ERROR CONTROL SCHEMES AND DIRECTIONAL ANTENNAS IN WIRELESS NETWORKS

By

Su Yi

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Electrical, Computer and System Engineering

Approved by the
Examining Committee:

____
Shivkumar Kalyanaraman, Thesis Adviser

____
Babak Azimi-Sadjadi, Co-Thesis Adviser

____
Biplab Sikdar, Member

____
Alhussein Abouzeid, Member

____
Christopher D. Carothers, Member

Rensselaer Polytechnic Institute
Troy, New York

December 2005
(For Graduation December, 2005)

ERROR CONTROL SCHEMES AND DIRECTIONAL ANTENNAS IN WIRELESS NETWORKS

By

Su Yi

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Electrical, Computer and System Engineering

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Examining Committee:

Shivkumar Kalyanaraman, Thesis Adviser

Babak Azimi-Sadjadi, Co-Thesis Adviser

Biplab Sikdar, Member

Alhussein Abouzeid, Member

Christopher D. Carothers, Member

Rensselaer Polytechnic Institute
Troy, New York

December 2005
(For Graduation December, 2005)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENT

First I would like to thank my thesis advisor Prof. Shivkumar Kalyanaraman, who was very understanding to me throughout my doctoral study. He opened the door to the computer networking area for me, a newcomer to this field at that time. It was his class that stimulated me to decide my area of study in my first year at RPI. He was inspiring and helpful in all aspects of my thesis. He was the person who showed me how to think, execute, and communicate. This thesis would not happen without his support and motivation.

I cannot forget my co-advisor Prof. Babak Azimi-Sadjadi's tremendous help on my Ph.D study. I am sincerely grateful to his advice and patience when discussing all the details of my research work and paper submissions. I also would like to thank all the other members in my dissertation committee, Prof. Biplab Sikdar, Prof. Alhussein Abouzeid, and Prof. Christopher D. Carothers for kindly reviewing my dissertation and giving valuable comments. I am also in debt to Prof. John W. Woods for supporting my last semester of study. Thanks are due to Prof. Alejandra Mercado for our fruitful discussion on the cooperative networks.

While at RPI, I had the privilege of interacting with wonderful bright and talented people. They have taught me much and their advice, feedback and friendship have made my Ph.D experience both more educational and more fun: Yufeng Shan, Fengji Ye, Jayasri Akella, Vijaynarayanan Subramanian, Hua Yang, Huaming Wu, Rajagopal Iyengar, Shivani Deshpande, Xin Wang, former graduates Yong Xia, Omesh Tickoo, Kartikeya Chandrayana, Jiang Li, Satish Raghunath, Aleck Liu, and all other people from our Networks Lab. It has been a pleasure to work with all of you. I especially would like to thank Dr. Yong Pei for all his useful discussion and guidance on the directional antenna project.

I also very much appreciate the help from our department administration, Ms. Ann Bruno and Ms. Priscilla Magilligan, especially for the help during the period when I struggled through my visa status problem.

Finally, to my dear husband, Chunyang, who is always there for me. There is no word that can express my gratitude to his help and his sacrifice all these years. Thanks and best wishes to my mom and my sister who gave me continuing support and encouragement. I could not have been able to accomplish this without my family.

ABSTRACT

The focus of this thesis is on finding better traffic-carrying capabilities of wireless ad hoc networks, with interdisciplinary study on wireless networking and communications.

First we investigate the performance of multimedia applications in wireless networks and the impact of various error control protocols. We propose a two stage error control scheme that improves the effective throughout of wireless networks. We apply error control to the packet header and packet load separately. The network intermediate nodes either use header FEC or header CRC checksum to successfully transport the packets from the source to the destination. Only at the destination, the error in the payload is corrected. We compare the proposed schemes with 802.11 protocol and show that header error protection strategy can effectively increase the video performance.

Then we propose a system design of cooperative networks. This cooperation technique is intended to give innovations at various layers as to improve overall system throughput and system reliability. Under a cluster-based network design, code combining is used together with FEC to improve the link reliability. The analytical results and the simulations show that with the cooperation of nodes in a clustering network, the link reliability will be greatly improved. We also show that not only transmission power is greatly reduced, but also the aggregate power consumption for a successful transmission and reception.

The capacity of wireless ad hoc networks is constrained by the interference caused by the neighboring nodes. Lastly we investigate the capacity of ad hoc wireless networks using directional antennas. Using directional antennas reduces the interference area caused by each node, thus increases the capacity of the network. We will give an expression for the capacity gain and we argue that in the limit, when the beam-width goes to zero the wireless network behaves like the wired network. In our analysis we consider both arbitrary networks and random networks where nodes are assumed to be static.

We conclude that by looking at some various novel aspects of the wireless networks and communications, the capacity and performance improvement of the network can be made a reality.

CHAPTER 1
Introduction

1.1 Wireless Ad Hoc Networks

The ability to communicate with anyone on the planet from anywhere on the planet has been mankind's dream for a long time. Wireless is the only medium that can enable such untethered communication. With the current advances in VLSI and wireless technologies, it is now possible to build high-speed wireless systems that are cheap as well as easy to install and operate.

Ad hoc networks are multi-hop wireless networks that are composed of mobile hosts communicating with each other through wireless links. These networks are typically characterized by scarce resources (e.g. bandwidth, battery power etc.), lack of any established backbone infrastructure and a dynamic topology. Some challenging but critical tasks that researchers have tried to address over the past years have been the fundamental limits of the network performance and the development of protocols that best suit the characteristics of ad hoc networks.

Wireless ad hoc networks consist of mobile nodes communicating over a wireless channel. Any node can in principle transmit data to any other; however, because of the nature of the wireless channel, each node can effectively transmit to only some of the others, typically those that lie in its vicinity. On the other hand, the traffic requirements are taken to be arbitrary, therefore it is necessary that nodes cooperate to forward each other's packets to their final destinations. If some of the nodes need to communicate with destinations on other networks, for example the Internet, they will do so by routing their traffic through nodes that lie on the boundary of the two networks and act as gateways. Figure 1.1 shows an example of the components and the topology of wireless ad hoc networks.

The concept of wireless ad hoc networks is certainly not new. Because of their suitability for the battle field environment, they had been the subject of intense research starting from 1970's. At the time, they were referred to as packet radio networks. By the middle of 1990's the interest of the research community in wireless

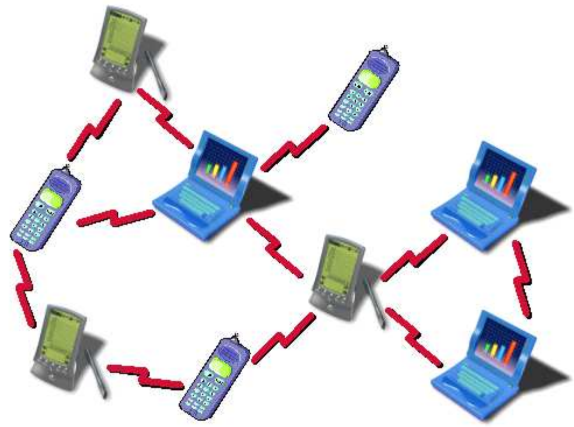


Figure 1.1: Wireless Ad Hoc Networks

ad hoc networks intensified.

Historically, wireless data communications was principally the domain of large companies with specialized needs; for example, large organizations that needed to stay in touch with their mobile sales force, or delivery services that needed to keep track of their vehicles and packages. However, this situation is steadily changing and wireless data communications is becoming as commonplace as its wired counterpart.

The need for wireless data communications arises partially because of the need for mobile computing and partially because of the need for specialized applications. Mobile computing, which aims to migrate the computing world onto a mobile environment, is affected primarily by two components: portability and connectivity.

Portability, i.e., the ability to untether computers from the conventional desktop environment, is getting increasingly feasible because with the continuous improvement in integration, miniaturization, and battery technology, the differences in performance and cost between desktop and portable computers is shrinking. Therefore, the processing power of desktop computing is becoming available to portable environments and this is highly desirable as far as productivity is concerned.

Regarding the connectivity, i.e., the ability to connect to external resources and have access to external data, wireless data technology plays a significant part because it can offer ubiquitous connectivity, that is, connectivity at any place, any time. For this reason, wireless data technology can be of real value to the business world since computer users become more productive when they exploit the benefits of connectivity. The explosive growth of local area network (LAN) installations over the past several years is ample evidence of the importance placed on connectivity by the business world. Usually, portability and connectivity are at odds: the more portability increases, the more difficult it becomes to connect to external resources. However, wireless data technology provides the means to effectively combine both capabilities and, therefore, it is an essential technology for mobile computing. For wide-area mobility there are mainly two available technologies: data transmission over cellular networks, whether analog or digital, and data transmission over mobile data networks. The main difference between these two technologies is the data transport mode. Cellular networks, being primarily voice oriented, utilize circuit switching technology and, therefore, are optimized to isochronous data traffic conditions, whereas mobile data networks employ packet switching technology and are ideal for asynchronous data traffic transmission.

As a special form of mobile data networks, wireless ad hoc networks have an obvious advantage over cellular networks on their robustness: If a node dissipates its power supply, or malfunctions, or otherwise disappears, the nodes in its vicinity will take over its routing responsibilities. On the other hand, if a base station becomes inoperable, all the nodes in its cell will lose their connection to the network.

A second advantage is the superior resource utilization that ad hoc networks achieve: in cellular networks, users that don't have a good wireless link with any base station are either denied service, or the system consumes a lot of resources (bandwidth and energy) to support their operation. On the other hand, in ad hoc networks there are many different paths with which a packet can reach its destination. If the channel between two nodes is in a deep fade, the network will work around that link (provided there are other nodes around to handle the traffic).

Contrary to cellular networks, wireless ad hoc networks are totally wireless. They also have a flat and redundant topology. Therefore, they can be deployed very fast, and so are a natural match for applications such as search and rescue operations. Moreover, there are many applications where it is simply not feasible or convenient to set up a wired infrastructure.

There are some other reasons for why the ad hoc topology may be more appropriate than the cellular topology. For example, in sensor networks where low node densities are preferable, we would need as many base stations as there are users. In home environments with a few smart appliances that only need to communicate among themselves, a gateway to other networks is not needed, so a base station would be redundant. Since ad hoc networks of hundreds or thousands of nodes are envisioned for some applications (for example networks of automobiles cruising on highways), it is imperative that all algorithms employed by the nodes be distributed. This will greatly improve their usefulness, however it will definitely complicate their design.

To conclude, we can summarize the most important features of wireless ad hoc networks, by describing them as mobile distributed multi-hop wireless networks without backbone support.

1.2 Wireless Link Issues

The unique properties of the wireless medium make the design of protocols very different from, and more challenging than, wireline networks. The unique features of wireless systems and their medium are:

• Half-Duplex Operation: In wireless systems it is very difficult to receive data when the transmitter is sending data. This is because when a node is transmitting data, a large fraction of the signal energy leaks into the receive path. This is referred to as self-interference. The transmitted and received power levels can differ by orders of magnitude. The leakage signal typically has much higher power than the received signal, which makes it impossible to detect a received signal while transmitting data. Hence, collision detection is not possible while sending data and so Ethernet-like protocols cannot be used. Due to the half-duplex mode of operation, the uplink and downlink need to be multiplexed in time (TDD) or frequency (FDD).

• Time Varying Channel: Radio signals propagate according to three mechanisms: reflection, diffraction, and scattering. The signal received by a node is a superposition of time-shifted and attenuated versions of the transmitted signal. As a result, the received signal power varies as a function of time. This phenomenon is called multipath propagation. The rate of variation of the channel is determined by the coherence time of the channel. When the received signal strength drops below a certain threshold, the node is said to be in fade. Handshaking is a widely used strategy to mitigate time-varying link quality. When two nodes want to communicate with each other, they exchange small messages that test the wireless channel between them. A successful handshake indicates a good communication link between the two nodes.

• Error-prone: As a consequence of the time-varying channel and varying signal strength, errors are more likely in wireless transmissions. In wireline networks, the bit error rates are typically less than 10^{-6} and as a result the probability of a packet error is small. In contrast, wireless channels may have bit-error rates as high as 10^{-3} or higher, resulting in a much higher probability of packet errors. In wired networks these errors are usually due to random noise. In contrast, the errors on a wireless link occur in long bursts when the node is in fade. Packet loss due to burst errors can be minimized by using smaller packets, forward error correcting codes or retransmission methods.

• Location-Dependent Carrier Sensing: It is well known that in free space, signal strength decays with the square of the path length [1]. As a result carrier sensing is a function of the position of the receiver relative to the transmitter. In the wireless medium, because of multipath propagation, signal strength decays according to a power law with distance. Only nodes within a specific radius of the transmitter can detect the carrier on the channel. This location-dependent carrier sensing results in three types of nodes when using carrier sensing: hidden nodes, ones that are within the range of the intended destination but out of range of the sender; exposed nodes, complementary to hidden nodes, are within the range of the sender but out of range of the destination; capture which is said to occur when a receiver can cleanly receive a transmission from one of two simultaneous transmissions, both within its range.

1.2.1 Medium Access Control

With the proliferation of mobile computers, their limited computing resources, and the popularity of Internet access, there is a growing need for these computers to be networked. In response, the standard which specifies the characteristics of the physical layer, as well as the Medium Access Control (MAC) protocols in the link layer for the wireless ad hoc networks need to be proposed [2].

Wireless ad hoc networks have been extensively studied for the past years. The design of these networks or protocols must take into consideration issues such as half-duplex operation, time-varying channel, bursty errors, and localized carrier sensing, which makes it a challenging problem. When doing link layer innovations, one problem that can never be overlooked is the medium access problem. Although our work does not involve new design of medium access control protocol, our link layer error control works closely with medium access control protocol. For this reason, we give a brief introduction on the basic concept of medium access control here.

Medium Access Control protocols can be classified on the basis of their network architecture into distributed and centralized MAC protocols. These protocols can be further classified into random access, guaranteed access, and hybrid protocols based on the mode of operation. Random access protocols are very efficient in multiplexing a large number of bursty sources. Hybrid protocols can give network guarantees and can support multimedia applications. Most MAC protocols have been analyzed assuming error-free channels, while in our work we assume the channel is error-prone. It is an additional challenge to provide Quality of Service in these networks.

The nature of the wireless channel brings new issues like location-dependent carrier sensing, time varying channel and channel errors. Low power requirements and half duplex operation of the wireless systems add to the challenge. Wireless MAC protocols have been heavily researched and a plethora of protocols have been proposed. Protocols have been devised for different types of architectures, different applications and different media.

The wireless medium is a broadcast medium, and therefore multiple devices can access the medium at the same time. Multiple simultaneous transmissions can result in garbled data, making communication impossible. A medium access control protocol moderates access to the shared medium by defining rules that allow these devices to communicate with each other in an orderly and efficient manner. MAC protocols therefore play a crucial role in enabling this paradigm by ensuring efficient and fair sharing of the scarce wireless bandwidth. Wireless MAC protocols have been studied extensively since the 1970s. The initial protocols were developed for data and satellite communications. We are now witnessing a convergence of the telephone, cable and data networks into a single unified network that supports multimedia and real-time applications like voice and video in addition to data. The multimedia applications require delay and jitter guarantees from the network. This demand of the network is known as the Quality of Service guarantee. These requirements have led to novel and complex MAC protocols that can support multimedia traffic.

1.3 Performance Metrics

It is necessary to understand the metrics that are used to compare the protocols designed at various layers for wireless networks. Delay, throughput, robustness against channel errors and battery power consumption are the widely used metrics to compare wireless network protocols. Following is a brief discussion of these metrics:

• Throughput: Throughput is the fraction of the channel capacity used for data transmission. A good protocol designer's objective is to maximize the throughput while minimizing the access delay. Average throughput is always evaluated to assess the overall throughput performance.

• Delay: Delay is defined as the average time spent by a packet from end to end, i.e., from the instant it is sent by the application till it reaches the destination process. Delay is a function of protocol and traffic characteristics. Therefore, when comparing protocols, it is necessary to compare them based on the same traffic parameters.

• Robustness against Channel Condition: The wireless channel is time-varying and error-prone. Channel fading can make the link between two nodes unusable for short periods of time. Such link failures should not result in unstable behavior.

• Power Consumption: Most wireless devices have limited battery power. Hence, it is important for any protocols to conserve power and provide some power saving features.

We will consider these metrics when we propose new network protocols and evaluate the network performance. Fairness, Quality of Service (QoS), supporting node mobility, and stability are additional metrics used to compare some specific wireless network protocols.

1.4 Contributions

The ultimate purpose of this thesis is to achieve better traffic-carrying capabilities of wireless ad hoc networks, with interdisciplinary study on wireless networking and communications. The theoretical investigation brings forward good design principles, therefore a part of the thesis is devoted to the design of wireless ad hoc networks based on these principles.

