nc\_A} = \bar{p_h}\bar{p_b}\{p'_x Q_{nc}(K^2)^{\lceil \frac{F}{f} \rceil} + p'_o Q_{nc}(K)^{\lceil \frac{F}{f} \rceil}\} + \{p_h + p_h p_b\} Q_{nc}(K^2)^{\lceil \frac{F}{f} \rceil} \quad (106)$$

$$P_{c\_A} = \bar{p_h}\bar{p_b}p'_x\{P_{rx}Q_{nc}(K^2)Q_c^{\dagger}(K,K) + \bar{P_{rx}}Q_{nc}(K^2)^{\lceil \frac{F}{f} \rceil}\} + \frac{\bar{p_h}}{2}\bar{p_b}p'_o\{P_{rx}Q_{nc}(K)Q_c^{\dagger}(K,1) + \bar{P_{rx}}Q_{nc}(K)^{\lceil \frac{F}{f} \rceil}\}$$
$$+ \frac{\bar{p_h}}{2}\bar{p_b}p'_o\{P_{ro}Q_{nc}(K)Q_c^{\dagger}(1,K) + \bar{P_{ro}}Q_{nc}(K)^{\lceil \frac{F}{f} \rceil}\} + p_h p_b\{P_{rx}Q_{nc}(K^2)Q_c^{\dagger}(K,K) + \bar{P_{rx}}Q_{nc}(K^2)^{\lceil \frac{F}{f} \rceil}\}$$

where $Q_c^{\dagger}(x,y)$ terms are similar to those in Equation 105 with p_x and p_o being replaced by p'_x and p'_o respectively. The corresponding SNR expressions can be obtained following the procedure adopted for random networks.

10.2.3 Properties

The main metric we use to evaluate the cooperation mechanism is the cooperation gain, which represents the amount of SNR gain (dB) obtained over the non-cooperation scheme. Let G_{c_R} and G_{c_A} represent the cooperation gain in random and arbitrary* networks

respectively. For a desired outage probability, the gains are given as $G_{c\_R} = SNR_{nc\_R} - SNR_{c\_R}$ and $G_{c\_A} = SNR_{nc\_A} - SNR_{c\_A}$. In establishing the properties, we use asymptotic analysis in the limit of large n and K wherever applicable for the proofs. However, the properties themselves are applicable to normal values of n and K.

Lemma 3 $G_{c\_R} \geq G_{c\_A}$

Proof: In the limit of large n, G_{c_R} and G_{c_A} can be shown to reduce to,

$$\begin{aligned} G_{c\_R} &= P_{ro} g_1(p_x, Q_{nc}, Q_c, K, f) + P_{rx} g_2(p_x, Q_{nc}, Q_c, K, f) \\ G_{c\_A} &= P_{rx} g_3(p_x, Q_{nc}, Q_c, K, f) \end{aligned}$$

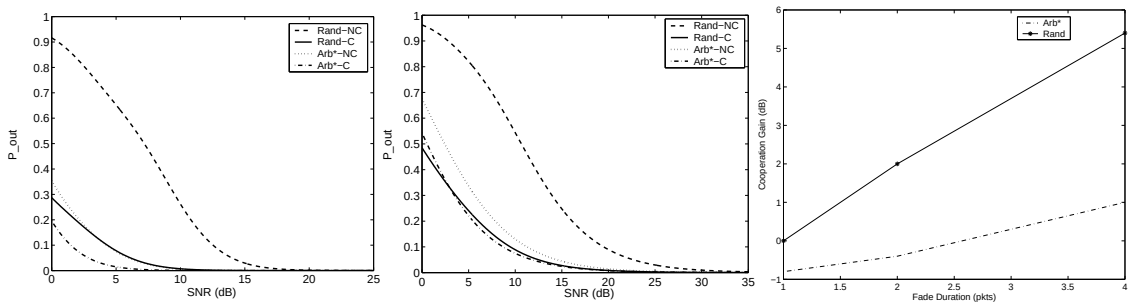
where g_1, g_2, g_3 are deterministic functions that are ≥ 0. Now, in the limit of large K, the spatial sensitivity of the communication pattern of smart nodes increases, thereby decreasing the probability of finding a relay, $P_{rx} \to 0$. However, P_{ro} remains unaffected since the omni directional patterns remains unaltered. Thus, in the limit of large n and K, we have,

$$\begin{aligned} G_{c\_R} &\to P_{ro} g_1(p_x, Q_{nc}, Q_c, K, f) \\ G_{c\_A} &\to 0 \end{aligned} \tag{107}$$

Thus, we have $G_{c\_R} \geq G_{c\_A}$ ■

Property 1 *The gains from cooperation are always more for random networks than for arbitrary* networks, indicating that random networks have a much larger potential for leveraging cooperation.*

Note that the gains from cooperation must not be confused with the absolute performance. It is straight-forward that arbitrary* networks will provide a much better (throughput) performance than random networks owing to their control over node placement and routing. However, when it comes to exploiting cooperation, the smart nodes are already exploited to the best possible extent in arbitrary* networks and hence there is not much room left for improvement through cooperation. Thus, leveraging cooperation also provides random networks a means to bridge the performance gap between the two networks without resorting to any autonomous control over node placement.



a) P_out: p=30, f=1 b) P_out: p=30, f=3 b) Gain: p=20, K=2

Figure 48: Outage and Gain Results

This can also be verified from the numerical results for outage probability presented in Figures 48(a) and (b), where the number of elements considered at each of the nodes is four. While there is a huge gap in performance between the non-cooperation schemes of arbitrary* and random networks, the gap is significantly reduced for the cooperation schemes. The cooperation gain as a function of fade duration for $K = 2$ and $p = 20\%$ in Figure 48(c) indicates the significant potential of random networks to exploit cooperation, which increases with increasing fade duration. Note that, all the numerical results presented in this chapter are averaged over numerous possible geometric configurations of the source, destination and relay.

Lemma 4 *G_{c_R} is a concave function in the (fractional) number of smart nodes in the network, $p_x \in (0, 1]$.*

Proof: In the limit of large K, we apply $P_{rx} \to 0$ to G_{c_R} and take the first $(\frac{d}{dp_x})$ and second $(\frac{d^2}{dp_x{}^2})$ derivatives of G_{c_R}. We then find the interval of p_x for which the $\frac{d^2}{dp_x{}^2} < 0$. Thus, the interval of p_x for which G_{c_R} is a concave function in the fraction smart nodes is obtained as,

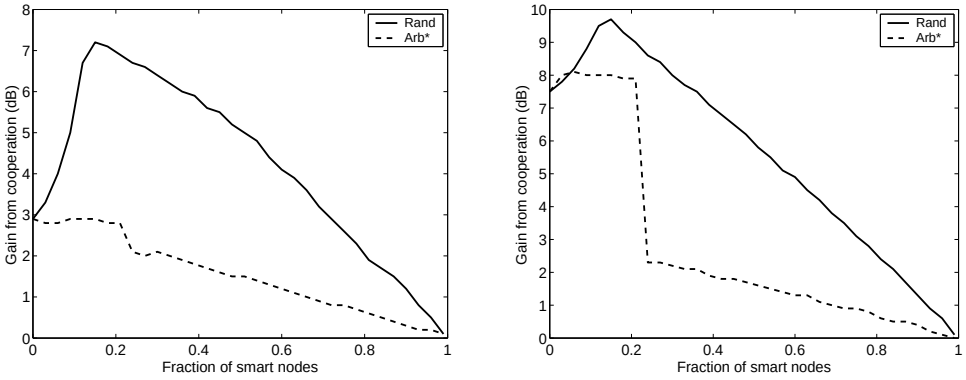
$$p_x > \frac{1}{3}\frac{Q_{nc}(1)A + Q_{nc}(K)B + S_{nc}(1) - S_{nc}(K)}{Q_{nc}(1)C + Q_{nc}(K)D} \tag{108}$$

$$A = 2S_c(K, K) - 3S_c(1,); \quad B = 2S_c(1, K) - S_c(K, K^2)$$

$$C = S_c(K, K) + S_c(1, 1); \quad D = S_c(1, K) - S_c(K, K^2)$$

The RHS in inequality 108 is either zero or a small positive constant for $f = 1$ and zero or negative for $f > 1$. Hence, the function G_{c_R} is concave for all practical purposes in $p_x \in (0, 1]$. ■

Property 2 *The higher the degree of heterogeneity in the network, the higher is the potential for cooperation. Thus, with increasing number of smart nodes, the cooperation gain increases initially. But beyond a certain fraction of smart nodes, the network tends towards a homogenous smart antenna network, thereby resulting in a decrease of cooperation gain.*



c) f=1 d) f=3

Figure 49: Cooperation Gain Results

This can be seen from the cooperation gain results for random networks presented in Figure 49 as a function of increasing fraction of smart nodes for $K = 6$. For arbitrary* networks, the gain from cooperation increases initially where some of the omni nodes acting as forwarders of traffic in the backbone can still try to exploit nearby smart nodes for cooperation. However, the probability of finding a nearby smart node and hence the gain is small since the available smart nodes are made to act as forwarders. Once the fraction of smart nodes is such that every forwarder in the backbone is a smart node, the disadvantage of not being able to exploit cooperation owing to spatially sensitive transmissions becomes significant, resulting in a large drop in gain.

Focussing on random networks, one of the key points to note in the results is that the gain from cooperation in a homogeneous smart antenna network ($p_x = 1$) is much lesser than the gain from a homogeneous omni antenna network ($p_x = 0$). While the absolute performance of homogeneous smart antenna networks is much higher than that of omni networks, the potential for cooperation is much smaller owing to the spatially sensitive transmissions that increases with increasing elements. Thus, while increased antenna gain

helps cooperation from the perspective of a node acting as a smart relay, it inhibits cooperation from the perspective of the node serving as a smart source and the inhibition dominates with increasing number of smart node sources. In order to favor cooperation, the smart node transmitters would have to sacrifice antenna gain to reduce the spatial sensitivity of transmissions. This leads to a fundamental tradeoff between exploiting antenna and cooperation gains, the relative importance of the two gains varying with the fading conditions and fraction of smart nodes available. Since both these network parameters cannot be estimated by a node in a distributed manner, we proceed to identify an adaptive cooperation mechanism that attempts to strike a balance between the two gains in random networks.

10.3 Adaptive Co-operation Mechanism

We begin by considering two versions of cooperation in random networks: (i) basic cooperation mechanism considered thus far (C_{ant}), which favors antenna gain and leverages cooperation only if a relay can be found within the communication pattern of the transmitter; and (ii) modified cooperation mechanism (C_{coop}), where a smart transmitter after experiencing a fading loss, switches to omni-directional mode of transmissions to favor cooperation. The tradeoff between the two gains is pronounced when $p_x \rightarrow 1$, under which conditions we have the following lemma.