In a multi-hop network, the throughput is essentially important for real-time applications due to their high bit rate requirement. We investigate the performance of multimedia applications in wireless networks and the impact of different error control protocols. In particular, we propose a two stage error control scheme that improves the effective throughout of wireless networks. We apply error control to the packet header and packet load separately. The network intermediate nodes either use header FEC or header CRC checksum to successfully transport the packets from the source to the destination. Only at the destination, the error of the load is corrected. We compare the proposed schemes with 802.11 protocol through extensive multi-hop simulations. Specifically, average throughput, end-to-end latency, and video PSNR results are analyzed. The performance comparison between each scheme is discussed in detail. It is shown that header error protection strategy can effectively increase the throughput and the video performance, via both theoretical analysis and simulation results.

Then we propose a system design of cooperative networks. This cooperation technique is intended to give innovations at a variety of layers as to improve overall system throughput and system reliability. More specifically, under a cluster-based network design, code combining is used together with FEC to improve the link layer reliability. This approach is different from how code combining is used in the conventional hybrid ARQ, which is in a sequential way. The analytical results and the simulations show that with the cooperation of nodes in a clustering network, the link reliability will be greatly improved with the same power consumption. We also show that not only transmission power is greatly reduced, but also the aggregate power consumption for a successful transmission and reception. This result is promising in that the reduced power requirement leads to less interference caused by a transmission, thus can improve the capacity of the wireless networks.

Most research on the network capacity, although from various aspects of the network performance, is motivated by the capacity analysis by Gupta and Kumar. The capacity of wireless ad hoc networks is constrained by the interference caused by the neighboring nodes. It is shown that the throughput for such networks is only $\Theta(\frac{W}{\sqrt{n}})$ bits per second per node in a unit area domain when omnidirectional antennas are used. Lastly we investigate the capacity of ad hoc wireless networks using directional antennas. Using directional antennas reduces the interference area caused by each node, thus increases the capacity of the network. We will give an expression for the capacity gain and we argue that in the limit, when the beamwidth goes to zero the wireless network behaves like the wired network. In our analysis we consider both arbitrary networks and random networks where nodes are assumed to be static. We have also analyzed hybrid beamform patterns that are a mix of omnidirectional/directional and a better model of real directional antennas. Simulation results for the network with the deployment of directional antennas are conducted for validation of our analytical results. We also propose some sketches on new designs to make use of the flexible beamforming, multi-beam antenna techniques to help design the MAC protocol, and help the node cooperation as well.

We conclude that by looking at some various novel aspects of the wireless networks and wireless communications, like head error protection, node cooperation, and the deployment of directional antennas, the capacity and performance improvement of the network can be made a reality.

1.5 Organization of the Thesis

This thesis looks at the question of improving capacity and performance in the wireless ad hoc networks through use of lower layer innovations. It first starts with the introduction of ad hoc networks and issues of our interests. Chapter 2 is dedicated to literature survey. Chapter 3 first outlines the problems with current error control technique, especially with multimedia applications, and then presents several error control algorithms, called Header Error Protection, for improving the performance of the multimedia transmissions. In Chapters 4 we propose a novel cooperative network where cooperation lies in various layers for the purpose of achieving a better system behavior. In this chapter, we put more focus on the analysis and simulations of link layer cooperation. In Chapter 5 we briefly review the existing work on capacity analysis and propose a model for analyzing the capacity improvement by using directional antennas. Finally in Chapters 6 we present the conclusions and the future work respectively.

CHAPTER 2
Related Work

2.1 Capacity and Performance Issues in Wireless Networks

Broadly speaking, research on capacity attempts to identify theoretical upper and lower bounds on the achievable performance of networks. Conceptually, network capacity cannot be associated with a particular layer, but rather occupies an extra dimension. Given its obvious importance, it is no surprise that this issue has attracted significant research interest over the years. In recent years topics on capacity of the wireless network have been studied from a variety of aspects of the network performance [3, 4, 5, 6, 7, 8, 9, 10, 11]. Most research work is motivated by the breakthrough of Gupta and Kumar on the capacity of wireless networks [12]. Their critical idea is to allow the number of nodes to go to infinity, thereby allowing statistical averaging to have an effect. Consequently, they are able to derive upper and lower bounds on the capacity that hold for all networks, in the limit of a large number of nodes, since, in that limit, all networks are essentially the same. Their major conclusion is that considering an ad hoc network with n nodes randomly located in a domain of area one square meter, under a Protocol Model of interference, such a network could provide a per node throughput of $\Theta(\frac{1}{\sqrt{n \log n}})$ bits/sec. It was also shown there that even under the best possible placement of nodes, such a network could not provide a per-node throughput of more than $O(\frac{1}{\sqrt{n}})$ bits/sec. In this case, the total end-to-end capacity is roughly $O(\frac{n}{\sqrt{n}})$, which is $O(\sqrt{n})$.

Works have been directed to study the factors that affect the performance of the ad hoc networks. Li et al.[3] examine interactions of the 802.11 MAC and ad hoc forwarding and the effect on capacity for several simple configurations and traffic patterns. It is shown that for total capacity to scale up with network size the average distance between source and destination nodes must remain small as the network grows. In [4], Grossglauser and Tse propose a scheme that takes advantage of the mobility of the nodes. By exploiting node mobility as a type of multiuser diversity, they show that the throughput can increase dramatically when nodes are mobile rather than fixed. Gastpar and Vetterli [5] study the capacity under a different traffic pattern. There is only one active source and destination pair, while all other nodes serve as relay, assisting the transmission between this source-destination pair. The capacity is shown to scale as $O(\log n)$. Both [6] and [9] study the throughput capacity of hybrid wireless networks formed by placing base stations in an ad hoc network. This is not a pure wireless ad hoc network since these base stations are connected by a high-bandwidth wired network. Recent work [10] gives throughput bound which demonstrates that throughput increases with node density. The reason why this result is contrasting with [12] is the assumption on finite power, large bandwidth, and the explicit use of link adaption.

Given the guidance of the work on capacity of wireless ad hoc networks, the system design from every perspective is examined to meet the practical needs. Thus the performance issue also draws deep attention from researchers. Unlike capacity problem, which could be pure theoretical, the performance of wireless networks is closely related to the protocol design, interfaces between various network layers, and so on. The focus of this thesis is on finding better traffic-carrying capabilities of wireless ad hoc networks, with interdisciplinary study on wireless networking and communication. For instance, the deployment of directional antennas at each node can achieve a higher capacity of the network.

In order to evaluate the performance and spatial reuse property of directional antennas, Nasipuri et al. [13] propose a MAC protocol for an ad hoc network that are equipped with multiple directional antennas. Their protocol uses a variation of the RTS/CTS exchange to let both source and destination nodes determine each other's directions. Simulation experiments indicate an average throughput improvement of 2 3 times over omnidirectional antennas. A DMAC protocol is presented in [14] that exploits the characteristics of both directional and omnidirectional antennas to allow simultaneous transmissions that are not allowed in the 802.11 protocol. Choudhury et. al. [15] design a MMAC protocol which uses multi hop RTS's to establish links between distant nodes, and then transmit CTS, DATA and ACK over a single hop. Simulation results show that MMAC outperforms DMAC as well as 802.11 on most of the topology configurations and the traffic patterns. Ramanathan [16] raises several interesting issues resulted from spatial reuse and larger transmission range of switched or steered beam-forming antenna. He evaluates the effectiveness of a number of enhancements, including channel access approaches, link power control, and directional neighbor discovery. Simulation results show that beam-forming can yield a 28% to 118% improvement in throughput and up to a factor-of-28 reduction in delay. Bao et. al. [17] propose ROMA (Receiver-Oriented Multiple Access), a distributed channel access scheduling protocol for ad hoc networks with directional antennas that are capable of forming multiple beams to carry out several simultaneous data communication sessions. Unlike random access schemes that use on-demand handshakes or signal scanning to resolve communication targets, ROMA determines a number of links for activation in every time slot using only two-hop topology information.

2.2 Error Control Schemes

An important characteristics of wireless medium is that transmission errors can occur from impairments. Possible wireless impairment includes attenuation, front end overload, natural background noise, multi-path interference, path loss and motion [1]. These transmission characteristic may affect the range of wireless devices, the level of coding needed to protect frames from error, the ability of wireless networks to meet the delay requirements, as well as the design of transport layer protocols. Errors in WLAN affect network capacity, protocol design and determining positions of stations.

Reliable communication protocols require that all the intended recipients of a message receive the message intact. Two categories of error control techniques are generally considered to cope with the erroneous wireless transmissions: Forward Error Correct (FEC) and Automatic Repeat reQuest (ARQ) [18, 19, 20, 21].

Forward Error Correction (FEC) is a method in which redundant information is included in a stream of data so that data lost in transit can be recovered at the destination [22, 23]. This method is especially popular in the areas of wireless communication. FEC codes introduce controlled amounts of redundancy into a transmitted data stream, enabling the receiver to make more accurate decisions about the transmitted sequence although it is corrupted by noise over the communication channel. Tornado codes introduced in 1998 have grasped the interests of networking community for several reasons [24]. Tornado codes are instances of a class of codes called Low-Density Parity Check (LDPC) codes. They are sparse graph codes and are rate less, i.e. almost infinite number of encoded packets can be generated from the original source message. Thereupon, the receiver only needs to collect a little more bits than the size of the source message to decode the original message (in practice 5% more). Moreover, the receiver can get any packets in the fountain, without specifying which packets. This property makes it most helpful in multicast applications. Each user may receive a different sequence of packets. They can make independent decision on how many packets are needed to correctly decode the source message.

Though FEC is helpful when bandwidth is plentiful, its benefits are not easily ascertained when the bandwidth is scarce. In other words, the problem of using FEC codes is the fixed bandwidth penalty associated with the code rate. Specifically, if the number of total packets lost is greater than the number of redundancy packets, then no data is recovered [25]. Furthermore, by FEC code alone, it can not guarantee the error-free transmission due to the channel error statistics instability. FEC schemes may be hard to implement and expensive for some system in order to attain high reliability.

ARQ or Automatic Repeat reQuest [18] for retransmission requires the transmitter to re-send a block of data, when errors are detected, upon request from the receiver. When data is transmitted in packets, an ARQ scheme can be used. Whenever a packet arrives, the receiver may choose not to accept the packet, but instead request a retransmission through a feedback channel. To determine whether or not a retransmission should be requested the receiver checks the quality or the reliability of the received packet. Usually this is done by means of an error detecting code, like a cyclic redundancy check code (CRC). ARQ schemes are simple and provide high system reliability but the throughput, or final output of the decoder, is not constant and it falls rapidly with increasing channel error rate. ARQ techniques are used in unicast protocols, but they do not scale well to multicast protocols with large groups of receivers, since segment losses tend to become uncorrelated thus greatly reducing the effectiveness of retransmissions. In such cases, FEC techniques can be used, consisting in the transmission of redundant packets (based on error correcting codes) to allow the receivers to recover from independent packet losses [26]. Another of ARQ's drawback is the need for a feedback channel and the time required recovering missing packets (a full RTT).

ARQ is commonly used where data reliability is needed, e.g. in TCP, wireless MACs etc. On the other hand, FEC can tolerate some amount of losses. Therefore, by using FEC codes, the packet loss at the receiver can be made arbitrarily small, at a price of sending redundant data. The special cases when there are too many losses can be handled then by usual ARQ techniques, which becomes a hybrid ARQ scheme.

It is evident that both FEC and ARQ have their advantages and disadvantages. The choose of FEC or ARQ also depends on the length of the packets/frames. Given the error property about the wireless channel, it is obvious that the longer a packet is, the more likely the packet will be erroneous. Retransmitting too often is not a good choice for most of the cases. So what kind of scheme is best leads to discussions based on the wireless network environment, the packet size, and so on.

A complementary means for optimally meeting the error-control requirements is through adaptive consolidation of MAC-layer ARQ and PHY-layer FEC. Such a hybrid FEC/ARQ (HARQ) scheme, first suggested in [27], uses an error control code in conjunction with the retransmission scheme.

Hybrid ARQ (HARQ) is an exciting development in 3G networks wherein information blocks are encoded for partial error correction at receiver (i.e. FEC) and additional uncorrected codes are retransmitted (i.e. ARQ). We propose to build off these developments and move these functionalities end-to-end where it correctly belongs. Two main types of HARQ have been proposed - type I and type II HARQ. A HARQ system of type-I implies that the same message, i.e. the same packet content, is sent each time that the receiver asks for a retransmission. Type II HARQ or Incremental Redundancy offers a nominal FEC protection, but then send additional parity bits during each retransmission [28, 29, 30, 31, 32]. Incremental

Redundancy requires larger size of buffer in a receiver than type I HARQ. With the success of turbo codes, HARQ using turbo codes has also drawn a great attention recently, [33, 34, 35, 36]. Some work on the comparison of ARQ protocol performance in different link levels is conducted [37].

Code combining is one of the ways of using the information in the previously received packets in order to improve performance[38, 39, 40]. Code combining involves sending a number of repetitions of coded data; decoders combine multiple coded packets before decoding. This scheme achieves gain with small buffer size in a receiver. We emphasize code combining because it is a great fit to our cooperative network design. For a cooperative network, the receiver side can have a buffer to store the copies for a same packet getting from different cooperative nodes. In this way the code combining approach can be used to decode the packet.

When the link layer detects that a frame is in error, conventionally, the frame is dropped and retransmission of the frame is requested. Based on the fact that the erroneous frames still contain useful information, packet combining and incremental redundancy can retain and utilize the erroneous frames to improve retransmission performance[41]. In this this, we address questions like: how to design a retransmission scheme to make efficient use of such information. We will consider taking advantages of both codes combining and header FEC protection to reduce the retransmission thus increase the bandwidth utilization.

We will explore the discovery of advantages of ARQ and FEC techniques with different approaches. We investigate the performance of multimedia applications in wireless networks and the impact of different error control protocols. In particular, we propose a two stage error control scheme that improves the effective throughout of wireless networks. We apply error control to the packet header and packet load separately. The network intermediate nodes either use header FEC or header CRC checksum to successfully transport the packets from the source to the destination. Only at the destination, the error of the load is corrected.

We also propose a system design of cooperative networks which give a link layer innovation as to improve overall system throughput and system reliability. Our model and analysis may propose implications toward the future link layer design.

Under a cluster-based network design, code combining is used together with FEC to improve the link layer reliability. This approach is different from how code combining is used in the conventional hybrid ARQ, which is in a sequential way.

CHAPTER 3

Improving Multimedia Transmission Performance by Using Header Error Protection

3.1 Introduction

Unlike general data transmission which needs error free delivery at each protocol layer, multimedia data can tolerate bit errors in a received packet. Some applications, such as voice over IP or video streaming, have a higher data rate requirement than accuracy requirement. In addition to congestion related packet loss and delay, that is seen in wired packet switched networks, wireless networks have to deal with a time varying, error prone, physical channel that in many instances is also severely bandwidth constrained. As such, the methods needed for wireless multimedia applications are fundamentally different from wired ones. Protocol design, such as link layer error control may impact the performance of the network and these applications.

As opposed to wired Internet streaming, to achieve a high level of acceptability and proliferation of wireless multimedia, in particular wireless video, several key requirements need to be satisfied in order to provide a reliable and efficient transmission: easy adaptability to wireless bandwidth fluctuations due to co-channel interference, multi-path fading, mobility, competing traffic; and robustness to partial data losses caused by the packetization of video frames and high packet error rate. One bit error in the link layer packet could cause the drop of the whole packet in the receiver side, even though the other bits of the packet are successfully received. This is acceptable for general data transmission, since one bit error in a file can make the whole file inaccessible. On the other hand, this may not be optimal for multimedia data transmission due to the loss tolerance of multimedia data. With partial data losses, the receiver may still decode the successfully transmitted part in a packet with desired visual quality. Therefore, at the receiver or the relays, instead of dropping the whole packet, a multimedia system can use the successfully transmitted bits in a received but corrupt packet, in order to reduce the bandwidth utilization.

Based on the above considerations, we found that error control in current 802.11 MAC protocol [42] is not efficient for supporting multimedia data transmission due to its bit error sensitivity. Therefore, in order to efficiently support multimedia data transmission we propose a new wireless link layer protocol. Even if the packet is received with some bit errors, the link layer still need to pass the packet to application layer. This approach is especially important in our proposed protocol, since we want to use the successfully received bits for multimedia applications. We call the proposed scheme HEP (Header Error Protection). A similar idea was previously used in ATM (Asynchronous Transfer Mode) which provides link-layer error correction for the packet header rather than for the entire packet [43]. A header error for both 802.11 MAC protocol and HEP based MAC protocol disrupts the transmission. Thus, the header information should be specially protected. Since the header is a small part of the packet the computational overhead of header error control is small. Error control techniques are used in this thesis to protect the header information from being corrupt. Two categories of error control techniques are considered: Forward Error Correction (FEC) and Automatic Repeat reQuest (ARQ) [18, 19, 20, 21].

There are some arguments on whether error control should reside at the link layer or at the application layer [44]. We provide another option - do part of the error control at the local level and leave some work done at the application level. Specifically we propose several header error protection schemes and analyze their impact on the throughput of the wireless networks. Recently some approaches of allowing some errors in data packets were proposed in speech transmission [45, 46]. However, video applications differ from voice applications in the bandwidth requirement, and the delay tolerance because of the buffering at the receiver side of the video streaming. Moreover, the segment size of voice applications is much smaller than that of video applications, thus error has a higher impact on voice segment than on video segment.

There are also some interesting works on performance enhancement of video transmission over wireless networks. For instance, Krishnamachari *et. al.* [47] evaluate different error control and adaptation mechanisms available in the different layers for robust transmission of video, namely MAC retransmission strategy, application-layer FEC, bandwidth-adaptive compression using scalable coding, and adaptive packetization strategies. Shan [48] propose a set of cross-layer techniques for adaptive real-time video streaming over wireless networks, including novel packetization scheme, FEC codes applied within an application packet, and a priority-based ARQ together with a scheduling algorithm. In [49], layered coding coupled with unequal error protection obtained by using different retry limits at the link level has been shown to balance the link erasure rate and the overflow rate. Zheng and Boyce [50] propose a modified version of the UDP protocol to accommodate Internet-to-wireless video traffic. A new protocol stack design is proposed to allow bi-directional information exchange so that the physical, link layers can communicate with the application layer. The improved UDP protocol can forward the physical frame error indication to the application for better error control. Other works suggest to limit the UDP Lite [51] partial checksum to the packet header [52], focusing on schemes that add redundancy at the radio link level, and allowing packets containing errors to be forwarded to the applications. However, the quality of the received multimedia streams is not measured.

In this chapter we would like to exploit the idea of protecting packet data and packet header differently in video transmissions. The rest of the chapter is organized as follows: Section 3.2 gives the background of 802.11 MAC protocol mechanism; header error protection strategies are introduced along with their throughput analysis in Section 3.3; in Section 3.4, we show our simulation results and followed by the conclusions in Section 3.6.

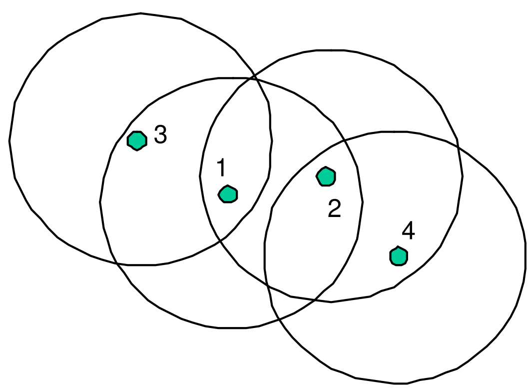
3.2 802.11 Background

Real-time streaming of audiovisual content over the Internet has been emerging as an important technology area in multimedia communications. Furthermore, with the advance of Wireless LAN (WLAN) technologies, this multimedia content will be redistributed in the future within the home to a variety of wireless devices. This WLAN architecture can be a pattern of multi-hop ad hoc networks that are composed of mobile hosts communicating with each other through wireless links. These networks are typically characterized by scarce resources (e.g. bandwidth, battery power etc.), lack of any established backbone infrastructure and a dynamic topology.

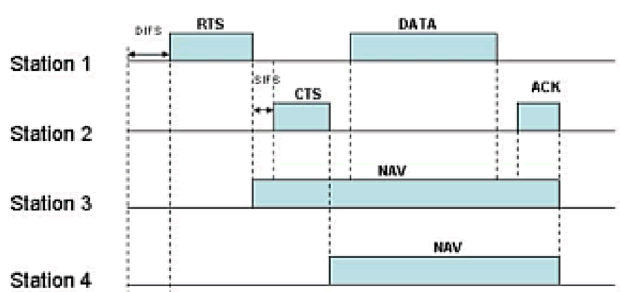
Since 802.11 is the current off-the-shelf MAC protocol for WLANs, an overview of 802.11 is presented in the following part.

The 802.11 standard ensures that all stations, both radio-based network interface cards (NICs) and access points, implement access methods for sharing the air medium. There are two kinds of medium access methods defined in 802.11 standard, the mandatory Distributed Coordination Function (DCF) mode, a form of carrier sense multiple access with collision avoidance (CSMA/CA) making certain that all stations first sense the medium before transmitting, and Point Coordination Function (PCF) mode, which is an optional access method based on polling. This thesis assumes use of the IEEE 802.11 DCF, the access method used in ad hoc mode.

The 802.11 MAC DCF is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) in which stations with data ready to transmit first determine the state of medium as busy or idle. To reduce collisions caused by hidden terminals in the network, 802.11 uses a four-way RTS/CTS/Data/Ack exchange. In brief, a node that wishes to send a data packet first sends a Request To Send (RTS) packet to the destination. If the destination believes the network is idle, it responds with a Clear To Send (CTS). The sender then transmits the data packet, and waits for an acknowledgment (ACK) from the receiver. If a node overhears an RTS or CTS, it knows the medium will be busy for some time, and avoids initiating new transmissions or sending any CTS packets. 802.11 RTS and CTS packets include the amount of time the medium will be busy for the remainder of the exchange. Each node uses these times to update its Network Allocation Vector (NAV). The NAV value indicates the amount of time remaining before the network will become available. Upon successful receipt of an RTS frame not addressed to itself, a node updates its NAV to the maximum of the time carried in the RTS frame and its current NAV value. Upon receiving an RTS addressed to itself, a node returns a CTS frame only if its NAV value is zero, otherwise no CTS is sent. Hence, a sender will see no CTS if its RTS packet has collided with another transmission at the receiver, or if the receiver's NAV indicates that the network is not available. Figure 3.1 illustrate how 802.11 DCF operate among vicinal nodes. Station 1 is the node which wants to send and Station 2 is the intended receiver. Looking through the time line, the handshake is successful so data is sent after RTS/CTS and followed by an ACK from the receiver. Station 3 overhears RTS from Station 1 and then updates its NAV to the proper value; Station 4 overhears CTS from Station 2 and updates its NAV as well. Both NAV values decrease to 0 at the time the ACK is received by Station 1.



(a) Mobile Nodes Placement



(b) DCF Operation

Figure 3.1: An Example of 802.11 DCF Operation

If a station senses the medium as busy it defers its transmission attempt by setting a backoff counter that only decrements after the medium is sensed idle for the DCF Interframe Space (DIFS). It continues to decrement as long as the medium remains idle, deferring its countdown during busy periods. The backoff counter is a uniformly distributed random number between 0 and contention window (CW). When the backoff counter reaches zero the station begins to transmit.

It is possible for multiple stations to begin transmissions at the same time, resulting in a collision. Collisions occur at the receiver, hence they must be inferred if a sender fails to receive a valid ACK from the receiver within a bounded time. All senders detecting a collision double their contention windows up to a maximum value, CW_{max}, and repeat the backoff algorithm. Reception of an ACK infers the channel access and data transmission were successful, hence the station resets its contention window CW to CW_{min} and waits for its next data to send.

In 802.11, each frame consists of the following basic components: a) a MAC header, which comprises frame control, duration, address, and sequence control information; b) a variable length frame body, which contains information specific to the frame type; c) a frame check sequence (FCS), which contains an IEEE 32-bit cyclic redundancy code (CRC). The MAC frame format comprises a set of fields that occur in a fixed order in all frames. Figure 3.2 depicts the general MAC frame format. The fields Address 2, Address 3, Sequence Control, Address 4, and Frame Body are only present in certain frame types.

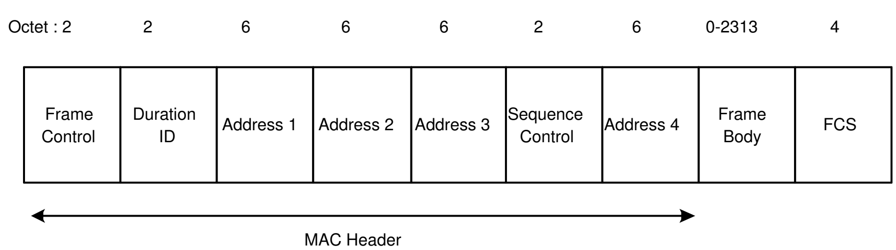


Figure 3.2: 802.11 MAC Frame Format

The ARQ protocol used by 802.11 is called the stop and wait protocol. Essentially, the transmitter sends the packet, and then waits for an ACK from the receiver before it tries to send the next MAC protocol data unit (MPDU). After transmitting a MPDU that requires an ACK frame as a response, the station shall wait for an $ACKTimeout$ interval. When timeout, the station concludes that the transmission of the MPDU has failed, and this station shall invoke its backoff procedure upon expiration of the ACKTimeout interval. If the source station does not receive an acknowledgment, it shall attempt to retransmit the failed MPDU or another eligible MPDU, according to the link retry limit, after performing the backoff procedure and the contention process.

3.3 Header Error Protection and Performance Evaluation

In wireless network the throughput is a key characteristic, especially for real-time applications, which require high bandwidth utilization to satisfy end users. Consider an ad hoc network with n nodes randomly located in a domain of area one square meter. It was shown in [12] that under a Protocol Model for interference, such a network could provide a per node throughput of $O(\frac{1}{\sqrt{n \log n}})$ bits/sec. In this case, the total end-to-end capacity of the entire network is $O(\sqrt{\frac{n}{\log n}})$. This result indicates a vanishing throughput performance as the network scales.

The *effective throughput* we discuss in this chapter is defined as the the fraction of channel bandwidth that is used to successfully transmit packets if every node is transmitting in full utilization of bandwidth. Also this effective throughput is under the impact of packet[1] error control. We will consider three packet error control schemes. We start with the ARQ scheme in the current Wireless LAN MAC layer protocol IEEE 802.11. After that, we propose two kinds of header error protection scheme: *header CRC* and *header FEC*. These two schemes are compared with the original ARQ strategy used in 802.11 protocol.

It was shown that under a Protocol Model for interference, if there are n nodes randomly placed in a network domain, the average hop number h is assumed to be $\sqrt{\frac{n}{\log n}}$; each node in the network can transmit at an average rate of $\frac{c}{\sqrt{n \log n}}$ bits/sec, where c is a constant. This thesis uses the mathematical approximations with these average values. The critical idea is to consider random topologies and traffic patterns, to allow the number of nodes to go to infinity, and to compute the performance asymptotically. In this manner, statistical averaging is introduced. All results will only hold with high probability, i.e., with probability approaching unity as the number of nodes approaches infinity. However this is a small concession to

[1] In this thesis we talk about link layer error control, yet we still use the term *packet* instead of *frame* for general use, and to differentiate with the term *frame* in video transmission.

make, given the wealth of results that can be obtained in this manner.

3.3.1 Error Models for Link Layer

Measurement and analysis of the error characteristics of a wireless network is itself a research topic. Even worse, the characteristics are very environment dependent. For this reason, varieties of experimental data is presented for wireless LANs [53]. We focus on the network performance rather than the modeling of the physical layer error properties. So we only think about a virtual bit channel, not a physical communication channel.

In this chapter we use the Binary Symmetric Channel (BSC) model with error probability p and a binary Markov channel model as our channel error models. For a BSC error model, data are transmitted on a channel with the error probability p. This is a memoryless model where the noise bits are produced by a sequence of independent trials. Each has the same probability $1 - p$ of producing a correct bit and probability p of producing a bit error. p is then the bit error rate (BER) for the wireless link.

Binary Markov channel is the first order binary Markov channel model (called Gilbert model [54, 55] for packet transmission). It is shown through analysis and simulation that a first-order Markov process is a good approximation for fading channels [56, 57]. The model is described by the transition matrix

$$\begin{bmatrix} 1 - p_{01} & p_{01} \\ p_{10} & 1 - p_{10} \end{bmatrix}$$

where p_{01} (p_{10}) is the probability that the transmission of current bit is unsuccessful (successful), given that the previous transmission was successful (unsuccessful). Note that $\frac{1}{p_{10}}$ represents the average length of a burst of errors, and the average BER is given by $\frac{p_{01}}{p_{01}+p_{10}}$.

3.3.2 Packet CRC in 802.11

In IEEE 802.11, the ARQ is a stop-and-wait ARQ with a positive ACK after each packet. The CRC checksum protects the whole packet. Usually there is a limit on the number of times that WLAN cards retransmit a packet (e.g., 4 times). Simply consider the single hop packet error probability, defined as P_{e_1} for this packet CRC scheme. For the BSC, the errors are independent, so

$$P_{e_1} = \sum_{i=1}^{q} (1-p)^{q-i} p^i \left(\begin{array}{c} q \\ i \end{array} \right) \tag{3.1}$$

where q is the packet length (in bits).

Under our assumptions, there are n nodes in the network, the aggregate throughput without considering packet dropping is $c\sqrt{\frac{n}{\log n}}$. First we assume there is no limit on the number of retransmissions. Given the probability of error P_{e_1}, the average number of retransmissions for a single hop has a geometric distribution with successful probability of $1 - P_{e_1}$. Thus the probability of number of retransmissions (excluding the first transmission) in one hop is:

$$\mathbf{P}\{ret = i\} = P_{e_1}^i (1 - P_{e_1}) \tag{3.2}$$

If a flow only has one hop distance and the bandwidth is W, then the effective throughput of this flow is

$$F(h=1) = \sum_{i=1}^{\infty} \frac{W}{i} \mathbf{P}\{ret = i-1\} = \sum_{i=1}^{\infty} \frac{W}{i} P_{e_1}^{i-1} (1 - P_{e_1}) \tag{3.3}$$

Note $\frac{d}{da}(\sum_{i=1}^{\infty} \frac{a^i}{i}) = \sum_{i=1}^{\infty} a^{i-1} = \frac{1}{1-a}$ when $|a| < 1$. So $\sum_{i=1}^{\infty} \frac{a^i}{i} = \int \frac{1}{1-a} = c_0 - \ln(1 - a)$. Let $a = 0$ to solve the constant value we get $c_0 = 0$. Then we have

$$F(h=1) = -\frac{(1 - P_{e_1}) \ln(1 - P_{e_1})}{P_{e_1}} W \tag{3.4}$$

For multi-hop networks, since the error statistics for each hop is independent, the retransmission for different wireless links are independent. Now suppose a flow has experienced h hops, consuming bandwidth $W_1, W_2, ..., W_h$ of each link respectively. Then the aggregate effective throughput of this flow is

$$\begin{aligned} F(h) &= \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{\infty}\cdots\sum_{i_h=1}^{\infty}(\frac{W_1}{i_1}+\frac{W_2}{i_2}+\cdots+\frac{W_h}{i_h}) \\ &\quad \cdot\mathbf{P}\{ret=i_1-1\}\mathbf{P}\{ret=i_2-1\}\cdots\mathbf{P}\{ret=i_h-1\} \\ &= \mathbf{E}[\frac{W_1}{i_1}]+\mathbf{E}[\frac{W_2}{i_2}]+\cdots+\mathbf{E}[\frac{W_h}{i_h}] \\ &= -\frac{(1-P_{e_1})\ln(1-P_{e_1})}{P_{e_1}}(W_1+W_2+\cdots+W_h) \end{aligned} \tag{3.5}$$

Since all the $W_1, W_2, ..., W_h$ add up to the network aggregate throughput $c\sqrt{\frac{n}{\log n}}$, summing up all the flows in the network adds up to the total aggregate effective throughput of the 802.11 protocol. Therefore, we have

$$A_1 = -c\sqrt{\frac{n}{\log n}}\frac{(1-P_{e_1})\ln(1-P_{e_1})}{P_{e_1}} \tag{3.6}$$

Note in Eqn(3.5), we set the retry limit to ∞. It is easy to see from the expression that $F(h)$ is an increasing function of the retry limit. Thus, when retry limit is less than ∞, the aggregate throughput A_1 will be less than the result give in Eqn(3.6). In fact, the effect of a positive retry limit has diminishing return. For typical BER it is easy to show that the 4 retry limit in 802.11 can be approximated with by Eqn(3.6). We use the results given by the infinity retry limit for the standard 802.11 protocol and header CRC.

3.3.3 Header CRC

Header CRC aims to protect the header, not the whole packet. There is a retransmission for the packet if error detected in header part. The main purpose of protecting header is the need to carry the correct destination address for IP forwarding, and source address for end-to-end ACK. A CRC is needed for detecting errors in this information. The probability that any error detected in a header is:

$$P_{e_2} = \sum_{i=1}^{k+r}(1-p)^{k+r-i}p^i \begin{pmatrix} k+r \\ i \end{pmatrix} \tag{3.7}$$

where k is the header size, and r is the CRC bits.

In a similar form with previously introduced packet CRC, the aggregate effective throughput of networks is:

$$A_2 = -c\sqrt{\frac{n}{\log n}}\frac{(1-P_{e_2})\ln(1-P_{e_2})}{P_{e_2}} \tag{3.8}$$

Note that the factor part $-\frac{(1-P_{e_2})\ln(1-P_{e_2})}{P_{e_2}}$ is a monotone decreasing function of P_{e_2}. This factor decreases from 1 to 0 as P_{e_2} increases from 0 to 1. This is consistent with heuristic expectations, because one expects the throughput to increase when packet error probability decreases.