Lemma 5 *When* $p_x \rightarrow 1$,

$$\begin{aligned} G_{C_{ant}} &\geq G_{C_{coop}}, \quad f = 1 \\ G_{C_{coop}} &\geq G_{C_{ant}}, \quad f = F \end{aligned} \tag{109}$$

Proof: Since the baseline non-cooperation mechanism is the same for both the schemes, it is sufficient to show,

$$\begin{aligned} SNR_{C_{ant}} &\leq SNR_{C_{coop}}, \quad f = 1 \\ SNR_{C_{coop}} &\leq SNR_{C_{ant}}, \quad f = F \end{aligned} \tag{110}$$

We outline the proof for $F = 4$. In the limit, $p_x \rightarrow 1$ and $f = 1$, we have

$$\begin{aligned} SNR_{C_{ant}} &= \bar{P_{rx}}[1 + Q_{nc}(K^2) + Q_{nc}^2(K^2) + Q_{nc}^3(K^2)] \\ SNR_{C_{coop}} &= \bar{P_{rx}}[1 + Q_{nc}(K^2) + Q_{nc}(K^2)Q_{nc}(1)\{1 + \\ & \quad \bar{P_{ro}}Q_{nc}(1) + P_{ro}Q_c(K, K^2)\}] \end{aligned}$$

$SNR_{C_{ant}} \leq SNR_{C_{coop}}$ now equivalently reduces to showing,

$$Q_{nc}(K^2)[1 + Q_{nc}(K^2)] \leq Q_{nc}(1)[1 + \bar{P_{ro}}Q_{nc}(1) + P_{ro}Q_c(K, K^2)]$$

which is true since the $Q_{nc}(K^2)$ term as a scaling factor on the LHS is a much more dominant factor than the $Q_c(K, K^2)$ term, which occurs as an additive factor scaled by P_{ro} on the RHS.

Similarly, for $f = F$, we have,

$$\begin{aligned} SNR_{C_{ant}} &= \bar{P_{rx}}[1 + Q_{nc}(K^2) + Q_{nc}(K^2) + Q_{nc}(K^2)] \\ SNR_{C_{coop}} &= \bar{P_{rx}}[1 + Q_{nc}(K^2) + Q_{nc}(K^2) \\ &+ Q_{nc}(K^2)\{\bar{P_{ro}} + P_{ro}Q_c(K, K^2)\}] \end{aligned}$$

Proving $SNR_{C_{coop}} \leq SNR_{C_{ant}}$, now equivalently reduces to showing,

$$Q_{nc}(K^2)[\bar{P_{ro}} + P_{ro}Q_c(K, K^2)] \leq Q_{nc}(K^2) \tag{111}$$

which is obvious. ■

Thus, we see that exploiting antenna gain and hence scheme C_{ant} is the best strategy under fast fading ($f = 1$) conditions. However, C_{coop} serves to be the best strategy under conditions of time-correlated fading ($f \geq 2$)) by virtue of exploiting cooperation. We shall also observe this subsequently in the results. Hence, what we ideally need is a strategy that can adapt between these two schemes based on the fading conditions. However, since this information is a dynamic network parameter that cannot be estimated, we cannot have a strategy that switches between the two schemes. Hence, we will have to devise a strategy that serves as a middle-ground between the two schemes but will deliver reasonably good performance gains irrespective of the nature of fading conditions. In this regard, we propose the following adaptive cooperation mechanism, which operates as follows.

10.3.1 Mechanism

The proposed adaptive cooperation mechanism is referred to as C_{adap}. The operation of the omni nodes is the same as in the other cooperative schemes. However, a smart node (transmitter) on experiencing a loss reduces the number of elements being used from K to *three*. This reduces the degree of spatial sensitivity and enables cooperation. If it still fails, then it switches to omni-directional transmission to enable cooperation. If $K \leq 3$ in the first place, then the switch is directly made to the omni mode on experiencing a fading loss. The intermediate stage of switching to a lower number of elements, specifically three, provides the following guarantee with respect to an omni-directional transmission.

Proposition 1 *If q_o and q_3 are probabilities of finding a relay in the case of an omni transmission and a beamformed tranmission made with three elements respectively, then* $\frac{1}{3} \leq \frac{q_3}{q_o} \leq 1$.

Proof: Consider a network on a square grid of side R. We assume a circular communication pattern for an omni-directional transmission.

The best case is achieved when the source and destination are separated by a distance equal to the maximum communication range r, where the probability of finding a relay node in the intersection area of the communication ranges of the (omni) transmitter and receiver is given by,

$$q_o = \frac{nr^2}{R^2}(\frac{2\pi}{3} - \frac{\sqrt{3}}{2}) \tag{112}$$

Using three elements for transmission is sufficient to cover more than the intersection area covered in the case of the omni-directional transmission, thereby resulting in $q_3 = q_o$.

However, when the distance between the source and destination starts decreasing and tends towards zero, the intersection area of the (omni) transmitter and receiver starts to increase, thereby increasing q_o. However, since the coverage of the beamformed transmitter is fixed, q_3 starts to decrease. In the limiting case of the transmitter and receiver being co-located, we have

$$q_3 = \frac{nr^2}{R^2}\frac{2\pi}{3} = \frac{1}{3}q_o \tag{113}$$

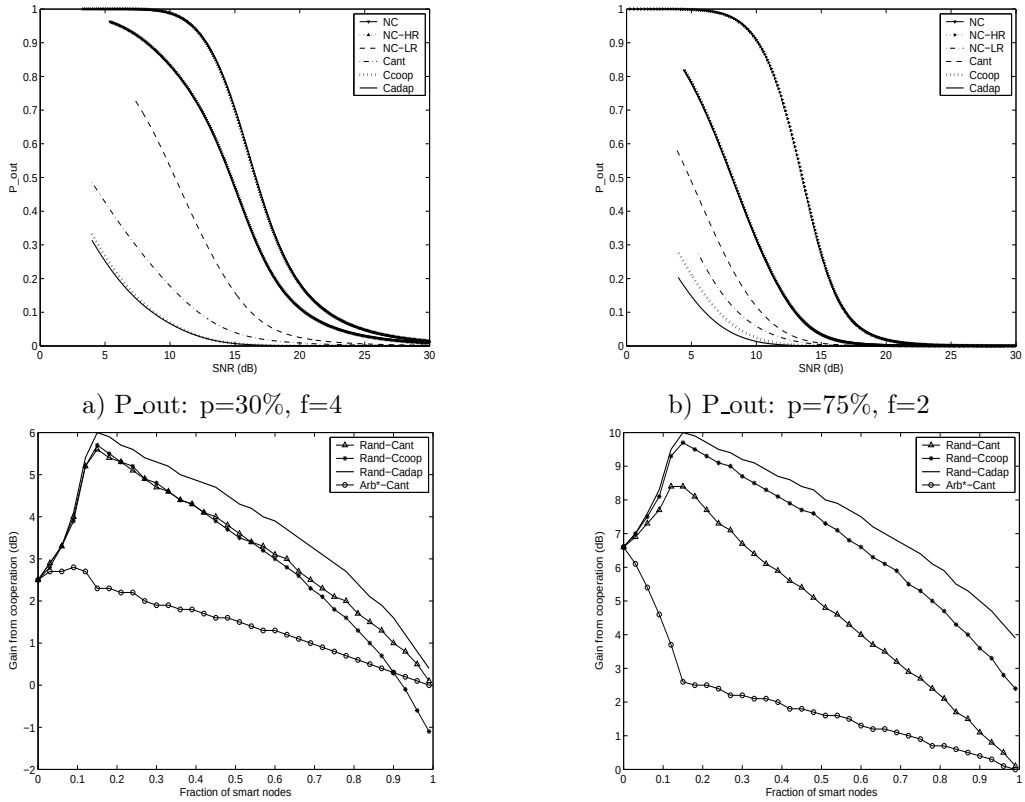
■

Note that all the above cooperation mechanisms considered can be easily extended to the case when the smart nodes are initially using their antenna gain for increased rate instead of reliability. In such cases, on experiencing a fading loss, the transmitter would first switch back to omni rate and use the available antenna gain for reliability. Thereafter, the procedure outlined in the cooperative mechanisms can be directly applied. Since most applications would tend to use the available antenna gain for increased rate, we present numerical results evaluating the different cooperative schemes under this scenario. In the comparison, we also consider two basic rate based mechanisms to counteract fading, namely:

Non-cooperative High Rate (NC-HR): Here, the antenna gain, exploited for increased data rate on a smart link, is retained even after experiencing a fading loss. While this is not a wise idea in the presence of (time) correlated fading, it still minimizes the amount of SNR transmitted in every trial due to lower transmission time.

Non-cooperative Low Rate (NC-LR): Here, on experiencing a fading loss, the transmitter reduces its transmission rate to a low value which helps improve BER performance. Any available antenna gain on the link contributes to reliability as well. However, this increases the average SNR consumed per transmission and also the delay (which impacts throughput directly).

The results are presented in Figure 50 with the smart nodes employing four elements each. From the outage results in Figures 50(a) and (b) it can be clearly seen that NC-HR performs the worst in the presence of fading since it continues to use up its available antenna gain for increased rate and hence has no protection to fading loses. Though the reduced transmission time can contribute to a potential throughput increase, this is outweighed by the large outage error to result in its worst performance. NC-LR uses a rate reduction factor of two on experiencing a fading loss in our experiments. It performs better than the NC scheme but worse than the simple cooperative (C) scheme for small fractions of special nodes. However, for large fractions of special nodes, it performs slightly better than the C strategy owing to the low rate coupled with the antenna gain on most of the links (being smart). However, the increase in delay by a factor of two and hence the reduction in throughput will make the simple cooperative scheme outperforms the NC-LR scheme.



a) P_out: p=30%, f=4

b) P_out: p=75%, f=2

c) Gain: f=1

d) Gain: f=3

Figure 50: Performance of Different Strategies

Thus, the results clearly indicate that rate adaptation mechanisms cannot provide gains as significant as those resulting from cooperation.

It can be clearly seen from the cooperation gain results in Figures 50(c) and (d) that while C_{ant} and C_{coop} perform the best in fast fading and time-correlated fading respectively, the adaptive cooperation mechanism strikes a balance in exploiting the antenna and cooperation gains to provide the best performance under all fading conditions. Further, the degradation in cooperation gain for a homogeneous smart antenna network ($p_x = 1$) is much more graceful in the adaptive cooperation mechanism (C_{adap}) than in the other cooperative schemes.

We now proceed to propose a simple MAC protocol called *MACH* (MAC for HSAN's) that incoporates the adaptive cooperation mechanism.