3.3.4 Header Error Control Coding

In header error control coding, the network nodes use an error control coding technique to transmit the header information without error. Therefore, the network always has the correct address of the destination. In this scheme, the network might deliver the packet through with error in payload.

We also call this scheme *header FEC*, since FEC is added to the header. In header FEC, a retransmission is issued when the redundancy fails to correct the errors in header. Therefore it is in fact a hybrid ARQ for only protecting header part. BCH codes are well known codes for binary data transmission, especially good for large block codes [19]. m protection bits are added to each header for error correction. For a t-error-correcting linear code, it is capable of correcting a total of 2^m error patterns, including those with t or fewer errors. So the probability that the decoder commits an erroneous decoding in one packet is upper bounded by:

$$P_{e_3} \leq \sum_{i=t+1}^{k+m}(1-p)^{k+m-i}p^i \begin{pmatrix} k+m \\ i \end{pmatrix} \tag{3.9}$$

A packet is likely to fail to reach the destination unless it succeeds during the transmission at each hop. Given the probability that a packet will be dropped in one-hop transmission P_{e_3}, it is easy to get the aggregate throughput of the network using header error coding:

$$A_3 = c\sqrt{\frac{n}{\log n}}(1 - P_{e_3})^h = c\sqrt{\frac{n}{\log n}}(1 - P_{e_3})^{\sqrt{\frac{n}{\log n}}} \tag{3.10}$$

We put the FEC protection bits to the tail of the packet, because errors tend to be in burst. If we let the redundant bits be faraway from the header, the header and the protection bits are less likely to be corrupt at the same time.

The efficiency of coding requires the information message to be as small as possible. On the other hand, the more redundancy bits added, the more reliable the transmission would be. The question is how many bytes exactly we would encode. Considering IP header is 20 bytes, we now suppose 30 bytes are to be protected by error detection or correction, since there are important information in headers from other layers as well. This header protection configuration can be adapted to different applications. For binary BCH codes, we choose codes that satisfy block length of $k + m = 255$ bits, $k = 247$ bits, and $t = 1$ bit. This combination is the closest to 30 bytes (240 bits) header. We then use 8 error correction bits to correct 1 bit error for 247 bits. So 1 byte extra can protect 30 bytes of header. Substituting these numbers in (3.9), we have $P_{e_3} = 2.9884 \times 10^{-4}$ and 3.0578×10^{-6} with $p = 10^{-4}$ and 10^{-5}, respectively. That means to protect the header that is no longer than 30 bytes, 1 byte is enough. Also since P_{e_3} is so small, A_3 in Eqn(3.10) could be seen as asymptotically approaching to $c\sqrt{\frac{n}{\log n}}$ when $n \to \infty$.

3.3.5 Comparison of the Effective Throughput of These Schemes

By comparing P_{e_1} with P_{e_2} and P_{e_3}, it is quite clear the latter two have much lower values thus the proposed HEP schemes have the advantage in terms of the effective throughput. The two ARQ schemes - packet CRC and header CRC - also have the same form of throughput expressions, with different factors that depend on the packet error rate. For a clear illustration, we use MATLAB to generate numerical simulation plots of the throughput with an increasing number of nodes for these four different schemes using the throughput expressions derived above. Factor c in the y-axis is the same as that in the above equations.

The parameters used in these plots are: $p = 5 \times 10^{-5}$, payload $q = 500 \times 8$

bits, header $k = 240$ bits, error correction bits $m = 8$, CRC bits $r = 8$, and error correcting capability $t = 1$ bit. Some intermediary results are: $P_{e_1} = 0.1910, P_{e_2} = 1.5217 \times 10^{-4}, P_{e_3} = 7.5634 \times 10^{-5}$. P_{e_3} used here is the upper bound number, i.e., the worst case scenario.

Figure 3.3 show the average per-node throughput as a function of the number of nodes in the network. Our analytical results are valid for large networks. We choose the amount of nodes from 50 to 100 in this plot also to meet the practical needs. Curves for header CRC and header FEC almost lap over each other. The difference between the performance of these two HEP schemes and 802.11 indicates that the gap between the not-so-good performance of the current protocols and the theoretical results can be reduced using header error control. The curves for header FEC and header CRC demonstrate a better scale property than the packet CRC scheme used in 802.11. This result may help in the design of different protocol stacks according to different requirements. For applications having high requirements for data rate and less requirements for accuracy of data, the header error control is especially helpful. We will discuss how to choose from header FEC and header CRC in later part.

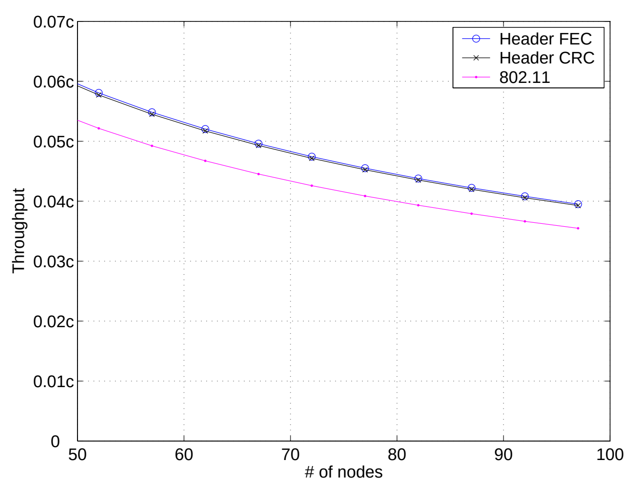


Figure 3.3: Per-node Throughput as a Function of n for Different Schemes

3.4 Simulations [2]

In this section, we evaluate our proposed header error protection schemes by multimedia simulations. The network simulator we use is *ns2* plus wireless extension [58]. We build new protocol models based on our proposed schemes and integrate them into *ns2*. We use the default packet retry limits - 4 for long packets and 7 for short packets - in both 802.11 and our proposed HEP protocols. We use the real-time transport protocol (RTP) over UDP/IP with disabled checksum at UDP segment (or we can call it an extreme case of UDPLite). The packet size used in all simulations is 500 bytes. In addition, all of our simulations use 2Mbps radio.

3.4.1 Ns2 Simulations

In order to model a scenario that is closer to reality, we simulate our protocol and 802.11 protocol on a random network. *Random network* is defined in section 3.3. Nodes are placed uniformly at random in a square domain, and the traffic pattern is random in this network. In *ns2*, the default setting of antenna parameters results in an effective transmission range of 250 meters. The average node density is set to 75 nodes per square kilometer to guarantee the connectivity of the network. In order to get the capacity of the network, we let each node send packets to a randomly chosen destination. The constant RTP rate of the traffic is chosen in order to place the network in a saturation state. In this state there is some slight packet loss and if RTP rate is increased the network aggregate throughput will not increase statistically.

We simulate random networks scaled from 50 nodes to 100 nodes under 802.11 protocol, header CRC protocol, and header FEC protocol with a Markov channel model. Figure 3.4 gives an example of random network node placement with 100 nodes. The parameters for the Markov model are $p_{01} = 2.5 \times 10^{-5}, p_{10} = 0.5$, which yield an average channel BER 5×10^{-5}. The duration of each simulation is 2 minutes and the result is averaged upon 200 runs for different node distributions. The average per-node throughput is shown in Figure 3.5. The simulation results reflect statistically significant analysis based on a 95% confidence interval. If we

[2] This part was done jointly with Yufeng Shan.

compare Figure 3.5 with Figure 3.3, we see they share the same decreasing trend. The sharper decrease in the simulation results indicate the inefficiency of the MAC scheduling. When network scales, the distributed MAC protocol can hardly give an optimal solution to achieve the theoretical capacity, which was observed in [59]. Nevertheless when nodes are around 100, the throughput improvement by using header CRC or header FEC upon 802.11 is about 18%.

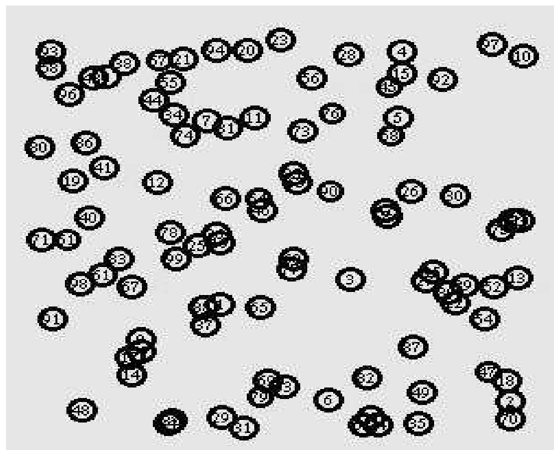


Figure 3.4: Random Network Model Where Nodes Randomly Spread around a Square

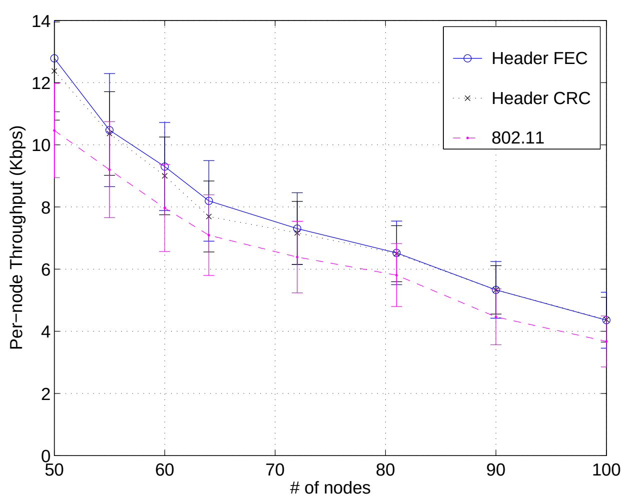


Figure 3.5: Simulation Results on Per-node Throughput of Random Networks as a Function of n for Different Schemes. The Error Bars Show the Data within 95% Confidence Interval around the Mean Value, the Pointed Data.

3.4.2 Video Simulations

Evaluation of the quality of the reconstructed image remains an important issue. We first introduce the standard image quality assessment which we use in our video simulations.

3.4.2.1 Image Quality Computation

Peak Signal to Noise Ratio (PSNR) is a commonly used picture quality measurement [60]. Signal-to-noise (SNR) measures are estimates of the quality of a reconstructed image compared with an original image. The basic idea is to compute a single number that reflects the quality of the reconstructed image. Reconstructed images with higher metrics are judged better. In fact, traditional SNR measures do not equate with human subjective perception. Several research groups are working on perceptual measures, but for now we will use the signal-to-noise measures because they are easier to compute. Just remember that higher measures do not always mean better quality.

The actual metric we will compute is the peak signal-to-reconstructed image measure which is called PSNR. Assume we are given a source image $f(i,j)$ that contains N by N pixels and a reconstructed image $F(i,j)$ where F is reconstructed by decoding the encoded version of $f(i,j)$. Error metrics are computed on the luminance signal only so the pixel values $f(i,j)$ range between black (0) and white (255).

First compute the Mean Squared Error (MSE) of the reconstructed image as follows

$$\mathrm{MSE} = \frac{\sum_i \sum_j [f(i,j) - F(i,j)]^2}{N^2} \tag{3.11}$$

The summation is over all pixels. The Root Mean Squared Error (RMSE) is the square root of MSE.

PSNR in decibels (dB) is computed by using

$$\mathrm{PSNR} = 20 \log_{10} \frac{255}{\mathrm{RMSE}} \tag{3.12}$$

Typical PSNR values range between 20 and 40. They are usually reported to two decimal points (e.g., 25.47). The actual value is not meaningful, but the comparison between two values for different reconstructed images gives one measure of quality.

3.4.2.2 A Single-hop Scenario

We use video experiments to show the different video quality of three schemes, and the difference in throughput capacity each scheme can achieve. We start our video simulation with a network with only 2 nodes. We put 2 nodes 200 meters apart. Due to the significant overhead added by the exchange of RTS/CTS/ACK, when packet size is 500 bytes, the maximum throughput achievable for the 2 nodes under an error-free wireless environment is slightly above 1Mbps.

We test our video simulation using a H.263+ coded bitstream. "Foreman" video sequence is used with 300 frames length, QCIF (Quarter Common Intermediate Format, Quarter CIF) format [60]. Given the maximum throughput achievable for this single-hop scenario, we use multimedia streaming experiments to evaluate the performance of the three schemes. The encoded bit stream is divided into 500 byte packets. These packets are transmitted evenly spaced over approximately 4 milliseconds, thus the data transmission rate is 1Mbps. In this test, different channel conditions are used. First scenario, $p_{01} = 0.5 \times 10^{-5}, p_{10} = 0.5$, then $p_{01} = 1.25 \times 10^{-5}, p_{10} = 0.5$; and last $p_{01} = 2.5 \times 10^{-5}, p_{10} = 0.5$. The average BER for these scenarios are 10^{-5}, 2.5×10^{-5}, and 5×10^{-5}, respectively. The simulation takes 100 runs in each scenario for all the schemes. Note in order to correct the residual bit errors at the receiver for the header CRC and header FEC strategies, we use a rate 1013/1023 BCH code at the application layer. This high rate code has a simple generator polynomial and has little overhead.

The quality of the three schemes are easy to differentiate visually. Header FEC performs best, without obvious discernable artifacts, especially in the first two scenarios. Header CRC is not as good but quite acceptable. 802.11 comes the last and gets much worse with bad channel condition. Figure 3.6 shows some sample frames of the video simulation for the last scenario. The averaged PSNR charts are shown in Figure 3.7. The reason that 802.11 performs poorly as channel condition gets worse is that it has packet losses due to the limited retransmissions. Packet loss is unlikely to be recovered by high rate FEC. In addition, packet loss affects much larger area in a video frame than bit errors, which makes it unaffordable.

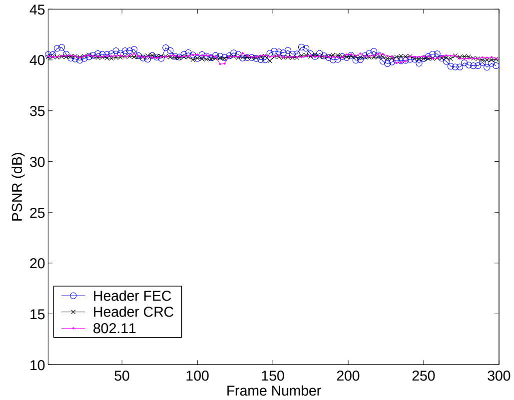
(a) Header FEC (b) Header CRC (c) 802.11

Figure 3.6: Sample Video Frames for Different Schemes

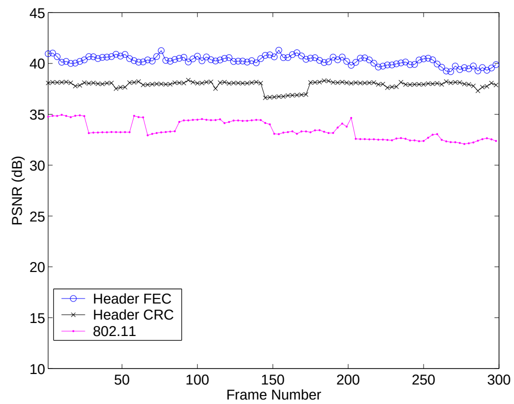
3.4.2.3 A Multi-hop Chain Scenario

The advantage of using header error protection is more obvious in a multi-hop network, since retransmissions increase the traffic load and limit the throughput. In the next set of simulations we intend to find out the effective throughput of multi-hop networks under the three schemes. We use a single traffic chain model to avoid the effect of the interference by other traffics. There are n nodes placed in a straight line, and each neighboring nodes are separated by 200 meters, shown in Figure 3.8. The video traffic is sent from the first node to the last node, traveling through all the intermediate nodes. Parameters of the Markov error model are: $p_{01} = 2.5 \times 10^{-5}, p_{10} = 0.5$. Figure 3.9 illustrates simulation results of the maximum throughput for a single chain. Let *chain length* be the number of nodes in a chain. Each curve represents the flow throughput of different protocols when the chain length n increases from 5 to 10. The curve for 802.11 throughput performance is basically consistent with that in [3] (their throughput is a little bit higher since they do not have error model in *ns2*).

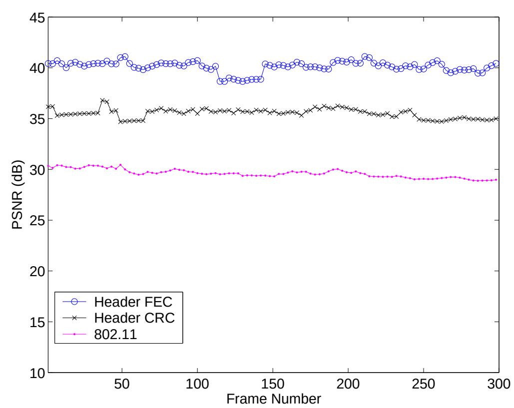
Header CRC performs not as good as header FEC, because headers cannot be recovered by FEC, there are still some packet drops due to too many retries. Even though header CRC and header FEC consume some extra bandwidth for the application FEC overhead, the effective throughput is higher than that of 802.11.