10.4 MACH Protocol

10.4.1 Mechanism Recap

In our strategy, passive (listening) neighbors of an omni-directional transmitter (S) exploit the wireless broadcast advantage to receive the packet being transmitted to a receiver (D). If the transmission from S fails due to fading, and if there is a neighbor (R) that has successfully received the packet and is within transmission range to D, then R will cooperate to retransmit the packet to D. If there are multiple such neighbors, the one with the largest link (antenna) gain (smart node as opposed to an omni node) will take part in the cooperation. If there are no such neighbors, the source continues to retransmit the packet until the maximum number of trials possible. Further, if the transmitter is exploiting any available antenna gain on the link for increased rate, then it reduces its rate to omni rate and starts exploiting the antenna gain for reliability from the second trial onwards. If the transmitter is a smart node (directional or adaptive), then the spatially sensitive (optimized beam patterns) transmission will prevent any available neighbors from cooperating. In this case, the transmitter first switches from using its antenna gain for rate to reliability using all available elements (if not doing so already). If this transmission also fails, then it switches to using three elements for reliability (if $K > 3$) before eventually resulting in omni-directional transmission to enable cooperation. Using the intermediate stage of operating on three elements would help retain some antenna gain while also ensuring that a neighbor within the range of both S and D receives the packet. Once a neighbor takes part in cooperation it assumes responsibility for the packet instead of S. In any case, the total number of trials for the packet (including those made by the cooperating node) is limited by the specification of MAC protocol (F).

10.4.2 Design Challenges

There arise several challenges in realizing the adaptive cooperation mechanism outlined above in a purely distributed manner.

(i) First, in order for the transmitter S to be able to change strategies and enable cooperation, it needs to be able to distinguish a contention packet loss from a wireless fading

packet loss. (ii) If there is a fading loss during a transmission, it becomes necessary for neighbors to identify this to take part in cooperation even if S has appropriately changed strategies to enable cooperation. Further, they must also identify if they are within communication range of D. (iii) Assuming, that the neighbors have a means to figure out if they can aid in cooperation, one still needs to have a distributed channel access mechanism that will ensure that only one of the neighbors assumes responsibility for retransmission and it is the one that can contribute maximum gain towards retransmission. (iv) Also, S needs to make sure that if no feasible cooperating neighbor is available then it continues to assume responsibility for the packet. (v) Finally, we also need to have an efficient channel access mechanism that can maximize utilization by exploiting the available smart nodes in the network well.

10.4.3 Channel Access Overview

Before deciding on the MAC mechanism, one needs to decide on the fairness model desired by the protocol operation. While max-min fairness has been shown to be unsuitable for ad-hoc networks from an efficiency perspective [43], proportional fairness proves to be a good compromise between efficiency and fairness and a suitable candidate for use in ad-hoc networks [37, 43]. Hence, the fairness model that we adopt for our MAC protocol is proportional fairness[2]. Further, since some links in our network contribute more to the channel capacity than others, we adopt a weighted proportional fairness model to maximize the aggregate network capacity where the smart links access the channel with a higher probability than the non-smart ones. Note that, trying to provide equal bandwidth fairness to the different links irrespective of their contribution to the network capacity, will not help us exploit the capabilities of the available smart nodes efficiently. If the smart links were to exploit the available antenna gain for reliability or transmit power reduction then this would not affect the packet transmission duration which would remain to be the same for all the links. However, since the smart links use the gain for rate increase, this reduces the channel usage time for a packet (fixed size) transmission by a factor that is proportional to

[2]One can also adopt other fairness models and our mechanisms are not specific to the fairness model.

the antenna gain. Hence, the higher priority (frequency) of the smart links in accessing the channel is needed to compensate for their reduced duration of usage of channel bandwidth, consequently resulting in a notion of fairness with respect to the channel access time. This is similar to the approach followed in [49] where channel utilization was maximized by allowing links with good channels to operate at higher rate while at the same time ensuring that all the links obtained a fair share of channel access time by making the good links transmit multiple back-back packets at a high rate.

10.4.4 Contention Mechanism

We use the distributed persistence algorithm for omni-directional antenna networks in [37] as the basic contention resolution mechanism in our MAC protocol. It has been shown in [37] that if the persistence adaptation of each flow follows,

$$\dot{a_i} = \alpha w_i - \beta p_i a_i \tag{114}$$

then the system will converge to the optimal point of maximizing aggregate network utilization while ensuring a proportional fairness model. α and β are system parameters corresponding to a utility constant and a penalty constant respectively. p_i is the loss probability experienced by the flow i; w_i is the weight assigned to the flow, which in our case is proportional to its capacity gain (SNR gain for directional and adaptive nodes) contribution, and a_i is the persistence probability with which the flow decides whether to contend for a slot or not with $\dot{a_i}$ representing the rate of change of persistence. The state maintained at each node consists of the persistence probability, the system constants α and β and the flow's (link's) weight w_i. There are four possible states in which a node can be, namely *Contend*, *No_Contend*, *Acquire*, and *Co_Sched*. Every node i having a packet to transmit, first decides to contend for the channel with a probability of a_i. If the node succeeds, it moves from the *No_Contend* state to the *Contend* state, where it chooses a waiting time uniformly distributed from the interval $(0,B_{max})$ where B_{max} is a constant. The node then waits for the backoff period (in mini-slots), after which it tries to access the channel to see if the channel is busy. If the node finds the channel to be busy (a transmission has already begun), it gives up the slot and decrements its persistence by β. Similarly, if the

channel is idle but if the node detects a collision, it decrements its persistence by β. On the other hand, if the node finds the channel to be idle, and does not experience any collision then it moves to the *Acquire* state where it transmits. At the end of the slot, all the nodes having a packet to transmit in the next slot increase their persistence by αw_i. Thus, the links that contribute more gain (eg. A-A, D-D links) will increment their persistence values more aggressively and hence contend with higher persistence to provide a larger gain, while maintaining a fair channel access time. The values for α and β are empirically set to 0.07 and 0.3 in our simulations.

Every node maintains persistence values and weights for all the links on which it is a transmitter. The weights used in the persistence mechanism must be chosen appropriately. While the weights are based on the link gains ($w_i = \{1, K, K^2\}$) thereby giving smart links higher priority in accessing the channel, we also need to ensure that the cooperating links obtain higher priority in accessing the channel than the source links. Further, among the cooperating links, the one with the higher link gain must have the highest priority. This is achieved by associating two constants with the weight of a link. When a link operates as a normal source (transmitting) link, it uses a weight of $c_1 w_i$ while it uses a weight of $c_2 w_i$ when it operates as a cooperating link. We have the following proposition,

Proposition: If $c_2 > c_1$, the desired channel access priorities will be achieved.

Proof: The normal weighting policy based on the link gains will achieve the desired priorities for normal operation. However, during cooperation, we would require any cooperating link to have higher channel access than the source link of the same type. This is achieved by ensuring $c_2 > c_1$. Further, since all candidate cooperating links use their link gains as weights with the same constant c_2, the cooperating link with the largest gain will have the highest priority. ■

10.4.5 Protocol Details

The MACH protocol uses persistence for channel access and follows a four way handshake (RTS-CTS-DATA-ACK) as in the CSMA/CA protocol. Further, it does not require any synchronization between the nodes. We now explain the sequence of operations at the

source (S), destination (D) and the cooperating (R) node.

Fading Loss Detection: The source, destination and relay (cooperating) node operations are outlined in Figures 51 and 52 respectively. The source S appends to the DATA packet a short preamble that is transmitted at a very low rate compared to the actual DATA transmission. This is similar to the short preamble in IEEE 802.11 which is transmitted at a default low rate of 1 Mbps while the DATA packet itself can be transmitted at 11 Mbps, etc so that the preamble has a higher probability of being decoded since it contains valuable information for channel access (non-intended receivers need not receive the entire packet). A similar strategy is used here although for a different purpose (for identifying fading loss). Hence, if the destination D is able to decode the short preamble but not the data, then the loss is due to fading and not due to contention. However, if fading is extremely severe resulting in loss of preamble as well, then this would not affect the correctness of operation of the protocol, but would make the protocol fall back to default operation, thereby not exploiting cooperation. However, the rate of the preamble is kept low enough to help identify a large fraction of fading loses while at the same time the preamble size is kept small enough (as required for minimal detection and decoding) to avoid excessive overhead.

Source and Destination Operations: RTS and DATA are transmitted using the type (smart/omni) of the source node unless S decides to favor cooperation wherein they are transmitted omni-directionally. However, CTS and ACK are transmitted following the nature of the transmission made by S and not according to D's type. This is because, if S is an omni node, then its transmission would automatically favor cooperation. The reason CTS needs to be omni is because, for a neighboring node to identify if it is within communication range of D, it needs to be able to receive the CTS from D. Hence, if a neighboring node receives both RTS and CTS, then it can assume itself to be a relay. Further, to decide if the relay should cooperate in retransmitting the packet, it needs to identify if the packet was lost in the first place due to fading. This is made possible by D transmitting ACK in the same mode as S's DATA transmission along with the loss information (ACK is sent even if the DATA was not decoded due to fading loss but the loss was detected). If S had transmitted DATA in directional/adaptive mode, then R would have anyway not received

Source and Destination Operation

1 S: Tx RTS based on S's type (O/X); D Rx based on D's type
2 D: Tx CTS using O(X) if RTS Rx using O(X) resply, S Rx using S's type
3 S: Tx DATA using S's type, available antenna gain used for rate/reliability
trials++; small preamble Tx along at low rate, D Rx using D's type
4 D: Detects fading loss if DATA not decode-able but preamble decode-able
Tx ACK on O(X) if DATA Rx on O(X), sends loss info, S Rx using S's type
5 S: If (trials == F) Drop Packet
S: Else Re-contend based on persistence using c_1
6 S: If (Channel_access())
/* If not fading loss, no change in strategy, Goto Step 1 */
7 If (Strategy == rate)
8 Strategy = reliability; first on K elements, then on 3 elements; goto Step 1
9 S: Tx RTS using O; D Rx based on D's type
10 D: Tx CTS using O, S Rx using S's type
11 S: Tx DATA using O, available antenna gain used for reliability; trails++
12 small preamble Tx along at low rate, D Rx using D's type
13 D: Detects fading loss, Tx ACK using O with loss info, S Rx using S's type
14 S: Re-contends based on persistence using c_1
15 S: If (Channel_access()) Goto Step 1, use antenna gain for reliability
16 S: ElseIf (Recv_CTS(R)) Drop pkt /* CTS for its pkt destined to R */
17 S: Else Re-contend for next slot, Goto Step 14
18 S: ElseIf (Recv_CTS(R)) Drop pkt /* CTS for its pkt destined to R */
19 S: Else Re-contend for next slot, Goto Step 5

Figure 51: Source/Destination Operation

the DATA packet with a high probability due to the spatially sensitive transmission by S. Hence, S would have to retransmit the DATA omni-directionally before R were to cooperate. Hence, it is sufficient for D to send back ACK omni-directionally only if the DATA was sent omni-directionally in the first place. Thus, a neighboring node can both identify if it is a relay and if it should cooperate. On experiencing a fading loss, S switches strategy from exploiting available antenna gain for increased rate to reliability (if not already using gain for reliability) first on K elements, then on 3 elements and eventually to omni-directional transmission. However, if it still experiences a fading loss in omni-directional mode with the number of trials not exhausted, then it switches back to exploiting all the available antenna gain (K elements) for reliability. In most practical cases, the maximum number of trials is three or four for the DATA packet.