(a) Average BER=10^{-5}



(b) Average BER=2.5×10^{-5}



(c) Average BER=5×10^{-5}

Figure 3.7: PSNR vs. Frame Number for Different Schemes

Simulation results show that there is some potential in the throughput improvement for the header error protection schemes, especially when network gets large and hop number increases.

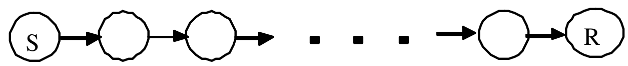


Figure 3.8: A Single Chain with Multi-hops from Sender S to Receiver R

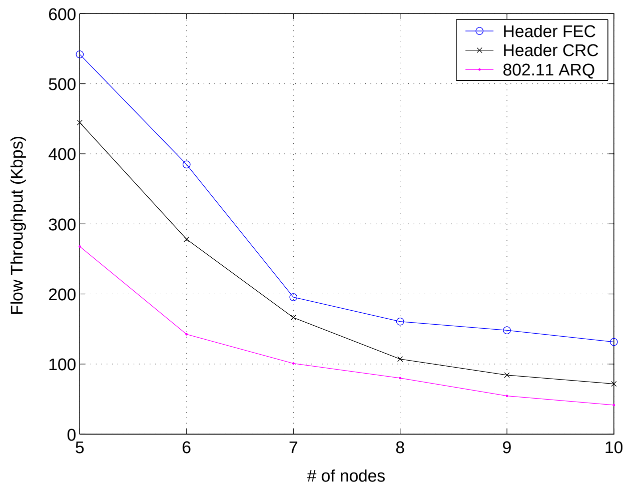


Figure 3.9: Throughput vs. Chain Length n for Different Schemes

End-to-end latency is also an important issue for video applications. We also evaluate the delay performance for this chain multi-hop network. Both the saturated and the unsaturated network performance is evaluated. In the saturation state, nodes are more likely to cope with other transmitting neighbors. For this reason, there will be more packet drops at the interface queue than for the unsaturated network. In the unsaturated state, the traffic load is light, so the collisions are rare. The results from Figure 3.10 and Figure 3.11 show that the average end-to-end latency of the saturated network is in fact slightly less than that of the unsaturated network, because more packets are dropped via collision. More discussion about the saturation is presented via the following simulations. The header FEC strategy is especially superior than the other two schemes in the unsaturated network, which applies to most of the real cases.

In the simulated unsaturated network, we study the trace file of our simulations and find that the packet drops, which may happen at any node, come from two causes: the interface queue overflow and the excess of the retry limit. Both causes are directly related to the link layer retransmission. More retransmissions lead to occupied interface queue. They also cause exceeding the retry limit of data packet or

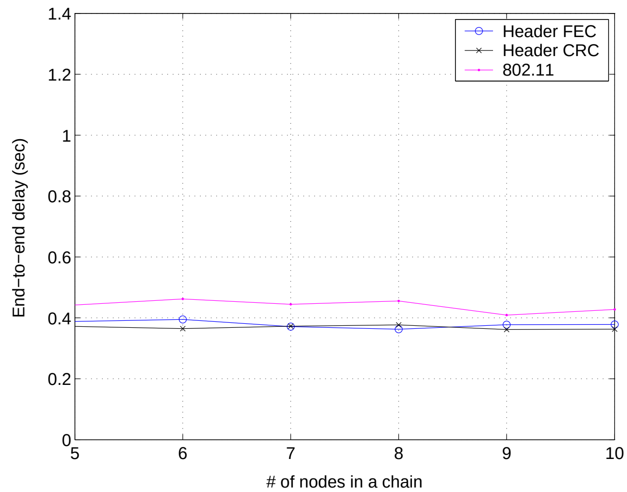


Figure 3.10: End-to-end Latency in a Saturated Network for Different Schemes

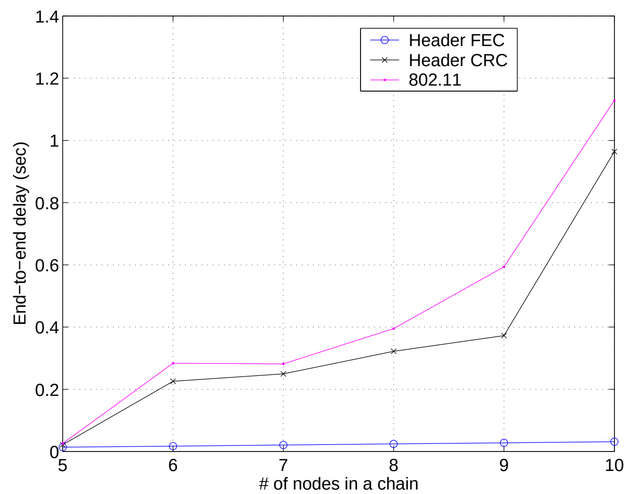


Figure 3.11: End-to-end Latency in an Unsaturated Network for Different Schemes

RTS message. Figure 3.12 shows the packet loss rate of the chain network for three schemes. The packet loss rates are always 0% for the header FEC in this single chain network, which further supports the above statements. The average PSNR for the video streaming versus the chain length is shown in Figure 3.13. The transmission rate is 40bps, which corresponds to the unsaturated network condition. This low data rate results in a relatively low source coding quality. As shown in Figure 3.13, the average PSNRs are lower than those in previous single hop network that uses a much higher data rate. It also shows that header FEC has a static performance on average PSNR with chain length, whereas performance of both header CRC and 802.11 decrease sharply after chain length reaches 8. This indicates that multi-hop scenario exaggerates the video quality discrepancies between those three schemes.

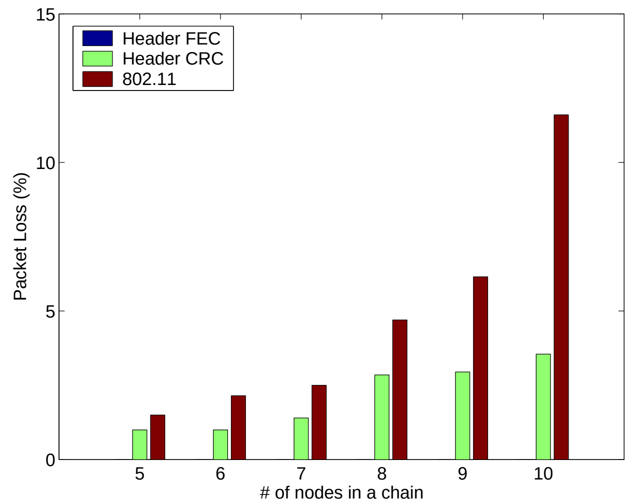


Figure 3.12: Packet Loss vs. Chain Length n for Different Schemes

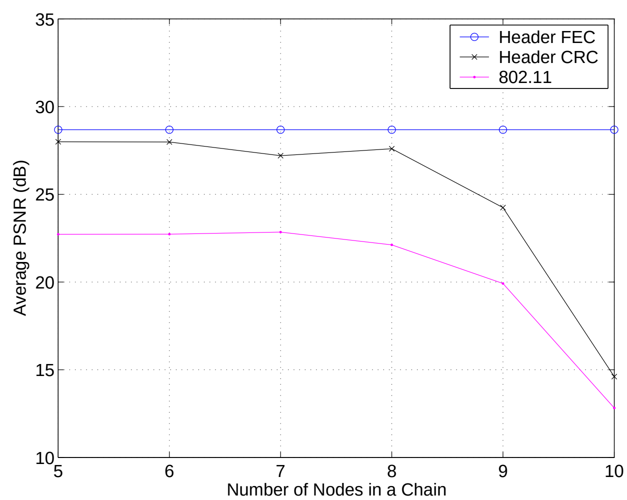


Figure 3.13: Average PSNR vs. Chain Length n for Different Schemes

3.4.2.4 A Multi-hop Chain Topology with Cross Traffic

In the previous chain topology simulation, the network does not have much load caused by contention, since only one single traffic is carried from the source to destination. The only contention is between the nearby relay nodes. In order to get more results for the performance of a multi-hop network, we place more ad hoc nodes to set up a linear topology with cross traffic. In this grid, shown in Figure 3.14, the main traffic is a RTP traffic carried from node S to node R. This node placement is the same as in the chain topology. Besides this main traffic, there are cross traffic from node S_1 to R_1, node S_2 to R_2, ..., and node S_n to R_n, with n as the main chain length. These cross traffic also needs the nodes in the main chain as relays. The distance between each horizontally and vertically adjacent nodes is 200m. Parameters of the Markov error model are: $p_{01} = 1.25 \times 10^{-5}, p_{10} = 0.5$.

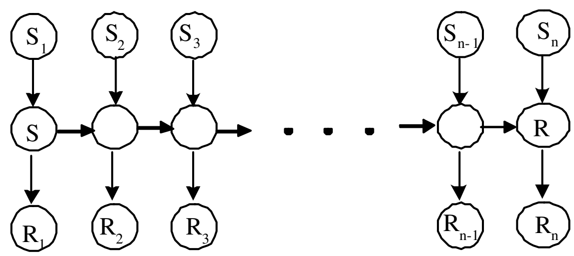


Figure 3.14: A Chain Topology with Cross Traffic

In this chain topology with cross traffic, the source of the main chain together with the relays of this chain have to compete with the nodes transmitting cross traffic. The RTS/CTS/DATA/ACK mechanism requires nodes near the sender or receiver to keep silent. Since the transmission range in our simulation set up is 250m, there will be much contention around each transmission pair. If a sender experiences a collision when sending RTS, it will choose an exponential random backoff time and retransmit after the backoff. The retransmission will continue if it fails, until it reaches the retry limit. It is obvious that the number of retries has impact on the end-to-end delay. It also increases the possibility of the collision for the surrounding nodes, which further increases the delay of each transmission. The proposed header protection strategy can largely reduce the number of retransmissions, thus it has the potential to reduce the end-to-end delay, which is important for a real time application.

Figure 3.15 gives the result for the end-to-end throughput performance of the chain topology with cross traffic. The video is in 40kbps, with the channel error rate

2.5×10^{-5}. The cross traffic is also in a 40kbps rate. It plots the flow throughput of the main chain (middle chain) traffic versus the chain length n. Figure 3.16 shows the delay performance of the main chain for the same setting. In fact, the end-to-end throughput and end-to-end delay performance are closely related. The higher the throughput, the shorter the delay. This is because that the longer delay indicates more retransmission, which leads to more collision and thus more possible packet drops. It is somewhat misleading that the delay performance in this network with heavy traffic is better than that in the previous single chain network (Figure 3.11). The reason of this is that the packet loss rate, which is shown in Figure 3.17, is much higher in this multi-flow network than in the single-flow network. Some packets are dropped instead of surviving after a long time delay. Hence the delay performance itself does not necessarily reflect the video performance.

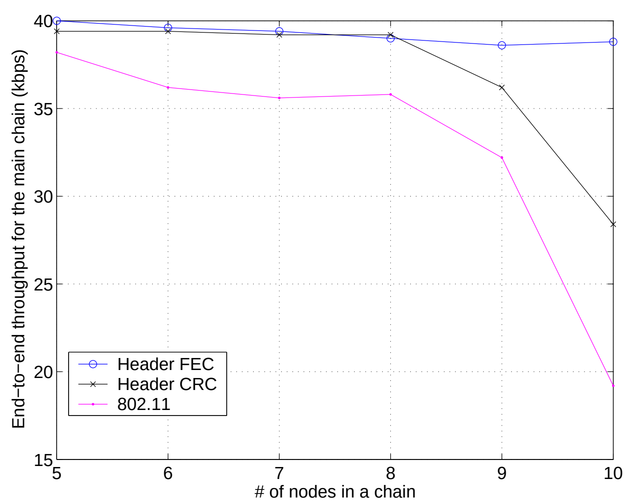


Figure 3.15: End-to-end Throughput vs. Main Chain Length n for Different Schemes

It is worth mentioning that the sending rate for the middle chain source, 40kbps, acts like a threshold of the optimal sending rate for this grid topology with main chain length 10. If the sending rate is increased, the end-to-end throughput will drop. We also regard this threshold as the saturation point. This point depends on the node placement, and the traffic pattern for the whole network. Below this point, the network is in an unsaturate state, and the throughput will be close to the sending rate. If the sending rate exceeds the saturation point, the throughput will

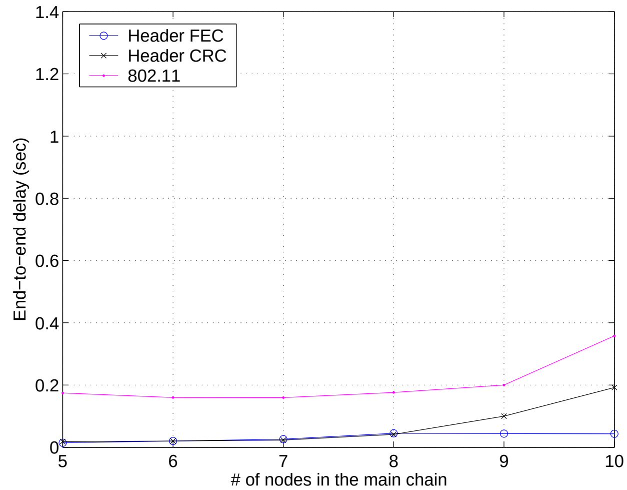


Figure 3.16: End-to-end Latency vs. Main Chain Length n for Different Schemes

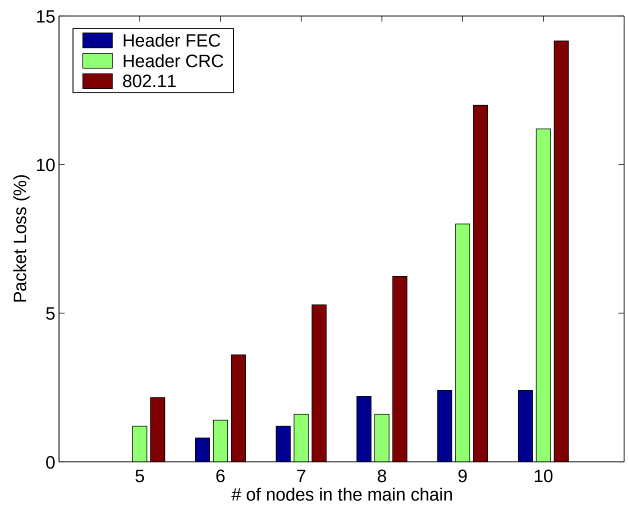


Figure 3.17: Packet Loss vs. Main Chain Length n for Different Schemes

drop, sometimes very sharply. This drop of throughput implies the MAC scheduling inefficiency. That is also to say, 802.11 MAC cannot discover the optimum schedule of transmissions on its own. The schedule of the MAC protocol is sub-optimal, and the fairness is not well addressed. Each node in a network experiences different degree of competition. For example, nodes at the edges of the grid have fewer competitors than those in the middle of the grid. So some bandwidth is wasted by transmitting packets that are eventually dropped at some nodes with higher degree of contention.

Another concern is that the interference range is always larger than the transmission range. Thus a node trying to transmit may not have received RTS or CTS that can inform how long it has to wait, yet it may fail because of the interference. In this case, it chooses a random backoff time and sends again. The next try might fail again because it does not has the information about the interfering nodes. Due to the limited retries and limited interface queue buffer length, too many retransmissions, either from contention or packet CRC check, can make things much worse than expected, especially near the saturation point. That is why the original 802.11 scheme has a much lower performance than the header protection schemes. In any case, the saturation of the network should be avoided.

Figure 3.18 show the average PSNR in dB as a function of number of hops, with different error control approaches used. Due to the introduction of the cross traffic, the degradation of the performance as the number of nodes increases is more evident than the single flow scenario for all three schemes. In this case, HEP, especially header FEC can effectively improve the PSNR values from the 802.11 scheme, thus prevent the video quality from degrading rapidly.