Relay Operation
1 R: If (Recv_RTS(D) & Recv_CTS(S)) Assume Relay
2 R: If (Recv_DATA(D)) Store DATA in HOL in MAC queue
3 R: If (!Recv_ACK(S) || (Recv_ACK(S) & !Fading)) Drop stored pkt
4 R: Else Contend using persistence using c_2
R: Obtain number of trials remaining (F') for pkt
5 R: If (Channel_access())
6 R: Tx RTS based on R's type (O/X); D Rx based on D's type
7 D: Tx CTS using O, R Rx using R's type
8 R: Tx DATA using R's type, available antenna gain used for reliability
R: trials++, D Rx using D's type
9 D: Tx ACK using D's type if (Recv_DATA(D)) successful, S Rx using S's type
10 R: If (Recv_ACK(R)) drop DATA
11 R: ElseIf (trials < F') Re-contend with persistence using c_1, Goto Step 5
12 R: Else Drop pkt
13 R: ElseIf (Recv_CTS(R') || Recv_CTS(S)) Drop pkt
14 R: Else Re-contend with persistence using c_2, Goto Step 5

Figure 52: Relay Operation

Relay Operations: Once a neighbor is aware that it is a relay, it stores the DATA packet (if received successfully) until it receives the ACK. If ACK is not received or if it does not indicate fading then the packet is dropped. Otherwise, the relay decides to cooperate in retransmission. It then stores the DATA packet at the head of line of its MAC queue to give it the same priority as in the source node. It then contends for the packet with persistence but using c_2 constant in its weight unlike the source that uses c_1. The appropriate choice of persistence values helps the relay with the largest link gain obtain channel access with a higher probability. The relay uses its link gain only for reliability and does not switch strategy for future trials. RTS, DATA and ACK are transmitted depending on the type of S and D. However, when D responds back to R with a CTS, the CTS is always made omni so that S as well as other contending relays (R') are made aware of R assuming responsibility for the packet for the remaining trials (F') for the packet. S and other relays R', then drop the packet from their queue. After the first of the remaining F' trials for the packet, R falls back to using c_1 in its link weight for the remaining trials (if needed) since it becomes the virtual source for that packet.

Thus, the operations at S, D and R address all the challenges raised earlier in exploiting the cooperative mechanism efficiently. More importantly, the cooperative mechanism is made possible without practically requiring any separate (control) communication amongst S, D and R. To put in simple words, instead of S retransmitting the data packet during fading losses, a neighboring relay with a larger link gain is exploited for the purpose without incurring any additional overhead (by exploiting the existing control packets in the handshake mechanism efficiently). The total number of trials is maintained to be the same as in a normal MAC protocol and hence does not affect its latency characteristics as well.

Note that, the cooperative mechanism is not specific to a persistence-based channel access mechanism and can also be implemented in a backoff domain such as the IEEE 802.11 MAC protocol. However, the priorities of channel access for the different links must be appropriately built into the backoff mechanism by weighting either constant waiting periods such *difs*, or contention window, etc. Also, the mechanism is not specific to a smart source node exploiting its antenna gain for rate, but can also be used with other forms of gain exploitation such as power reduction etc. More importantly, MACH is a simple MAC protocol for the purpose of incorporating the adaptive cooperation mechanism. Hence, it can be used in conjunction with any of the existing or newly proposed MAC protocols for multi-hop wireless networks with homogeneous or heterogeneous smart antenna nodes.

10.5 Performance Evaluation of MACH

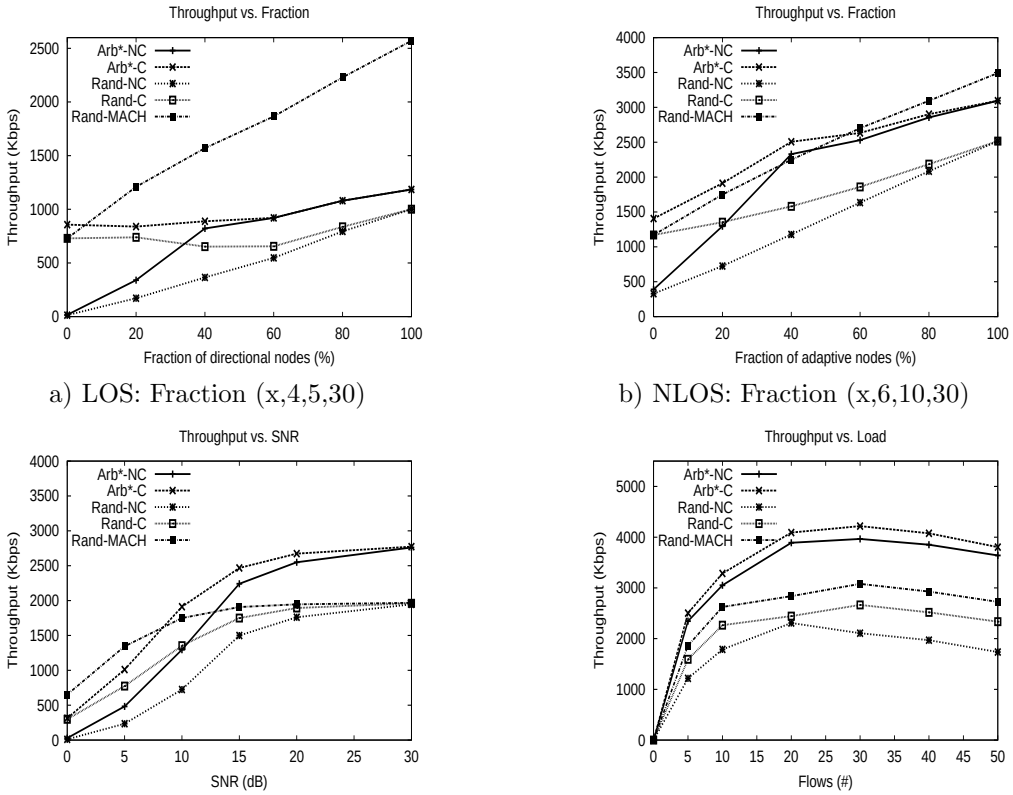
We evaluate the performance of the proposed MAC protocol in a variety of network settings. We use *ns2* network simulator for all our simulations. The topologies considered are static consisting of 100 nodes in a 1000m by 1000m grid. The transmission range used is 100m. The parameters used in the study are (i) fraction of smart nodes, (ii) transmitted SNR, and (iii) load (# flows). The number of flows is maintained at 30 unless otherwise specified. Aggregate Throughput, measuring the number of packets successfully received by the destinations of all multi-hop flows in Kbits/sec, is the primary metric of interest. Gain in throughput resulting from collaboration is also explicitly profiled. We also use normalized standard deviation (standard deviation normalized to mean) of the flows' throughputs to

observe the fairness properties of MACH. The sources and destinations are chosen at random. UDP is used as the transport protocol. Every source generates traffic on a channel of 2 Mbps at a high rate (20 pkts/s with 1 KB packet size) to remain back-logged for the entire simulation duration. We consider the basic non-cooperative (NC) and cooperative (C_{ant} or simply C) schemes in random and arbitrary* networks as baselines for comparison with MACH in random networks. The simulations are run for 100 secs and each of the data point in the graphs presented is averaged over 10 seeds.

We consider two environments of study: LOS and NLOS environments. In NLOS environments, we consider a combination of adaptive and omni nodes while in LOS we consider a combination of directional and omni nodes. We use the two ray propagation physical model available in *ns2* for the LOS environment. However, for the NLOS environment where we consider multipath scattering, such a simple model would not suffice. To emulate the link characteristics present in multipath environments, we run Matlab simulations for the scenarios considered with different types of smart antenna processing at the nodes. Thus, we obtain the link-level packet loss statistics based on the different SINR's for different (i) configurations (positions) of the (desired) transmitter-receiver pair and interferers, (ii) receiver processings (LMMSE, SIC, etc.) used, and (iii) number of elements used in the MEA. We then incorporate these statistics into the physical layer model of *ns2* for our simulations. We consider Rayleigh fading in our channel model. Every node uses a constant transmit power which is determined by the transmit SNR budget. The SINR threshold on every link is set to 5 dB. The number of trials for packet retransmission are limited to four. The duration of fade (in packets) is assumed to be a random variable even within a single simulation run, taking in values $\{1, 2, 3, 4\}$ with decreasing probability in that order.

10.5.1 Throughput

The throughput results are presented in Figure 53. Figure 53(a) presents the throughput results for LOS environments with four elements and a transmit SNR of 5 dB as a function of the number of smart nodes in the network. It can be seen that the simple cooperative scheme is able to provide better gains in random networks than in arbitrary* networks, thereby



a) LOS: Fraction (x,4,5,30)

b) NLOS: Fraction (x,6,10,30)

c) NLOS: SNR (20,6,x,30)

d) LOS: Load (40,4,15,x)