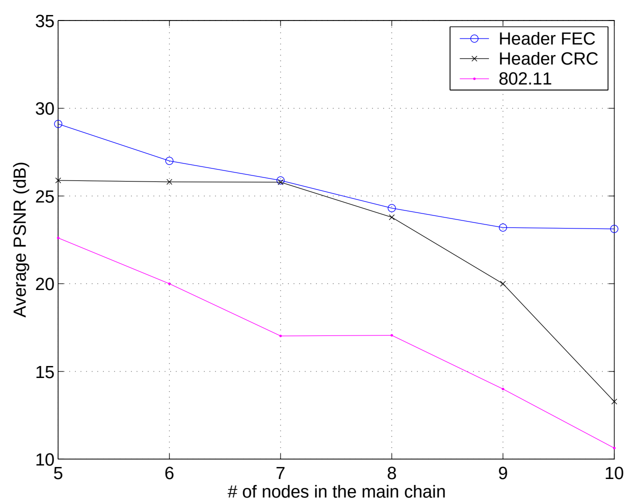


Figure 3.18: Average PSNR vs. Main Chain Length n for Different Schemes

3.5 Design Guidelines

By evaluating the throughput performance, video performance of the different schemes under some specific link conditions, we have come to the conclusion that header error protection works better than the original packet CRC strategy in 802.11. It is valuable to further investigate the working range of the new schemes, and how to choose the right scheme under different environment. In this section, we give out some design guidelines on error control for supporting the multimedia applications.

The major factors that affect the selection of the error control schemes are as follows:

• Channel Error: Channel error rate is the most important and sensitive factor for the video performance. Following the two-state Markov error model, we use the average BER and average burst error length to describe the channel characteristics.

• FEC or CRC: The choice between header error coding and header CRC depends on questions like whether or not the coding and decoding energy is a factor, if the reverse link is desired, etc. Other concerns may affect the choice as well, which include ARQ is better for handling burst errors and header error coding can be adaptive to the link error environment (e.g., if the link error rate increases, the protection bits can be added to correct more errors with little cost).

In the choice of FEC codes, we find BCH a simple and efficient block code for correcting the bit errors occurred in packet header. To keep the complexity of encoding and decoding minimal, BCH codes (255,247,1) and (255,239,2) are of our interests. The latter can correct any errors up to 2 bit length and have a 2 bytes overhead. We suggest that if the average burst error length is equal to or greater than 3 bits, header CRC is preferred.

• Packet Size: Packet size is also an important factor for considering the error detection or correction problem. Given the channel BER and the packet size, the packet error rate can be simply approximated. Packet error rate determines the packet drop probability of each error control scheme. If the packet length is too long, then the packet error rate is high, where conventional ARQ is not effective.

Since channel condition and packet size should be jointly considered, Table 3.1

gives suggested error control schemes based on the channel average BER and packet length values. In this table, we assume the average burst error length is less than 3 bits. In a low error rate channel, the difference between the header error protection and 802.11 is negligible, so any of them can be chosen. When channel error rate arises to above 10^{-4}, neither ARQ or header error protection can work well for large packet size. In this case, FEC codes need to be added to the packet to cover the whole packet.

Average BER	Pkt size $\leq$100 bytes	Pkt size 500 bytes	Pkt size 1000 bytes	Pkt size $\geq$1500 bytes
$\leq 10^{-6}$	any	any	any	any
10^{-5}	any	header FEC or CRC	header FEC	header FEC
$\geq 10^{-4}$	header FEC	header FEC	FEC for whole pkt	FEC for whole pkt

Table 3.1: Guidelines on Choosing Error Control Schemes

3.6 Summary

This thesis proposes two header error protection schemes, header CRC and header FEC, in order to give a solution to improve the performance of multimedia transmissions. Both header CRC and header FEC only need insignificant changes at the packet header if 802.11 is kept as the MAC protocol. Network simulation results show that under a random network scenario header error protection takes advantage of FEC or ARQ to reduce the number of dropped packets at relaying nodes, thus can improve the throughput of the network. We also examine their video streaming performances together with 802.11 protocol under a single hop network, a multi-hop chain and a cross traffic abundant multi-hop chain network. They present better qualities than 802.11 does in terms of the visual effects observed by experimenters and the PSNR results. Packet losses induced by bit error checking not only impair the video quality but also diminish the maximum throughput a network can achieve.

Since the link layer does not perform any error correction or detection for the whole packet, the payload error at the destination may be higher than the acceptable limit. Therefore, we propose to use end-to-end error control coding for the application layer, wherever it is needed. Application layer FEC is needed not only because of the channel errors, but also because of the packet losses caused by congestion. This is one of the reasons we propose not to do local error protection for the whole packet. Whether to do end-to-end error control coding or not, and how efficient the codes should be depend on the requirement of the application.

A cross layer approach that integrates application and link layer should be considered, such as how to protect data information according to the priority of classes of the data, and choose FEC packets adaptively to the application requirement, etc. Towards this direction, in a recent work, we propose a application layer two-stage FEC scheme for scalable video communication over wireless LANs [61]. The proposed scheme enables the joint optimization of protection strategies across the protocol stack, and packets with errors are delivered to the application layer for correction or drop. In stage 1, packet-level FEC is added across packets at the application layer to correct packet losses due to congestion and route disruption. In stage 2, bit-level FEC is processed within both application packets and stage-one FEC packets to recover from bit errors in the MAC/PHY layer. Header CRC/FEC are also used to enhance the MAC/PHY layer and to cooperate with the two stage FEC scheme. Thus, we add FEC for application messages only at the application layer, but can correct both application layer packet drops and MAC/PHY layer bit errors. We explore both the efficiency of bandwidth utilization and video performance using the scalable video coder MC-EZBC and *ns2* simulations. Simulation results show that the proposed scheme outperforms conventional IEEE 802.11.

CHAPTER 4
Cluster-based Cooperative Wireless Networks [3]

4.1 Introduction

This thesis aims towards adaptively using a wide variety of distributed cooperation techniques in wireless multi-hop ad hoc or sensor networks. These techniques are intended to improve overall system throughput, reduce the cost of node elements, and extend the units' service lives. To this end, tools from multi-disciplinary areas of signal processing, transmit and receiver diversity, error correction and detection (FEC), cluster-based forwarding and routing, are to be employed.

4.1.1 Overview

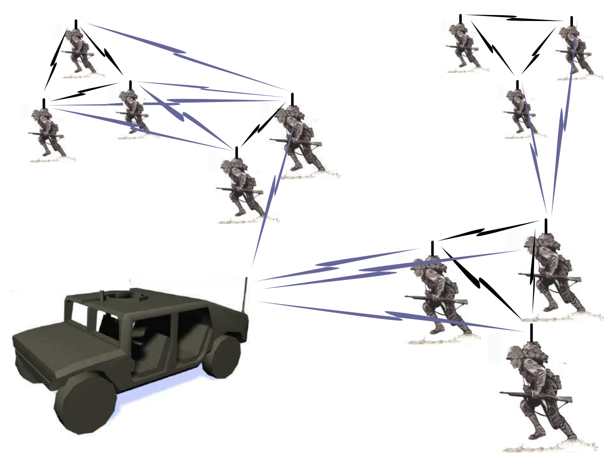
This work proposes to extract diversity gain out of the redundancy inherently present in all broadcast network transmission, such as wireless sensor networks, and direct those gains for chosen receiver nodes. The redundancy in such systems is present since the signal carried over such a channel is received (if not necessarily detected) by all nodes within transmission radius. Thus, in this distributed cooperative paradigm, packets are not relayed from one network node to the next, but from one cluster of nodes to the next cluster of nodes, until it reaches its destination. Grouping the network nodes into collaborating groups enables the system to gain from cooperative diversity, cooperative error recovery, network-layer cluster auto-configuration and cooperative cluster-based routing.

This research strives to design cooperation techniques that permit the nodes for such networks to be as simple, small, flexible, long-lasting, and affordable as possible. This work addresses a wide variety of cooperation techniques in all physical, link, and network layers for wireless multi-hop networks.

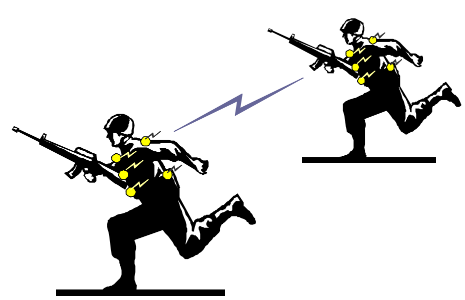
Multi-hop ad hoc networks or sensor networks are increasingly essential for commercial infrastructures, military settings, crisis monitoring, and public safety.

[3] The work on the physical layer cooperation was previously done by Prof. Babak Azimi-Sadjadi and Prof. Alejandra Mercado

Our objective in this thesis is to make these networks more powerful by increasing their overall throughput, service capability, and individual unit flexibility. This work provides further benefits by improving the power efficiency of the network nodes, which implies smaller, more affordable units, permitting the deployment of more units. For example, consider Figure 4.1(a) where soldiers in small groups are separated from each other. The distance between soldiers in each group is small enough to allow short range communication link between them. On the other hand the communication between different groups of soldiers requires considerable energy. An automatically configurable communication system detects the communication units capable of cooperation and sets up an optimum cooperating scheme (for example cooperation in physical layer or link layer) to reduce the transmission power and enhance the communication channel.



(a) The automatically configurable cooperative communication system detects the communication devices of the nearby soldiers. In this figure each group of soldiers use a short distance radio among themselves to share their information and cooperate to reduce the long range radio transmission power.



(b) As a simple example, one soldier with multiple antennas collaborating in a cluster communicates with another soldier with a cluster of antennas.

Figure 4.1: Soldiers with Sensor Nodes

The same idea can be applied to the case with a natural cluster in a soldier's suit (e.g.: multiple antennas per suit), shown in Figure 4.1(b). The cluster formation will be simplified, i.e., all the working antennas of a suit can collaborate. We could also have the option of two clusters - i.e. two soldiers' suits coming together - rather than arbitrary clustering. Let us assume a soldier will have multiple antennas on his front and back - even if some fail, others can cooperate - giving a new advantage: resilience to failures of individual antennas. GPS on the soldier plus simple location systems on each individual node on the soldier would simplify the auto-configuration without needing each node to have GPS.

Such a system will benefit from cooperation by the following means. For the same bit error rate, the total transmission power among all nodes is reduced. Equivalently, for the same total power the bit error rate is reduced, hence the throughput is increased. It also has the benefit of lowering ambience interference levels.

The individual node transmission power is further reduced because the burden of transmission is divided among all cooperating nodes, thus extending the nodes' battery lives. This increases the sustainable number of active users in the network which makes the service more affordable.

The cooperating nodes can act as *virtual* antenna array. This diversity inherently provides protection against channel variations and transmitter/receiver failures.

For mobile networks, the smaller and lighter cooperating nodes are easier to mobilize.

Cooperation increases the connectivity of the network. For example, a group of sensor nodes that are isolated cannot individually reach the nearest connected node even if each node uses its maximum transmission power. By cooperating they can increase their transmission range, hence increase the connectivity of the network.

Trajectory Routing offers flexibility and improved reliability. Since the sensor nodes are low-power and low-memory devices, it may not be practical for them to store routing tables. This routing protocol combines location with clustering information, and it also makes implementing traffic engineering possible.

Nodes which cooperate in this distributed way can use additional bandwidth that is otherwise unavailable, for the purpose of improving signal quality at the receivers. For example, consider the case of smart suits, which collect a variety of information from the wearer and transmit it to a monitoring base. Rather than having a single transmitter on the suit collecting the data and relaying it to the base, several distributed transmitters can use bluetooth technology (using a distinct bandwidth and extremely low powers) to share their data among themselves. Then they can cooperate to relay the information to the base with a reduced BER. This precludes wires (which are heavy and may be severed), reduces the transmitter size and weight, reduces the power supply size, and increases reliability (in the sense that a transmitter may be disabled without silencing the suit).

Under a cluster-based network design, code combining is used together with FEC to improve the link layer reliability, without increasing power consumption.

In some sensor network set ups, the power level at each sensor is fixed and unadjustable. It is observed that in large area fields, the received packets may always be corrupt for some receiving nodes which are in a critical distance away. This is due to the fact that at physical layer, the weak signal received is distorted by noise. In this case, the effectiveness of retransmission is very minimal. Because of the multi-path effect, not only the packets have too many errors in it, these errors exhibit independent features in terms of location and distribution. Letting nodes located in a nearby neighborhood help each other is a proper way to solve the problem of unreliable links and it is the major inspiration for the cooperation design.

4.1.2 Physical Layer Cooperation

We consider a large collection of autonomous nodes or terminals that communicate with each other by forming a multi-hop wireless network and maintaining connectivity in a decentralized manner, as depicted in Figure 4.2. Any pair of connected nodes are linked through a multipath fading channel, which reflects the effects of path loss, slow and fast fading, and interference from other nodes. Cooperation among nodes can be done in different communication layers. Figure 4.2 shows cooperation in the physical layer and in the link layer.

In the physical layer, cooperative nodes share their information to improve the

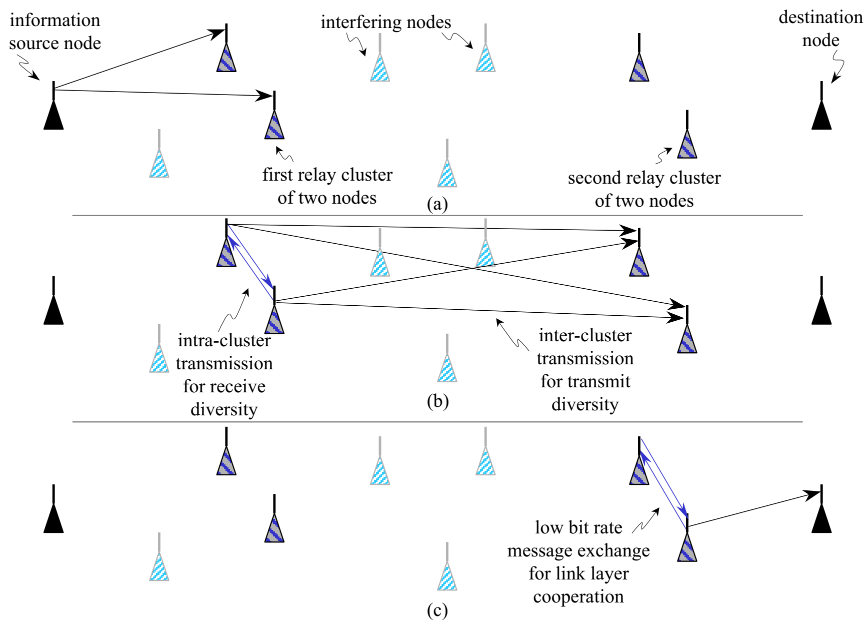


Figure 4.2: Transmitting Nodes Group into Cooperative Clusters to Relay the Information from the Source to the Destination [62]. (a) The information source reaches the first relay cluster. (b) The nodes in the relay cluster share their information for diversity gain. Then they relay the information to the next cluster. (c) The next cluster has a reliable channel with the destination node, hence there is no need of physical layer cooperation. A single node can relay the information to the final destination node.

channel quality using transmit and/or receive diversity (Figure 4.2(a) and 4.2(b)). Physical layer cooperation has been studied recently under the subject name of "cooperative diversity." In cooperative diversity the transmitting nodes use the nodes in the neighborhood of the transmitter and the receiver as relays [63, 64, 65, 66, 67, 68, 69, 70], active scatterers [71], or simply clusters of cooperating nodes [72, 62], to reduce the adverse effect of multipath fading in the wireless channel.

Different from most of the previous works where only cooperative transmit diversity is studied, Mercado *et al.* [72, 62] propose a diversity technique that is capable of implementing distributed cooperative receive diversity as well as distributed cooperative transmit diversity. This automatically translates to distributed cooperative MIMO diversity which can be used for em signal transmission under extreme power constraints such as those of sensor networks. They have shown that by using a CDMA platform and a coarse synchronization among cooperating transmitters, MIMO cooperative diversity is possible.

4.1.3 Link Layer Cooperation

In this chapter we present a new link layer cooperation scheme for multi-hop wireless networks and sensor networks to improve the overall channel quality for each transmitter/receiver pair [73]. Unlike physical layer cooperation, we take a different approach and we use cooperation in the link layer. If the SNR of the received signal is moderately high, one can avoid physical layer cooperation to save on the bandwidth used for information sharing and synchronization [72, 62] and instead use the link layer cooperation to increase the overall throughput of the network. In the link layer cooperative transmission the cooperating nodes decode the received packets (instead of the individual bits/symbols done in the physical layer cooperation) and participate in the cooperative transmission of the error-free packets. The link layer cooperation can be implemented in following stages depending on the quality of the link:

Stage 1: Cluster head decides if cooperation is necessary. Unlike the node to node cooperative cluster transmission, a packet is successfully received if at least one node in the cluster receives the packet without error. The nodes with the error free packet send their status to the cluster head using a low bit rate message. The cluster head chooses one of the nodes with the error free packet to forward that packet to the next cluster.

Stage 2: FEC and Code combining among cluster nodes. If no node receives the packet successfully, the cooperating nodes can combine their erroneous packets and use code combining techniques [38, 39, 40] to reconstruct the packet. FEC can be designed over the entire frame to facilitate code combining.

Stage 3: ARQ or transmit diversity. If the reconstruction is unsuccessful the master node sends an ARQ to the previous cluster for the packet retransmission. Or, if the forward channel quality is too low for any individual transmitter, the master node can recruit several transmitting nodes to use cooperative transmit diversity. If the number of nodes with error free packets is not enough to satisfy the desired BER the cooperating nodes share their data with other nodes in the cluster to recruit them for cooperation.