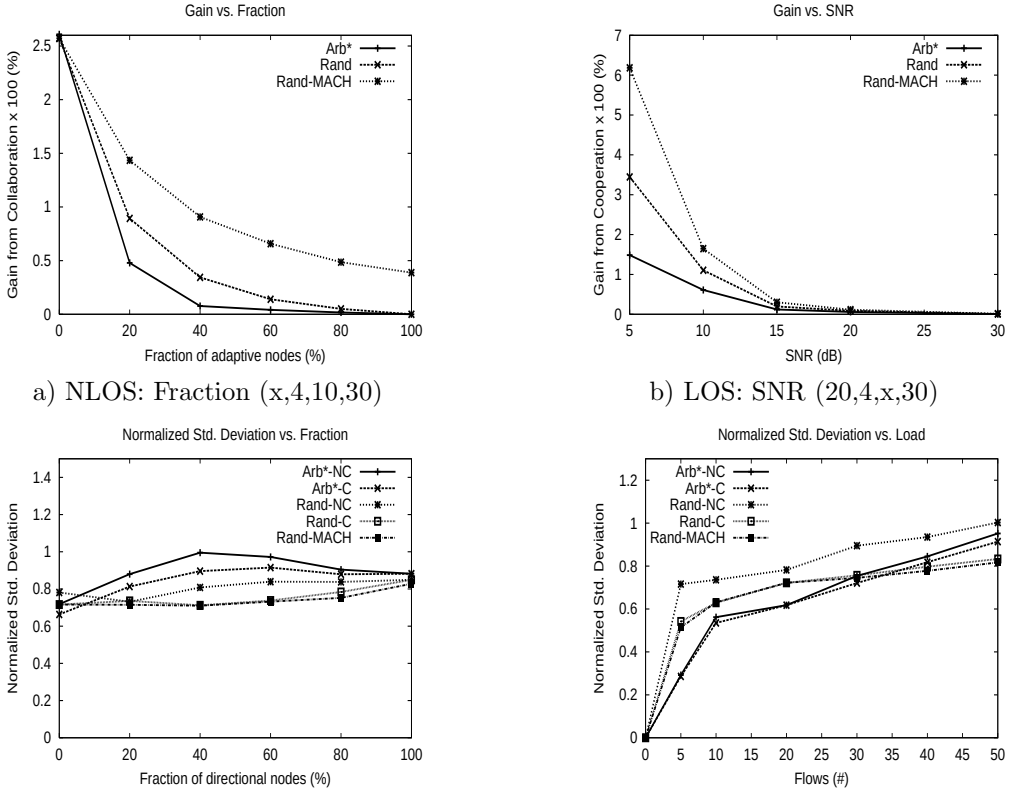
Figure 53: Throughput Results: (Fraction,Elements,T_SNR,Flows)

bridging the performance gap between the two networks without requiring for smart node placement. Further, MACH is able to provide significant gains over the simple cooperative scheme upto about 100%, especially when the fraction of directional nodes is large due to its ability to exploit cooperation efficiently and hence scale well in the presence of increasing fraction of smart nodes. Similar trends can be observed in Figure 53(b) which presents the throughput result for NLOS environments with six elements and a transmit SNR of 10 dB.

Figure 53(c) presents the throughput results measured as a function of increasing transmit SNR for a smart node fraction of 20% with six elements each in NLOS environments. It can be seen that the gain from cooperation is maximum at low SNR's where the impact of fading is maximum and then starts to decrease with increasing SNR's. MACH is able to provide a two fold performance improvement over the naive cooperative scheme.

Figure 53(d) presents throughput results as a function of load - increasing flows in the network for LOS environments with 40% smart nodes with six elements each and a transmit

SNR of 15 dB. The throughput increases with increasing flows initially and then starts to decrease at high loads due to increased contention in the network. Since (i) the transmit SNR is moderate, (ii) fraction of smart nodes is almost half and (iii) the number of antenna elements is also large, most of the links in the network have a high link gain and sufficient protection against fading losses thereby reducing the potential for cooperation. However, even in this case MACH provides a gain of about 50% over the no cooperation scheme and about 25% over the simple cooperation scheme.



a) NLOS: Fraction (x,4,10,30)

b) LOS: SNR (20,4,x,30)

c) LOS: Fraction (x,4,10,30)

d) NLOS: Load (40,4,10,x)

Figure 54: Gain and Fairness Results: (Fraction,Elements,T_SNR,Flows)

10.5.2 Cooperation Gain

Figures 54(a) and (b) present results profiling the gain in throughput obtained from cooperation. This is measured by the difference in throughput between the cooperative and non-cooperative schemes, as a fraction of the throughput of the non-cooperative scheme. Figure 54(a) presents the gain as a function of the number of smart nodes in the network for

NLOS environments, with a transmit SNR of 10 dB, and four elements at each of the smart nodes. It can be seen that the gain from cooperation decreases with increasing fraction of smart nodes for all the schemes in both random and arbitrary* networks. While it is understandable for the simple cooperative schemes which are unable to exploit cooperation with the smart node transmitters, even MACH suffers a decrease in gain. This is because, while MACH helps even the smart nodes to exploit cooperation, it does so at the cost of reducing the number of antenna elements being used for beamforming, thereby losing out on the antenna gain. Hence, as the fraction of smart nodes increases, the price to pay for exploiting cooperation increases, thereby decreasing the net gain from cooperation. However, the gain from cooperation in MACH is still about 40% in the worst case and several folds in the best case (lower fractions).

Figure 54(b) presents the gain as a function of transmit SNR for LOS environments with 20% smart nodes having four elements each. It can be seen that as transmit SNR increases, the gain from cooperation decreases rapidly since the received SNR would also increase and thereby provide greater protection against fading losses. The gain from cooperation is large at small SNR, which makes cooperation a highly useful mechanism in energy constrained applications such as sensor networks. The gain from simple cooperation can be as high as a factor of two while the gain from MACH over simple cooperation can be as high as a factor of three.

10.5.3 Fairness

While MACH does provide significant gains in throughput, we need to ensure that the improvement comes purely from cooperation and does not come from aggressive channel access by the smart links over the omni links than that existing in non-cooperation. Hence, we consider deviation in the distribution of throughputs obtained by MACH and simple cooperation schemes with respect to that of the non-cooperation scheme. Standard deviation normalized to the mean is used for the purpose. The normalized standard deviation (NSTD) results are presented as a function of the fraction of smart nodes and load in Figures 54(c) and (d) for the LOS and NLOS environments.

Increasing the fraction of smart nodes does not impact the NSTD significantly, indicating that the smart nodes are not exploited more aggressively than in non-cooperation scheme (Figure 54(c)). In fact, the measure is lower for the cooperative schemes than the non-cooperative scheme by about 10% in both random and arbitrary* networks. The reason for the decrease in NSTD is due to the fact that fading losses skew up the throughput distribution from the desired one (in the absence of fading) in non-cooperative schemes which is alleviated in the cooperative schemes. Further, the NSTD measures for MACH are similar to that of the simple cooperation scheme, indicating that the gain of MACH over simple cooperation is also purely due to leveraging cooperation efficiently for the smart nodes. Increasing the load increases contention in the network causing the unfairness index to increase with load (Figure 54(d)). However, the cooperative schemes still provide a lower NSTD than the non-cooperative ones by about 25%.

CHAPTER XI

CONCLUSIONS AND OPEN PROBLEMS

The focus of this thesis has thus been to investigate the use of smart antennas in ad-hoc networks and hence efficiently design network protocols that best leverage their capabilities in communication. This has been achieved by dividing the problem domain into two components of designing efficient network protocols for homogeneous and heterogeneous smart antenna networks.

In homogeneous smart antenna networks, we first considered the most sophisticated of the smart antenna technologies, namely multiple-input multiple-output (MIMO) links and designed a fair stream-controlled medium access control (SCMA) protocol called *SCMA* and a MIMO routing protocol called *MIR*. These protocols exploited the benefits of MIMO links to the best possible extent at the MAC and routing layers respectively. Having identified the optimization considerations with respect to MIMO links helped us understand the relative benefits of other antenna technologies as well and gave us the potential to look at smart antennas as a whole. In this direction, we designed 'unified" frameworks that formulate and solve the MAC and routing problems for smart antennas in general. The unified frameworks were used to derive MAC and routing protocols for different kinds of smart antenna technologies like switched beam, adaptive arrays and MIMO links. The frameworks were also used to evaluate the different smart antenna technologies and their corresponding strategies in different network conditions, thereby helping us identify design rules indicating the optimal smart antenna technology and strategy to be used in specific network conditions depending on application requirements.

Building on the strength of the insights and inferences gained from protocol design in homogeneous environments, we then moved on to consider ad-hoc networks with nodes having heterogeneous antenna technologies. We first studied the impact of increasing degree of antenna heterogeneity on the throughput capacity of the network through analytic tools and

identified several valuable inferences. We also designed efficient MAC and routing protocols that help exploit the heterogeneous link capabilities available in the network. Finally, we designed simple yet effective node-cooperation strategies that helped the communication protocols further exploit the heterogeneous capabilities to the maximum extent.

While the thesis identifies and addresses several problems pertaining to the exploitation of smart antenna capabilities in ad-hoc networks and the consequent need for efficient network protocol design, there still exist several open problems that are of practical relevance and hence need attention.

- *Exploitation of interference suppression capabilities:* The thesis has predominantly focused on exploiting the link gains of the different smart antenna technologies, except in the design of the stream-controlled MAC protocol. Interference suppression capabilities of smart antennas have significant potential to improve the spatial reuse of the network, allowing multiple conversations to occur in tandem though they are in the vicinity. This can help improve throughput as well reduce the latency for several applications. However, before being able to design efficient network algorithms and protocols that can leverage these interference suppression capabilities, we would also need a sound theoretical understanding of the maximum possible benefits from such a capability. We have made some preliminary effort in this direction [63].

- *Energy efficient protocol design:* The smart antenna capabilities have been utilized mainly for the purpose of increased rate, range or reliability and seldom for the purpose of reduced power consumption on the links. Energy efficiency is a critical factor for several applications in wireless ad-hoc and sensor networks. Hence, designing energy-efficient communication protocols, where throughput may not always be "the" most important metric, also forms an important aspect. The design rules identified in this thesis with respect to the operation of different smart antenna technologies and their strategies can in fact be used to design energy-efficient network protocols where the focus will be to optimize *throughput/energy* as opposed to simply *throughput*.

- *Secure network protocol design:* Security has become a vital aspect of all wireless

applications, including both commercial and military, given the shared nature of the wireless medium and consequent potential to maliciously tap the wireless channel. While the usage of coding (for security) and cryptographic techniques can provide a layer of security, they come at the cost of increased complexity, overhead and resource (energy) usage. Smart antennas, which are primarily employed for improved communication capabilities, can also provide a layer of security as a by-product without having to incur the disadvantages of coding and cryptographic techniques, whose sole purpose is to provide security. The ability of smart antennas to couple transmissions onto specific spatial channel modes as well as their interference suppression capabilities are examples of how the data can made more spatially selective/sensitive and jamming can be prevented respectively. However, one would need efficient protocols that can leverage and execute the smart antenna capabilities for the purpose of providing security.

- *Enabling quality of service applications:* Multi-hop wireless networks are becoming popular in the commercial domain with the popularity of mesh network applications. However, for applications to thrive in these networks, one would need to provide quality of service to applications. While providing quality of service (QOS) guarantees has been an especially challenging task in multi-hop wireless networks with omni-directional antennas, the employment of smart antennas can help alleviate this burden. However, the requirements of several applications are no longer restricted to simply throughput, which in turn has been the primary focus of this thesis. For eg., voice-over IP (VOIP) and video streaming are going to be two important applications, whose primary metric of interest are delay (and jitter) and peak signal-to-noise ratio (PSNR) respectively. Hence, it becomes necessary to design efficient network protocols that will help exploit the capabilities of smart antennas in a manner that will optimize QOS metrics specific to different applications.

APPENDIX A

TIME COMPLEXITY OF SCMA

Theorem 1 *If the network is represented as a graph, $G = (V, E)$, where V is the set of nodes and E is the set communication links, then the worst-case time complexity of the link classification component of SCMA can be bounded by $O(E \log E)$, while the worst-case time complexity of the schedule-per-slot component can be bounded by $O(E)$.*

Proof: Given the node and flow contention (G') graphs, the number of vertices in the flow contention graph is the number of communication edges in the node graph. The perfect elimination ordering (PEO) of the vertices in G' can be obtained using the LexBFS algorithm in linear time [47], thereby requiring $O(E)$ operations. Given the PEO, all the maximal cliques in G' can also be obtained in linear time using a theorem by Fulkerson and Gross [18], thereby requiring $O(E)$ operations. The ranking and coloring of the vertices in G' can be accomplished in $O(E \log E)$. This is achieved in the following two steps: (i) The links are first partitioned into two sets - bottleneck links, belonging to multiple contention regions along with their rank; and non-bottleneck links, belonging to a single contention region. This incurs $O(E \log E)$ due to the ranking process involved. (ii) In the set of bottleneck links, the lowest ranked one is chosen first. If there are no non-bottleneck links in the set of maximal cliques containing the bottleneck link, that can benefit from spatial reuse, then the bottleneck link is moved to the set of non-bottleneck links. This re-classification is repeated for all the bottleneck links in the order of increasing ranks, resulting in $O(E)$ operations. Thus, the entire process of ranking and classification/coloring requires $O(E \log E)$ operations. This in combination with the $O(E)$ operations required for obtaining the maximal cliques, will continue to incur a worst case time complexity of $O(E \log E)$ for the link-classification component.

For the schedule-per-slot component, the red/white (depending on the scheduling phase)

link with the minimum service is chosen from the entire set of red and white links. This would require $O(E)$ operations. The neighbors of the serviced link(s) are determined and removed from the schedule of the upcoming slot. This would also require $O(E)$ operations. If the link chosen was white, then multiple white links in the same maximal clique will be serviced, resulting in all their neighbors to be removed from the schedule of the upcoming slot. This however incurs the same complexity of $O(E)$. The above steps have to be repeated for the remaining set of potential links that can be scheduled in the same slot. The number of such iterations will atmost be $\#$ *maximal cliques* (c), or can be bounded by the spatial reuse factor, which is a constant for a given node graph (depends on the network size and communication/interference range). Hence, the total time complexity of the schedule-per-slot component is given by $O(c \cdot E) = O(E)$.

■

APPENDIX B

MIR PROOFS

B.1 Proof of Lemma 1

The average lossy hop length $\bar{h}$ over which a packet travels in the case of lossy links with a PER of P is given by,

$$\bar{h} = h + \frac{[1-(1-P)^h][1-(1+Ph)(1-P)^h]}{P(1-P)^h} \tag{115}$$

where h is the average non-lossy hop length or the true average hop length in the underlying topology.

Proof: Let l be the probability of a packet successfully reaching the destination. This can be given as,

$$l = (1-P)^h \tag{116}$$

This assumes that once a packet is lost at a hop i, the resources utilized till hop i are wasted. One may think that the presence of semi-reliable MAC layer or FEC could potentially help the packet in reaching the destination inspite of loss along the path, thereby reducing the wastage of resources. But note that, in multipath environments especially in time-correlated fading that lasts for several tens to hundreds of milliseconds [71], a few retransmissions at the MAC layer or protection for a few bits through FEC will not help. Hence, for all practical purposes, a loss on a link can be assumed to be irrecoverable in terms of the wastage of resources till that hop.

The average hop a length a packet travels before experiencing a loss h_l can be given as,

$$h_l = \sum_{i=1}^{h} (1-P)^{i-1} Pi = \frac{1-(1+Ph)(1-P)^h}{P} \tag{117}$$

Now the average hop length a packet travels before successfully reaching its destination $\bar{h}$ can be given as,

$$\bar{h} = lh + (1-l)\{l(h_l+h) + (1-l)\{l(h_l+h_l+h) + \ldots$$

$$= \sum_{i=0}^{\infty} l(1-l)^i(h+ih_l) = h + h_l\frac{1-l}{l}$$
$$= h + \frac{[1-(1+Ph)(1-P)^h][1-(1-P)^h]}{P(1-P)^h} \quad (118)$$

Hence, based on the loss probability P, it can be concluded that,

$$\bar{h} = h + \frac{[1-(1+Ph)(1-P)^h][1-(1-P)^h]}{P(1-P)^h},\ 0 < P < 1$$
$$= h, \qquad P = 0 \quad (119)$$

■

B.2 Proof of Lemma 2

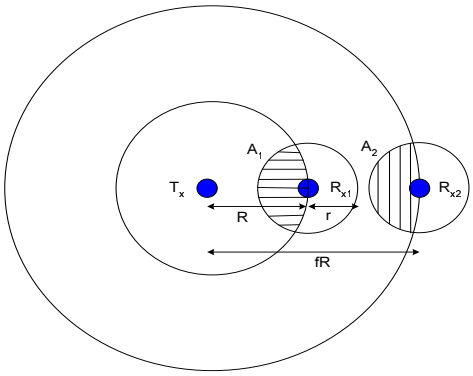


Figure 55: Link failure during mobility

$$P_{l\_db} < P_{l\_dr} < P_{l\_s} \quad (120)$$

where P_{l_db}, P_{l_dr} and P_{l_s} are the probabilities of a link going down due to mobility in DIV-BER, DIV-RANGE and MUX respectively.

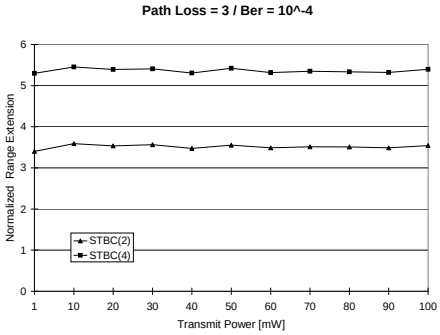
Proof: The increased reliability (BERs lower than the threshold) on the link in DIV-BER is able to sustain extended ranges on the link due to mobility as long as the BER does not fall below the threshold. This extendable range is given by the maximum range extension that is possible with diversity for a given BER threshold (as in DIV-RANGE). In DIV-RANGE, the (already) increased communication range gives the node more area to move about without breaking the link, thereby reducing the probability of the link going down.

To understand better, consider a link between a transmitter T_x and receiver R_x as in Figure 55. Without loss of generality, assume that R_x is at the fringe of the communication range in the link. This would correspond to R_{x1} at a distance R from T_x in MUX and DIV-BER and to R_{x2} at a distance fR in DIV-RANGE, where f is the range extension factor resulting from diversity gain. For simplicity, consider T_x to be stationary, while R_x moves with a speed v. Further, consider a time window of observation T. This time window could correspond to the least time within which the route failure can be prevented proactively by obtaining an alternate route, etc. In this time window, R_x can be equally likely anywhere in the circular area (assuming a random waypoint mobility model) of radius $r = vT$ with R_x's initial position (R_{x1} in MUX and DIV-BER and R_{x2} in DIV-RANGE) being the center. Now, the probability of being able to sustain the link is given by the area of intersection of the circles representing the region of mobility and the region of sustainable communication as a fraction of the region of mobility. The intersection area would correspond to the intersection of the circles with radii r and R in MUX, circles with radii r and fR in DIV-RANGE, and circles with radii r and fR in DIV-BER. The complement of the intersection area would give the probability of the link going down.

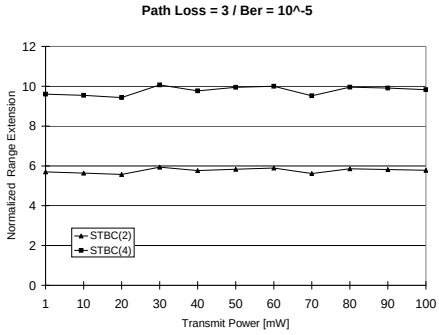
The centers of the region of sustainable communication is the same in all strategies, being the transmitter. However, the center of the region of mobility is the receiver which is at distance R in MUX and DIV-BER and fR in DIV-RANGE as shown in Figure 55. Thus, looking at the two regions for DIV-BER and DIV-RANGE, one can easily see that the areas of the two regions are the same, while the center for the mobility region is much closer to transmitter in DIV-BER than in DIV-RANGE, thereby giving the receiver larger area and hence probability to sustain the link during mobility. Thus, it is straightforward to show that,

$$\begin{aligned} 0 < P_{l\_db} < P_{l\_dr}, \quad & r > (f-1)R \\ P_{l\_db} = 0, \quad & r \leq (f-1)R \end{aligned} \tag{121}$$

In the case of DIV-RANGE and MUX, the receivers are at the fringe of the communication range and the region of mobility is also the same. However, the region of sustainable

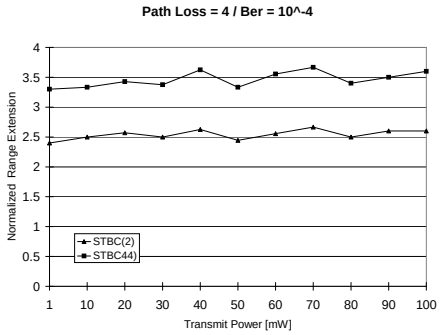


a) path loss = 3, p = 1e-4

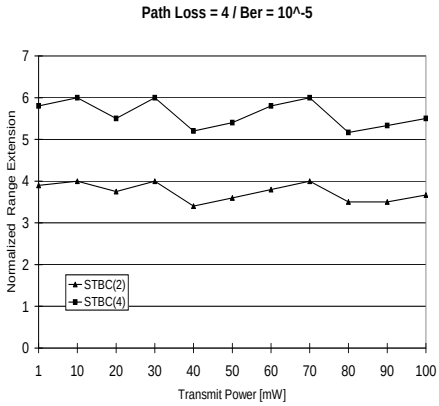


b) path loss = 3, p = 1e-5

Figure 56: Range extension with path loss = 3



a) path loss = 4, p = 1e-4



b) path loss = 4, p = 1e-5

Figure 57: Range extension with path loss = 4

communication is more in DIV-RANGE, thereby resulting in a higher ability and probability to sustain the link during mobility. This can also be verified by calculating the areas of intersection of the two regions in both the cases. Using simple trigonometry, it can be shown that the probability (of link going down) P_{l_s} is given by,

$$\begin{aligned} P_{l\_s} &= 1 - \frac{R^2}{\pi r^2}\cos^{-1}(1 - \frac{r^2}{2R^2}) - \frac{1}{\pi}\cos^{-1}(\frac{r}{2R}) \\ &\quad + \frac{R}{\pi r}\sqrt{1 - \frac{r^2}{4R^2}} \end{aligned} \tag{122}$$