The main technique in this link layer cooperation is the use of the well-known code combining. In the conventional type I hybrid ARQ scheme with code combining, the repeated packets are sent upon each request [27]. This retransmission based method can be considered a redundancy in time. In our new cooperative link layer paradigm, retransmission can be greatly reduced or avoided by making use of the wireless broadcast nature. In fact, the retransmission is replaced by information sharing among the nodes in the receiving cluster. In other words, we use the existing parallel channels between the transmitting node and the receiving nodes for code combining. This can be called redundancy in space. This method is well-suited for interactive real-time communication streams where waiting for retransmission introduces unacceptable delay and jitter. However, the cost for the node cooperation is the extra power and bandwidth used for the intra-cluster communication.

4.1.4 Network Layer Cooperation

The concept of cooperative diversity can be used to consider an interesting paradigm shift in layer-3 forwarding and routing: "cooperative cluster-based routing." Traditionally forwarding of packets at the network layer happens from node to node. There has been a copious amount of research into various aspects of ad-hoc wireless network routing, especially routing scalability [74, 75, 76, 77, 78, 79], geographic and trajectory routing [80, 81, 82, 83, 84, 85]. The key difference in our model is that packets will be forwarded ad-hoc on a cluster-to-cluster basis and not on a node-by-node basis. Moreover, we propose to give the source freedom to explicitly choose source-routes (eg: as an extension of DSR [86]) or trajectory routes (eg: as in location-driven or geographic routing) around which the cooperative cluster-based forwarding is performed. Figure 4.3 illustrates the cooperative cluster-based forwarding ideas. The figure shows a source that can explicitly pick a path along the trajectory and form a cluster as its first relay cluster.

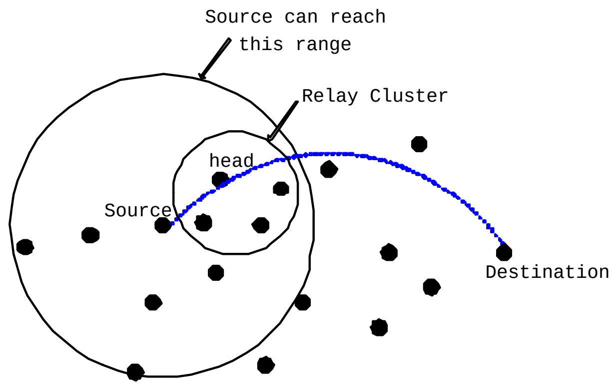


Figure 4.3: Routing Example: Forming the Cluster along the Trajectory.

An interesting aspect of our proposed method is that these forwarding cooperative clusters are formed on the fly. In particular, the source (or the previous-cluster-hop) can actually encode a cluster boundary into the packet header.

There is an important difference between our proposed forwarding protocol and other routing protocols that introduce the notion of a "cluster" (eg: [87, 88]). Our use of clusters is for enabling cooperative diversity and cooperative error recovery (at layer 1 and layer 2 respectively). In the conventional routing schemes, "clusters" are used for routing scalability, i.e. to facilitate hierarchial routing (just like "areas" in Open Shortest Path First (OSPF) routing protocol [89] or "peer groups" in Private Network-to-Network Interface (PNNI) [**?**]).

In this section we just sketch the concept and mechanics of our cooperative cluster-based forwarding concept. We discuss this concept assuming that addresses of nodes are location-based (e.g.: using GPS or GPS-free location techniques [90, 85, 91, 92, 93, 94, 95, 96]). The purpose of using a location-based address is to facilitate auto-configuration and to simplify and scale the routing protocol. We propose auto-configuration of the entire cooperative clustering protocol stack based upon location-based identifiers (e.g.: GPS or other GPS-free location IDs). Besides auto-enabling the dramatic power savings due to cooperative diversity techniques, this network-layer forwarding model allows flexibility in path choices, cluster sizing, increased reachability, increased delivery probabilities and increased network lifetimes even in sparsely connected or non-uniformly spread (eg: due to nodes shutting down to save power) sensor networks.

The rest of this chapter focuses on link layer cooperation and is organized as follows: The performance analysis for the link layer cooperation is given in Section 4.2. In Section 4.3 we present our simulations and results and in Section 4.4 we give our concluding remarks and we lay out future work.

4.2 Performance Analysis for Link Layer Cooperation

4.2.1 Preliminaries

We now assume nodes are already clustered using some existing clustering protocol, like LEACH [97] and there are enough nodes in one cluster to cooperate. The packets in each cooperative node will be sent to the cluster head for code combining. So the number of repeated packets is identical to the number of cooperative nodes. Throughout the whole chapter, L represents the number of nodes joining the cooperation. This is equivalent to the repeated packets in code combining.

In the cooperative cluster, the member nodes will transmit their received packets to the cluster head if necessary. The distance between the nodes in the cluster is much smaller than the distance between the transmitter and the receiver from different clusters. Therefore, the required intra-cluster transmission power is much smaller than the power of the inter-cluster transmission. In general the bit error rate for inter-cluster channel and intra-cluster channels are different. Let p_1 and p_0 be the bit error rate for the inter-cluster channel and intra-cluster channel, respectively. Therefore, a single bit traveling from the source to the cluster head via a member node, has the bit error probability equal to $p = p_1 + p_0 - p_1 p_0$.

How the code combining technique is used among the cooperation procedure can be illustrated in Figure 4.4.

Code combining [38] represents a technique for combining L repeated packets encoded with a code of rate R to obtain a lower rate, R/L, and thus more powerful, error-correcting code, capable of allowing more channel errors. Code combining is designed to work in a very noisy environment, where conventional diversity combining concepts [98, 99, 100] can easily break down. One feature of code combining is that the maximum-likelihood (ML) decoder will select the codeword m which maxi-

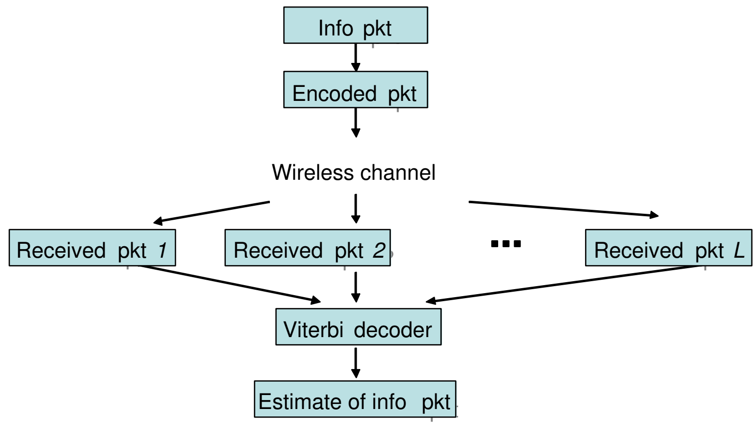


Figure 4.4: Block Diagram of Link Layer Cooperation with Code Combining Technique

mizes the conditional probability between the received sequence $\mathbf{r}$ and the repeated codeword denoted by $\mathbf{v}_m$. Repeated codewords are transmitted over BSC channels with bit error rate p_i for $i = 1, 2, \cdots, L$. The decoding function can be written as

$$\max_m \left\{ \mathbf{P}[\mathbf{r}|\mathbf{v}_m] = \prod_{i=1}^{L} (1 - p_i)^{N - d_{mi}} p_i^{d_{mi}} \right\} \tag{4.1}$$

where d_{mi} is the number of bit disagreements for the ith codeword, and N is the pre-combined codeword length. An alternate way to write (4.1) is

$$\min_m \sum_{i=1}^{L} w_i d_{mi} \tag{4.2}$$

where weight (reliability factor) $w_i = \log \frac{1-p_i}{p_i}$.

4.2.2 Code Combining with Block Codes

To show the performance of code combining, it is easier to start with block codes. We use Golay code with maximum-likelihood decoding as an example. Suppose the original code is (24,12), a rate 1/2 block code.

The minimum Hamming distance d_{min} of the new set of codewords will increase linearly with the number of repeats. Error correcting capability can be given as $t = \lfloor \frac{d_{min}-1}{2} \rfloor$. The minimum distance and the error correcting capability of the repeated Golay code is listed in Table 4.1.

number of repeats L	code dimensions (Ln, k)	minimum distance d_{min}	error correcting capability t
1	(24,12)	8	3
2	(48,12)	16	7
3	(72,12)	24	11
4	(96,12)	32	15
8	(192,12)	64	31
16	(384,12)	128	63

Table 4.1: Minimum Distance of a Repeated Golay Code and its Algebraic Error-Correction Capabilities

In general, for a binary symmetric channel (BSC) with and error probability of p_1 we can define the random variable

$$U_i^1 = \begin{cases} 1, & \text{with probability } p_1 \\ 0, & \text{with probability } 1 - p_1. \end{cases}$$

The number of errors in a codeword of length n is

$$\mu_1 = \sum_{i=1}^{n} U_i^1 \tag{4.3}$$

We define the random variable ν_1 as

$$\nu_1 = \frac{\mu_1}{n} \tag{4.4}$$

So random variable ν_1 has a mean of p_1 and a variance of $(1 - p_1)p_1/n$. When $n \to \infty$, $\nu_1 \sim N(p_1, \frac{(1-p_1)p_1}{n})$.

In the cooperative FEC, the member nodes will transmit their received packet to the cluster head if necessary. Therefore all the repeated packets will travel from the sender to the cooperative node and then to the cluster head, except that the packets directly received by cluster head, only have to travel from sender to the cluster head. Thus among L repeated packets, one of them has error probability p_1, and the other $L - 1$ have error probability $p = p_1 + p_0 - p_1 p_0$. Similarly with U_i, we define another Bernoulli trail as

$$U_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p. \end{cases}$$

There are $(L-1)n$ bits in the packets transmitted from all the member nodes. The head node has it's own packet with n bits. Let $\mu = \sum_{i=1}^{(L-1)n} U_i$, and $\nu = \frac{\mu}{(L-1)n}$. So $\nu \sim N(p, \frac{(1-p)p}{(L-1)n})$.

Now let's calculate the decode error rate for the original codes and the combined codes. Golay code is a perfect code, so for a combined code capable of correcting t errors in the combined packet, the probability of decoded error (word error rate) is

$$p(e) \quad = \quad \mathbf{P}[\mu_1 + \mu > t] = \mathbf{P}[n\nu_1 + (L-1)n\nu > t] \tag{4.5}$$

Let $\theta = n\nu_1 + (L-1)n\nu$. The linear function of independent Gaussian random variables is still Gaussian.

$$\mathbf{E}[\theta] = n\mathbf{E}[\nu_1] + (L-1)n\mathbf{E}[\nu] = np_1 + (L-1)n(p_1 + p_0 - p_1p_0) \tag{4.6}$$

$$\begin{aligned} \mathbf{Var}[\theta] &= n^2\mathbf{Var}[\nu_1] + (L-1)^2n^2\mathbf{Var}[\nu] \\ &= np_1(1-p_1) + (L-1)n(1-p_1)(1-p_0)(p_1+p_0-p_1p_0) \end{aligned} \tag{4.7}$$

Therefore $\theta \sim N(np_1 + (L-1)n(p_1+p_0-p_1p_0), np_1(1-p_1) + (L-1)n(1-p_1)(1-p_0)(p_1+p_0-p_1p_0))$. So

$$\begin{aligned} p(e) &= \mathbf{P}[\theta > t] = Q(\frac{t-\mathbf{E}[\theta]}{\sqrt{\mathbf{Var}[\theta]}}) \\ &= Q(\frac{t - np_1 - (L-1)n(p_1+p_0-p_1p_0)}{\sqrt{np_1(1-p_1) + (L-1)n(1-p_1)(1-p_0)(p_1+p_0-p_1p_0)}}) \end{aligned} \tag{4.8}$$

where $n = 24$ in this example and the function $Q(x)$ is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \tag{4.9}$$

Fig.4.5 plots the decoded error rate as a function of the number of cooperative nodes, L, under different channel error rates. In this example, channel error rate around 8.5% is a threshold that code combining can work well or not. If the channel error rate is greater than 8.5%, code combining can not handle the errors any more. Let's take a look at the algebraic error-correction ratio t/Ln of the combined code.

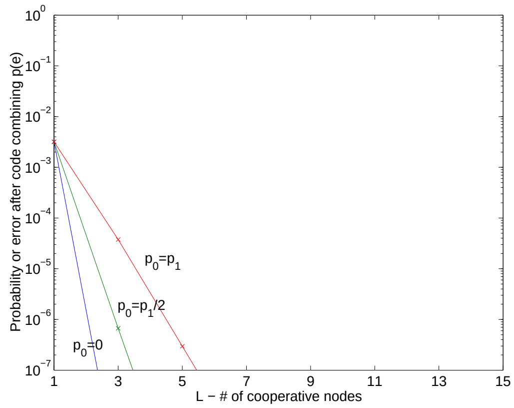
$$\frac{t}{Ln} \approx \frac{d_{min}}{2Ln} = \frac{Ld^o_{min}}{2Ln} = \frac{d^o_{min}}{2n} \tag{4.10}$$

where d^o_{min} corresponds to the minimum distance of the original code (no combining).

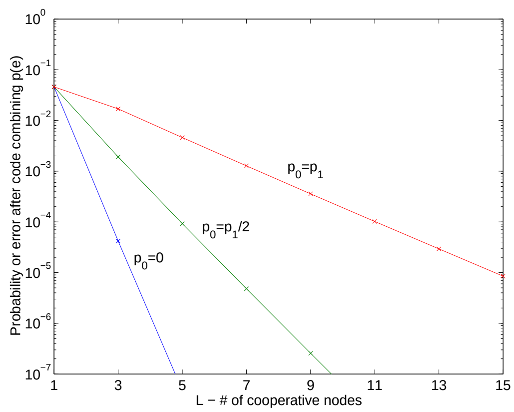
So this ratio will depend on the code chosen. In this example, $d^o_{min} = 8, n = 24$, so the algebraic error-correction ratio is roughly 1/6(16.7%). Our threshold on p_1 is lower than this due to the error accumulation at the relaying link ($p \approx 2p_1 = 17\%$). If the channel condition is better than the threshold (8.5%), cooperation among nodes can reduce the decode error rate greatly and thus avoid retransmission. To our knowledge, 8.5% bit error rate is generally far above the ordinary wireless channel. Depending on the channel error rate of noisy channel, the choice of the number of cooperative nodes can be made according to the performance prediction.

4.2.3 Code Combining with Convolutional Codes in a Uniform Channel Condition

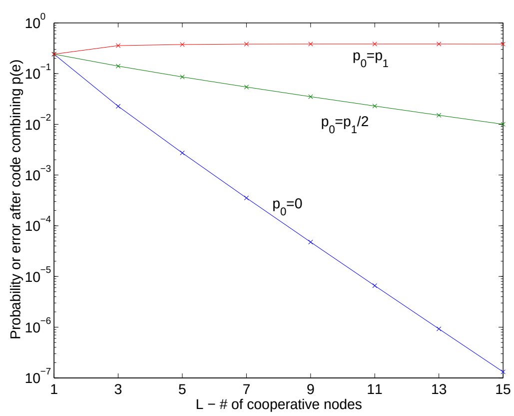
It can be seen from (4.1) that if a block code is used for code combining, the complexity of the decoder depends greatly on the number of codewords, which increases exponentially with the codeword length n. Therefore, to reduce the decoding complexity, we want the codeword length to be small. This will limit the use of block codes, since block codes are efficient in large blocks. For this reason, code combining is generally used for convolutional codes or for short block codes. Due to the large amount of redundancy short block codes deliver for data networks, we propose convolutional codes for the cooperative FEC in this thesis. For the



(a) $p_1 = 3\%$



(b) $p_1 = 5\%$



(c) $p_1 = 8.5\%$

Figure 4.5: Decoded Word Error Rate $p(e)$ vs. Number of Cooperative Nodes L, where $L = 1$ Corresponds to the Scenario with No Code Combining

rest of this section we analyze the performance of the code combining technique for convolutional codes. We adopt the notation used in [19].

Let's use a rate 1/3 nonsystematic feedforward convolutional encoder with memory order $m = 2$ as example. The block diagram of the encoder is shown in Figure 4.6. This encoder consists of $k = 1$ shift register with $m = 2$ delay elements and with $n = 3$ modulo-2 adders. The modulo-2 adders can be implemented as Exclusive-OR (XOR) gates. Since modulo-2 addition is a linear operation, the encoder is a linear system.