In DIV-RANGE, the corresponding probability P_{l_dr} it is given by,

$$\begin{aligned} P_{l\_dr} &= 1 - \frac{(Rf)^2}{\pi r^2}\cos^{-1}(1 - \frac{r^2}{2(Rf)^2}) - \frac{1}{\pi}\cos^{-1}(\frac{r}{2Rf}) \\ &\quad + \frac{Rf}{\pi r}\sqrt{1 - \frac{r^2}{4(Rf)^2}} \end{aligned} \tag{123}$$

Observing the impact of f in Eqn. 123, one can easily verify that $P_{l\_dr} < P_{l\_s}$. This combined with Eqn. 121 proves the result. ■

B.3 Range Extension Using Diversity

Figures 56 and 57 present the range extension factor (normalized to omni-directional range) as a function of transmit power for path loss components of 3 and 4 respectively. Each of the figures presents results for BER's of 1e-4 and 1e-5. A Rayleigh fading model was used for the simulations in MATLAB. Note that 1e-5 is a lower BER than 1e-4 and would hence correspond to a relatively higher SNR regime. Hence, it can be seen that range extension is more than linear in the high SNR regime, namely factors of 6 and 4 for path loss components of 3 and 4 respectively as we go from one to two elements. Further, the gain decreases with increasing path loss component. Also, the gain is close to linear in the relatively moderate SNR regime corresponding to a BER of 1e-4. However, as the number of elements is increased from two to four the gain decreases. This is due to the diminishing returns provided by the diversity gain. But note that MIR predominantly uses only two elements in all its diversity based mechanisms and hence the assumption of a linear range extension factor (with respect to number of elements) is reasonable. Though we have considered transmit powers only upto 100mW while IEEE 802.11b is known to use 280mW, the trends do not change much with increasing transmit power as can be seen from the results and hence the inferences would still be valid.

REFERENCES

[1] Alamouti, S., "A simple transmit diveristy technique for wireless communications," *IEEE JSAC*, vol. 16, pp. 1451–1458, Oct 1998.

[2] Anderson, J. B., "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE JSAC*, Nov 2000.

[3] Anderson, J., "Antenna arrays in mobile communications: Gain, diversity, and channel capacity," *IEEE Antennas and Propagation Magazine*, vol. 42, pp. 12–16, Apr 2000.

[4] Bao, L. and Garcia-Luna-Aceves, J. J., "Transmission scheduling in ad-hoc networks with directional antennas," in *Proceedings of ACM MOBICOM*, (Atlanta, GA), Sep 2002.

[5] Bliss, D., Forsythe, K. W., Hero, A. O., and Swindlehurst, A. L., "MIMO environmental capacity sensitivity," in *Proceedings of IEEE Asilomar Conference on Signals, Systems, and Computers*, Oct 2000.

[6] Blogh, J. and Hanzo, L., *Third-generation systems and intelligent wireless networking: Smart antennas and adaptive modulation.* IEEE Press, John Wiley and Sons, Ltd.

[7] Catreux, S., Driessen, P., and Greestein, L., "Simulation results for an interference-limited MIMO cellular system," *IEEE Communication Letters*, vol. 4, Nov 2000.

[8] Catruex, S., Greenstein, J., and Erceg, V., "Some results and insights on the performance gains of MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 839–847, Jun 2003.

[9] Choudhury, R. R. and Vaidya, N. H., "Impact of directional antennas on ad-hoc routing," in *8th Conference on Personal and Wireless Communcations*, (Venice), Sept 2003.

[10] Choudhury, R. R., Yang, X., Ramanathan, R., and Vaidya, N. H., "Using directional antennas for medium access control in ad hoc networks," in *Proceedings of ACM MOBICOM*, (Atlanta), 2002.

[11] Cui, S., Goldsmith, A. J., and Bahai, A., "Energy efficiency of MIMO and cooperative MIMO in sensor networks," *IEEE JSAC*, vol. 22, no. 6, 2004.

[12] Demirkol, M. and Ingram, M. A., "Stream Control in Networks with Interfering MIMO Links," *Proc. of IEEE WCNC*, Mar. 2003.

[13] Demirkol, M. F. and Ingram, M. A., "Control Using Capacity Constraints for Interfering MIMO Links," *Proc. of the Int. Symp. on Personal, Indoor and Mobile Radio Communications*, vol. 3, pp. 1032–1036, Sep. 2002.

[14] Doufexi, A., Armour, S. M., Nix, A. R., Karlsson, P., and Bull, D. R., "Range and throughput enhancement of wireless local area networks using smart sectorised antennas," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1437–1443, Sept 2004.

[15] Fahmy, N. S., Todd, T. D., and Kezys, V., "Ad hoc networks with smart antennas using IEEE 802.11-based protocols," in *IEEE ICC*, 2002.

[16] Foschini, G. J., "Layered space-time architecture for wireless communication," *Bell Labs Technical Journal*, vol. 6, pp. 311–335, Mar. 1998.

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

[18] Fulkerson, D. R. and Gross, O. A., "Incidence Matrices and Interval Graphs," *Pac. J. Math*, vol. 15, pp. 835–855, 1965.

[19] Gesbert, D., Shafi, M., Shiu, D., Smith, P. J., and Naguib, A., "From Theory to Practice: An Overview of MIMO Space-Time Coded Wireless Systems," *IEEE JSAC*, vol. 21, pp. 281–301, Apr 2003.

[20] Gupta, P. and Kumar, P. R., "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388–404, Mar 2000.

[21] Hottinen, A., Wichman, R., and Tirkkonen, O., *Multiantenna Transceiver Techniques for 3G and Beyond*. Wiley, John and Sons, Feb. 2003.

[22] http://www.isi.edu/nsnam/ns/, *The network simulator: ns2*. Jun 2001.

[23] Hu, M. and Zhang, J., "MIMO Ad Hoc Networks: Medium Access Control, Satuation Throughput, and Optimal Hop Distance," *Special Issue on Mobile Ad Hoc Networks, Journal of Communications and Networks*, 2004.

[24] Jiang, J.-S., Demirkol, M., and Ingram, M., "Measured capacities at 5.8 GHz of indoor MIMO systems with MIMO interference," in *Proceedings of the IEEE Fall Vehicular Technology Conference*, (Orlando), Oct 2003.

[25] Kelly, F. P., Maulloo, A., and Tan, D., "Rate control in communication networks: Shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, Mar 1998.

[26] Ko, Y., Shankarkumar, V., and Vaidya, N., "Medium access control protocols using directional antennas in ad hoc networks," in *Proceedings of IEEE INFOCOM*, Mar. 2000.

[27] Korakis, T., Jakllari, G., and Tassiulas, L., "A MAC protocol for full exploitation of directional antennas in ad-hoc wireless networks," in *Proceedings of ACM MOBIHOC*, (Annapolis, MD), Jun 2003.

[28] Krishnamurthy, S., Acampora, A., and Zorzi, M., "Polling based media access protocols for use with smart adaptive array antennas," in *IEEE Trans. on Networking*, Apr 2001.

[29] Lal, D., Toshniwal, R., Radhakrishnan, R., Agrawal, D. P., and Caffery, J., "A novel MAC layer protocol for space division multiple access in wireless ad hoc networks," in *International Conference on Computer Communications and Networks*, Oct 2002.

[30] Laneman, J. N. and Wornell, G. W., "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. on Information Theory*, vol. 49, pp. 2415–2425, Oct 2003.

[31] Lee, J.-W., Chiang, M., and Calderbank, R., "Optimal MAC design based on network utility maximization: reverse and forward engineering," in *Proc. of IEEE INFOCOM*, Apr 2006.

[32] Liberti, J. and Rappaport, T., "Smart antennas for wireless communciations," in *Prentice Hall PTR*, 1999.

[33] Mailloux, R. J., *Phased array antenna handbook*. Artech House, Nov 1993.

[34] Martin, C. C., Winters, J. H., and Sollenberger, N. R., "Multiple-input multiple-output (MIMO) radio channel measurements," in *IEEE VTS - Fall VTC*, (Boston, MA), Sept 2000.

[35] Mundarath, J. and Ramanathan, P., "NULLHOC: A MAC Protocol for Adaptive Antenna Arrays based Wireless Ad-Hoc Networks in Multipath Environments," in *IEEE Globecom*, 2004.

[36] Naguib, A. F., Paulraj, A., and Kailath, T., "Capacity improvement with base-station antenna arrays in cellular CDMA," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 691–698, Aug 1994.

[37] Nandagopal, T., Kim, T.-E., Gao, X., and Bhargavan, V., "Achieving MAC Layer Fairness in Wireless Packet Networks," (Proceedings of ACM MOBICOM 2000), Aug 2000.

[38] Nasipuri, A., Li, K., and Sappidi, U. R., "Power consumption and throughput in mobile ad hoc networks using directional antennas," in *Proc. of IEEE IC3N*, Oct. 2002.

[39] Nasipuri, A., Ye, S., and Hiromoto, R., "A MAC protocol for mobile ad hoc networks using directional antennas," in *IEEE Wireless Communication and Networking Conference(WCNC)*, 2000.

[40] Ochiai, H., Mitran, P., and Tarokh, V., "Design and analysis of collaborative diversity protocols for wireless sensor networks," in *IEEE VTC*, 2004.

[41] Peraki, C. and Servetto, S. D., "On the maximum stable throughput problem in random networks with directional antennas," in *Proceedings of ACM MOBIHOC*, Jun 2003.

[42] Radhakrishnan, R., Dhananjay, L., Caffery, J., and Agrawal, D. P., "Performance comparison of smart antenna techniques for spatial multiplexing in wireless ad hoc networks," in *IEEE WPMC*, Oct 2002.

[43] Radunovic, B. and Boudec, J. L., "Rate performance objectives of multi-hop wireless networks," in *Proceedings of IEEE INFOCOM*, (Hong Kong), Mar 2004.

[44] Ramanathan, R., "On the performance of ad hoc networks with beamforming antennas," in *Proc. of ACM MOBIHOC*, Oct. 2001.

[45] Rappaport, T., *Smart antennas: Adaptive arrays, algorithms, and wireless position location*. IEEE Press, Sept 1998.

[46] Ren, T., Koutsopoulos, I., and Tassiulas, L., "Efficient media access protocols for wireless LANs with smart antennas," in *IEEE WCNC*, Mar 2003.

[47] Rose, D., Tarjan, R. E., and Lueker, G., "Algorithmic Aspects of Vertex Elimination on Graphs," *SIAM Journal*, pp. 5:146–160, 1976.

[48] Roy, S., Saha, D., Bandhyopadhyay, S., Ueda, T., and Tanaka, S., "A network-aware MAC and routing protocol for effective load balancing in ad-hoc wireless networks with directional antenna," in *Proceedings of ACM MOBIHOC*, Jun 2003.

[49] Sadeghi, B., Kanodia, V., Sabharwal, A., and Knightly, E., "Opportunistic media access for multirate ad-hoc networks," in *ACM Mobicom*, 2002.

[50] Sakr, C. and Todd, T. D., "Carrier-sense protocols for packet-switched smart antenna basestations," *International Journal of Wireless Information Networks*, vol. 7, pp. 133–148, July 2000.

[51] Sanchez, M., Giles, T., and Zander, J., "CSMA/CA with beam forming antennas in multi-hop packet radio networks," in *Swedish Workshop on Wireless Ad-hoc Networks*, (Stockholm), Mar. 2001.

[52] Seshadri, N. and Winters, J. H., "Two signaling schemes for improving the error performance of FDD transmission systems using transmitter antenna diversity," *International Journal of Wireless Information Networks*, Jan. 1994.

[53] Shad, F., Todd, T., Kezys, V., and Litva, J., "Indoor SDMA capacity using a smart antenna basestation," in *IEEE ICUPC*, pp. 868–872, 1997.

[54] Sheikh, K., Gesbert, D., Gore, D., and Paulraj, A., "Smart antennas for broadband wireless access networks," *IEEE Communications Magazine*, vol. 37, pp. 100–105, Nov 1999.

[55] Shekar, H., Sundaresan, K., and Ingram, M., "Linear and non-linear receiver prcessings in MIMO ad-hoc networks," in *Proceedings of IEEE WPMC*, (Aalborg, Denmark), Oct 2005.

[56] Shiu, D., Foschini, G., Gans, M., and Kahn, J., "Fading correlation and its effect on the capacity of multiple-element antennas," *IEEE Transactions on Communications*, vol. 48, Mar 2000.

[57] Singh, H. and Singh, S., "A MAC protocol based on adaptive beamforming for ad hoc networks," in *IEEE PIMRC*, Sept 2003.

[58] Singh, H. and Singh, S., "Tone based MAC protocol for use with adaptive array antennas," in *IEEE WCNC*, Mar 2004.

[59] Spyropoulos, A. and Raghavendra, C. S., "Energy efficient communications in ad hoc networks using directional antennas," in *IEEE INFOCOM*, June 2002.

[60] Spyropoulos, A. and Raghavendra, C. S., "Asympotic capacity bounds for ad-hoc networks revisited: the directional and smart antenna cases," in *IEEE GLOBECOM*, Dec 2003.

[61] Sundaresan, K. and Sivakumar, R., "A unified MAC layer framework for ad-hoc networks with smart antennas," in *Proceedings of ACM MOBIHOC*, (Tokyo, Japan), May 2004.

[62] Sundaresan, K. and Sivakumar, R., "Routing in ad-hoc networks with MIMO links," in *GNAN Technical Report*, (http://www.ece.gatech.edu/research/GNAN/archive/tr-routing.pdf), Jun 2004.

[63] Sundaresan, K. and Sivakumar, R., "Algorithmic aspects of communication in ad-hoc networks with smart antennas," in *Proceedings of ACM MOBIHOC*, (Florence, Italy), May 2006.

[64] Sundaresan, K. and Sivakumar, R., "A unified MAC layer framework for ad-hoc networks with smart antennas," *To Appear in IEEE/ACM Transactions on Networking*, 2007.

[65] Sundaresan, K., Sivakumar, R., Ingram, M., and Chang, T., "A fair medium access control protocol for ad-hoc networks with MIMO links," in *Proceedings of IEEE INFOCOM*, (Hong Kong), Mar 2004.

[66] T. Tang, M. Park, R. W. H. J. and Nettles, S. M., "A MIMO-OFDM Transceiver for Ad Hoc Networking," in *Intnl. Workshop on Ad-hoc Wireless Networks*, 2004.

[67] Takai, M., Martin, J., Bagrodia, R., and Ren, A., "Directional virtual carrier sensing for directional antennas in mobile ad hoc networks," in *Proceedings of ACM MOBIHOC*, Jun 2002.

[68] Takata, M., Nagashima, K., and Watanabe, T., "A dual access mode MAC protocol for ad hoc networks using smart antennas," in *IEEE ICC*, June 2004.

[69] Thomas, G., "Random access with multi-beam antenna arrays," in *IEEE WCNC*, Mar 2002.

[70] Ueda, T., Masayama, K., Horisawa, S., Kosuga, M., and Hasuike, K., "Evaluating the performance of wireless ad hoc network testbed with smart antenna," in *International Workshop on Mobile and Wireless Communications Network*, pp. 135–139, 2002.

[71] V.V.Phan, S.G.Glisic, and D.D.Luong, "Packet-length adaptive CLSP/DS-CDMA: Performance in burst-error correlated fading channels," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 147–158, Jan 2004.

[72] Wang, X. and Kar, K., "Cross layer rate control for end-end proportional fairness in wireless networks with random access," in *Proc. of ACM MOBIHOC*, May 2005.

[73] Winters, J. H. and Gans, M. J., "The range increase of adaptive versus phased arrays in mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 353–362, Mar 1999.

[74] Winters, J. H. and Gans, M. J., "The range increase of adaptive versus phased arrays in mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 353–362, Mar 1999.

[75] Winters, J. H., Salz, J., and Gitlin, R. D., "The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems," *IEEE Transactions on Communications*, vol. 42, no. 2, 1994.

[76] Yamada, Y., Kagoshima, K., and Tsunekawa, K., "Diversity antennas for base and mobile stations in land mobile communication systems," *IEICE Transactions*, vol. E 74, pp. 3202–3209, Aug 1980.

[77] Yi, S., Pei, Y., and Kalyanaraman, S., "On the capacity improvement of ad hoc wireless networks using directional antennas," in *Proc. of ACM MOBIHOC*, Jun 2003.

[78] Zheng, L. and Tse, D., "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[79] Zorzi, M., "On the capture performance of smart antennas in a multicellular environment," *IEEE Transactions on Communications*, vol. 50, pp. 536–539, Apr 2002.

PUBLICATIONS

- *Conference Publications:*

1. K. Sundaresan, W. Wang and S. Eidenbenz, "Algorithmic Aspects of Communication in Ad-hoc Networks with Smart Antennas," *Proceedings of ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, May 2006.

2. K. Sundaresan and R. Sivakumar, "Routing in ad-hoc networks with MIMO links," **Best Paper Award** *Proceedings of IEEE International Conference on Network Protocols (ICNP)*, Nov 2005.

3. H. Shekhar, K. Sundaresan and M. A. Ingram, "Linear and non-linear receiver processing for MIMO ad-hoc wireless networks," *Proceedings of International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Sept 2005. "

4. Y. Zhu, K. Sundaresan and R. Sivakumar, "Practical limits on achievable energy improvements and useable delay tolerance in correlation-aware data gathering in wireless sensor networks," **Best Paper Award** *Proceedings of IEEE International Conference on Sensor and Ad Hoc Communications and Networks (SECON)*, Sept 2005.

5. K. Sundaresan and R. Sivakumar, "Rate vs. Range vs. Reliability: How to best exploit MIMO in ad-hoc networks," Poster Paper *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, May 2005.

6. Y. Zhu, K. Sundaresan and R. Sivakumar, "Exposing two critical myths about correlation-aware data aggregation," Poster Paper *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*,

May 2005.

7. A. Velayutham, K. Sundaresan, and R. Sivakumar, "Non-pipelined relay increases throughput capacity of wireless ad-hoc networks," *Proceedings of IEEE Intl. Conference on Computer Communications (INFOCOM)*, Mar. 2005.
8. K. Sundaresan and R. Sivakumar, "A unified MAC layer framework for ad-hoc networks with smart antennas," *Proceedings of ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, May 2004.
9. K. Sundaresan, R. Sivakumar, and M. A. Ingram, "A fair medium access control protocol for ad-hoc networks with MIMO links," *Proceedings of IEEE Intl. Conference on Computer Communications (INFOCOM)*, Mar. 2004.
10. K. Sundaresan, V. Anantharaman, H.-Y. Hsieh, and R. Sivakumar, "ATP: a reliable transport protocol for ad-hoc networks," *Proceedings of ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, Jun. 2003.
11. K. Sundaresan and R. Sivakumar, "On the medium access control problem in ad-hoc networks with smart antennas," *Extended Abstract in Proceedings of ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, Jun. 2003.

- *Journal Publications:*

1. K. Sundaresan and R. Sivakumar, "A unified MAC layer framework for ad-hoc networks with smart antennas," To Appear in *IEEE/ACM Transactions on Networking*, Aug. 2007.
2. K. Sundaresan, and R. Sivakumar, "Ad-hoc Networks with heterogeneous smart antennas: performance analysis and protocols," To Appear in *Wireless Communications and Mobile Computing Journal (WCMC)*, Special Issue on Ad-Hoc Networks: Technologies and Challenges, Nov. 2006.
3. K. Sundaresan, V. Anantharaman, H.-Y. Hsieh, and R. Sivakumar, "ATP: a reliable transport protocol for ad-hoc networks," in *IEEE Transactions on Mobile Computing*, Nov-Dec. 2005.

4. K. Sundaresan, R. Sivakumar, and M. A. Ingram, "Medium access control in ad-hoc networks with MIMO links: optimization considerations and algorithms," *IEEE Transactions on Mobile Computing*, Nov. 2004.

5. V. Anantharaman, S.-J. Park, K. Sundaresan, and R. Sivakumar, "TCP performance over mobile ad-hoc networks: a quantitative study," *Wireless Communications and Mobile Computing Journal (WCMC)*, Special Issue on Performance Evaluation of Wireless Networks, Mar. 2004.

6. K. Sundaresan, H.-Y. Hsieh, and R. Sivakumar, "IEEE 802.11 over multi-hop wireless networks: problems and new perspectives," *Ad-hoc Networks Journal*, Elsevier Publishers, Feb. 2004.

- *Book Chapters*

1. K. Sundaresan, S-J. Park, and R. Sivakumar, "Transport layer solutions for ad-hoc networks," To appear in *Ad-hoc Networks: Technologies and Protocols*, Kluwer, 2005.

VITA

Karthikeyan Sundaresan was born and brought up Chennai, a coastal city in the southern part of India. Chennai is a wonderful coastal town, always bustling with activity. He did his schooling and under-graduation from the same city. He obtained his bachelors degree in engineering from Anna University in Aug 2001, a premier institution in southern India. His major was in Electronics and Communication. He then moved to Atlanta, USA in August 2001 to pursue his interest in graduate studies at Georgia Institute of Technology, where he obtained his masters degree in April 2003 and is currently working towards in Ph.D degree in the school of electrical and computer engineering. He has also been the recipient of several awards at both the under-graduate and graduate level like the University Gold Medal, Ramanujam Centennial Gold Medal, Best Masters Thesis Award, Best Paper Awards, etc.