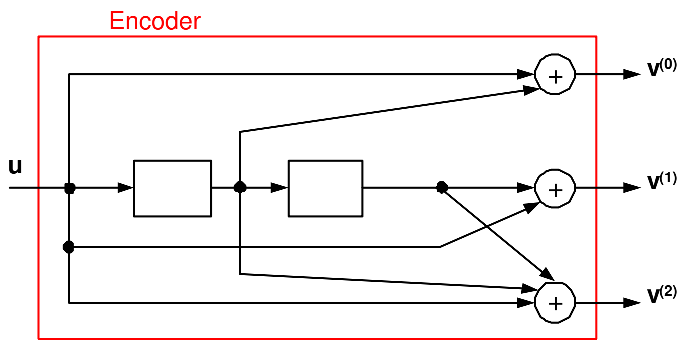


Figure 4.6: A (3,1,2) Binary Nonsystematic Feedforward Convolutional Encoder

If we look the encoder in Figure 4.6 as a black box with $k = 1$ input bit and $n = 3$ output bits, L identical encoders then form a virtual low rate $1/3L$ nonsystematic feedforward convolutional encoder. This 1-input $3L$-output virtual encoder is shown in Figure 4.7. The interesting feature is that the output bits of each identical sub-encoder are identical too. This encoder does not actually exist since each packet is encoded using the (3,1,2) sub-encoder. It only represents the encoding structure when we put the L repeated packets together. When some error occurs in one or more of the repeated packets, the output bits will be no longer the same.

If the cooperating nodes are close (relative to the distance between the transmitting node and the cluster head) to each other and close to the cluster head, the signal to noise ratios for all nodes are almost the same. In this case, the received packet weights w_i used in the code combining technique are the same for all the cooperative nodes, thus can be ignored. This scenario is referred as *uniform channel*

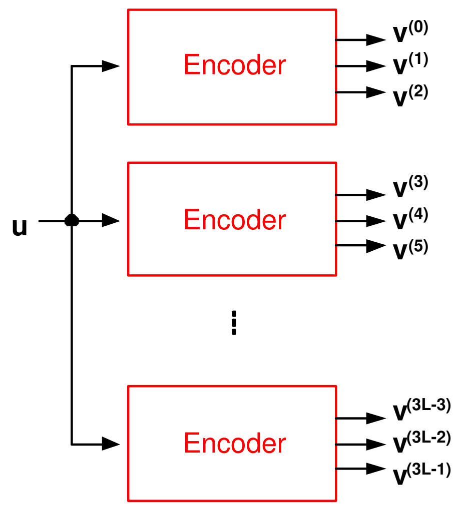


Figure 4.7: The Virtual ($3L$,1,2) Convolutional Encoder with L Combined Encoder

condition.

For general convolutional codes with maximum likelihood decoding (Viterbi algorithm), the *bit error probability*, P_b, that is, the expected number of information bit errors per decoded information bit, is used to evaluate the performance of Viterbi algorithm. This bit error probability can be approximated by (upper bound):

$$P_b \approx B_{d_{free}} \left[2\sqrt{p(1-p)} \right]^{d_{free}} \tag{4.11}$$

where $B_{d_{free}}$ is the coefficient of $X^{d_{free}}$ in the bit *weight*[4] *enumerating function* (WEF) $B(X)$, and d_{free} is the *minimum free distance*.

In code combining the decoder receives L corrupted copies of the transmitted packets. A k-input n-output convolutional code with rate $R = k/n$ with L repeated packets, can be modelled by a k-input nL-output convolutional code with rate R/L. The Viterbi decoder for this rate R/L convolutional code has exactly the same trellis structure as the original rate R convolutional code. The only difference is how the metric for each branch of the trellis is calculated. Therefore, the decoder for the code combiner and the decoder for the original convolutional code have the same order of complexity. Note that for both block and convolutional codes, the minimum

[4]It is unfortunate that we use the term "weight" both for the measure of the quality of a link (w_i) and for the number of ones in a binary sequence (d or $W(\cdot)$).

distance for a code repeated L times is simply L times the minimum (free) distance of the basic code. Furthermore, it is easy to see that the WEF of the R/L rate convolutional code, $B_L(X)$, has the following relation with the WEF of the original code:

$$B_L(X) = B(X^L) \tag{4.12}$$

Hence the lowest power of X in $B_L(X)$ is Ld_{free}, i.e., $d_{free}(L) = Ld_{free}$, and

$$P_b(L) = B_{d_{free}} \left[2\sqrt{p(1-p)} \right]^{Ld_{free}} \tag{4.13}$$

In this expression, p refers to the transition probability of a BSC channel.

We now use an optimum rate $R = 1/2$ convolutional code (2,1,3) with $d_{free} = 6, B_{d_{free}} = 2$ to simulate the bit error probability P_b vs. L. The result is shown in Fig.4.8.

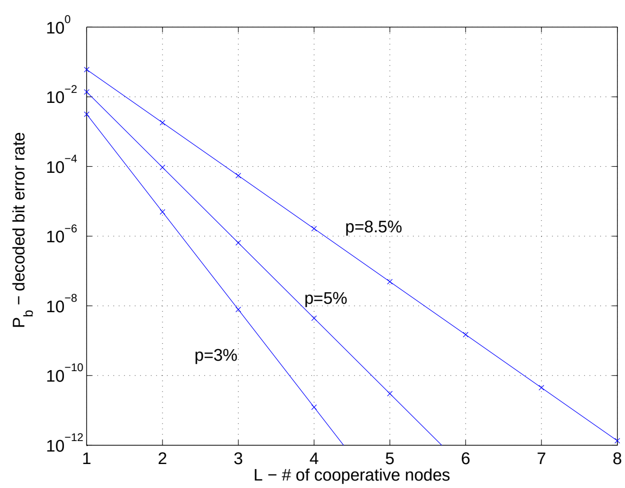


Figure 4.8: Decoded Bit-error Rate P_b vs. L with Different Channel Error Rates

4.2.4 Code Combining with Convolutional Codes in Different Channel Conditions

The assumption made in Section 4.2.3 is mainly valid when code combining is used together with hybrid ARQ, where the same channel is used for packet retransmission. However, in a cluster-based cooperation system, the channel condition can vary significantly among nodes. This is due to the different path losses caused by the different distances between receiver nodes and the transmitter. For this reason, the packets received with higher SNR should have higher weights in the decoder at the master node. The following part in this section will discuss the performance analysis of the weighted code combining. The results depend on the well-known performance bound for convolutional codes using Viterbi decoding, which is described in the following fact:

Fact 1 *Using the analysis of the maximum-likelihood path selection on a trellis diagram, the error probability of a convolutional code with optimum decoding can be upper-bounded using a union bound, by the sum of the error probabilities of each of the paths. The bit-error probability, that is, the expected number of information bit errors per decoded information bit, can be approximated by:*

$$P_b < \sum_{d=d_{free}}^{\infty} B_d P_d \tag{4.14}$$

B_d *is the total number of nonzero information bits on all weight-d paths, divided by the number of information bits k per unit time (i.e., the coefficient of the weight-d term in the bit WEF* $B(X) = \sum_{d=d_{free}}^{\infty} B_d X^d$ *of the decoder).* P_d *is the event error probability for the weight-d path. This bound is tight, because* P_d *is very small. Therefore the union bound is the dominant part for the whole probability of error.*

B_d is determined by the encoder. P_d has a nice expression for ordinary Viterbi decoding over a BSC channel. In weighted code combining, the result for P_d is more complicated. We assume that the decoder is aware of the channel condition for each cooperative node (this can be achieved by piggybacking extra bits during intra-cluster transmission process). Using the channel conditions, the decoder assigns the weight $w_i = \log \frac{1-p_i}{p_i}$ to the ith repeated packet according to the channel error rate p_i, for $i = 1, ..., L$.

A path with weight d would have the weight Ld when the code combining of order L is used. Let the *pseudo codeword* made of bits in these d positions for the correct path be $\mathbf{v}$, the corresponding pseudo codeword for the incorrect path be $\mathbf{v}'$, and the received set of packets be $\mathbf{r} = \{\mathbf{r}_1, \cdots, \mathbf{r}_L\}$. $\mathbf{r}_i$ is the ith received repeated packet. The path metric for $\mathbf{r}$ and $\mathbf{v}$ is given by

$$M(\mathbf{r}|\mathbf{v}) = \sum_{i=1}^{L} w_i d(\mathbf{r}_i, \mathbf{v}) \tag{4.15}$$

where $d(\mathbf{x}, \mathbf{y})$ is the Hamming distance between codewords $\mathbf{x}$ and $\mathbf{y}$.

For a weight-Ld path, a first event error will be made if, in the Ld positions in which the correct and incorrect path differ, the path metric for the incorrect path is less than that of the correct path (so the decoder wrongly chooses the incorrect path). The probability of such event is given by

$$\mathbf{P}[M(\mathbf{r}|\mathbf{v}') < M(\mathbf{r}|\mathbf{v})] = \mathbf{P}\left[\sum_{i=1}^{L} w_i d(\mathbf{r}_i, \mathbf{v}') < \sum_{i=1}^{L} w_i d(\mathbf{r}_i, \mathbf{v})\right] \tag{4.16}$$

From the linear property of the convolutional codes, the all-zero path is always assumed to be the correct path and the non all-zero path is the incorrect path. Therefore, $\mathbf{v}$ consists of d zeros and $\mathbf{v}'$ consists of d ones. Thus, $d(\mathbf{r}_i, \mathbf{v}) = W(\mathbf{r}_i)$ and $d(\mathbf{r}_i, \mathbf{v}') = d - W(\mathbf{r}_i)$, where $W(\mathbf{r})$ represents the Hamming weight of the received packet $\mathbf{r}$. So we have

$$\begin{aligned} \mathbf{P}[M(\mathbf{r}|\mathbf{v}') < M(\mathbf{r}|\mathbf{v})] &= \mathbf{P}\left[\sum_{i=1}^{L} w_i (d - 2W(\mathbf{r}_i)) < 0\right] \\ &= \mathbf{P}\left[\sum_{i=1}^{L} w_i W(\mathbf{r}_i) > \frac{d}{2}\sum_{i=1}^{L} w_i\right] \end{aligned} \tag{4.17}$$

If there is a tie between the metrics of the paths, decoder will randomly choose one. Let $c_{Ld} = \frac{d}{2}\sum_{i=1}^{L} w_i$, and $S = \sum_{i=1}^{L} w_i W(\mathbf{r}_i)$. Therefore, the probability of decoding error is given by

$$P_{Ld} = \mathbf{P}[S > c_{Ld}] + \frac{1}{2}\mathbf{P}[S = c_{Ld}] \tag{4.18}$$

S is the weighted sum of L binomial random variables with different parameter sets (d, p_i). We make use of the generating function to calculate the Probability Mass

Function (PMF) of random variable S:

$$\begin{aligned} G_s(z) &= \mathbf{E}\left[z^{\sum_{i=1}^{L} w_i W(\mathbf{r}_i)}\right] = \prod_{i=1}^{L} G_{W(\mathbf{r}_i)}(z^{w_i}) \\ &= \prod_{i=1}^{L}(1 - p_i + p_i z^{w_i})^d = \sum_k p_S(k) z^k \end{aligned} \tag{4.19}$$

The coefficient $p_S(k)$ is the probability of $S = k$. Therefore,

$$P_{Ld} = \sum_{k > c_{Ld}} p_S(k) + \frac{1}{2} p_S(c_{Ld}) \tag{4.20}$$

Thus, based on Fact 1 and (4.12), we have the following theorem:

Theorem 1 *The upper bound for the bit-error probability of the distributed code combining method, P_b, is given by:*

$$P_b < \sum_{d=d_{free}}^{\infty} B_d P_{Ld} \tag{4.21}$$

where B_d is the coefficient of the weight-d term in the bit WEF $B(X)$ of the original convolutional code, and B_{Ld} is given by (4.20).

Since p_i is small, P_{Ld} decreases greatly as d increases. P_b is generally dominated by the first several terms of the summation in (4.21), or even the first term $B_{d_{free}} P_{d_{free}}$. So this union bound is tight, and numerical results show the first several terms of the summation in (4.21) can be a good estimation of real P_b. There are $(d+1)^L$ terms in the right hand side of (4.19). For $L \leq 10$, the computation time of P_{Ld} is quite tolerable. Some results will be shown in the simulation section.

An example of the Viterbi decoding procedure in an unequal error condition environment is shown in Figure 4.9. This trellis diagram is for a (3,1,2) code with an information sequence of length $h = 3$. In this example, we assume $L = 3$. For each output, since the original code has $n = 3$ bits, so the combined code in one time unit has 9 output bits, in the order of channel No.1, No.2, and No.3. Assume the reliable factor (weight) $w_1 = 1$, $w_2 = 2$, and $w_3 = 3$. Suppose an all-zero sequence is sent, and the received sequence $\mathbf{r}$ = (000 001 111, 000 100 101, 001 010 001, 000

000 000, 000 000 000). Since the codeword in the trellis structure is a repeated (3 times) codeword, we use one copy - the first 3 bits on each branch - in the following notation. let $\mathbf{v}$ represent the correct path sequence (000, 000, 000, 000, 000), and $\mathbf{v}'$ represent the sequence of the highlighted path in Figure 4.9 (111, 101, 011, 000, 000). Split $\mathbf{r}$ into $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ according to channels, then we will have

$$M(\mathbf{r}|\mathbf{v}) = w_1 \cdot d(\mathbf{r}_1, \mathbf{v}) + w_2 \cdot d(\mathbf{r}_2, \mathbf{v}) + w_3 \cdot d(\mathbf{r}_3, \mathbf{v}) = 1 \cdot 3 + 2 \cdot 3 + 3 \cdot 6 = 24$$

Likewise

$$M(\mathbf{r}|\mathbf{v}') = 1 \cdot 6 + 2 \cdot 4 + 3 \cdot 1 = 17 < M(\mathbf{r}|\mathbf{v})$$

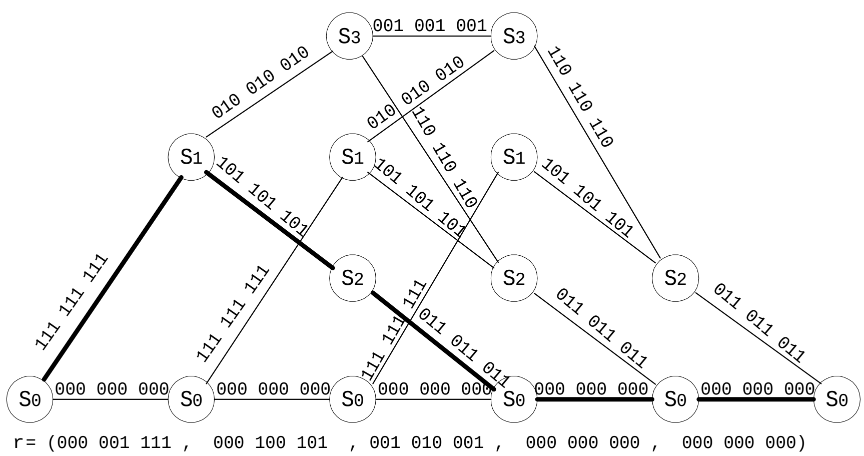


Figure 4.9: The Viterbi Algorithm for Code Combining with Unequal Error Probability

Thus the all-zero path is eliminated. Note if otherwise the channel condition were equal ($w_1 = w_2 = w_3$), then $M(\mathbf{r}|\mathbf{v}) < M(\mathbf{r}|\mathbf{v}')$ and the highlighted path would be eliminated. Using the same algorithm to check the other paths we can decide that the highlighted path is the final survivor, $\hat{\mathbf{v}} = (111, 101, 011, 000, 000)$. This surviving path corresponds to the decoded information sequence $\hat{\mathbf{u}} = (100)$. Note that the final $m = 2$ branches in any trellis path correspond to 0 inputs and hence are not considered part of the information sequence.

4.3 Simulations

In order to evaluate the performance of the cooperative networks, a set of random nodes representing the networks nodes are chosen according to the network topology as follows: the transmitter and the receiver cluster head are fixed nodes and are 250 meters apart. The cluster is formed around the cluster head in a circle with radius of 50 meters. The cooperative nodes are randomly placed as a uniform distribution inside the cluster. The topology of the simulated network is shown in Fig.4.10.

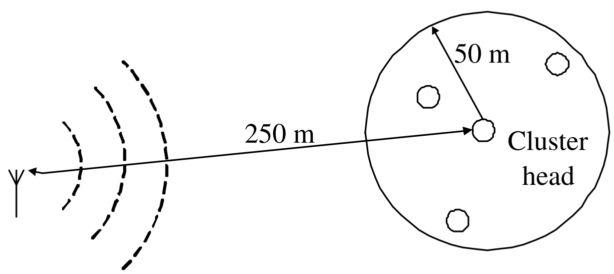


Figure 4.10: Topology of the Simulated Network

In the following simulations the decoded bit-error rate P_b, is calculated using Theorem 1 from section 4.2. The channel model used is Rayleigh fading channel, and Binary Phase Shift Keying (BPSK) is used for modulation [1]. It is shown in [101] that in a frequency-nonselective Rayleigh fading channel with Additive White Gaussian Noise (AWGN), the probability of error for using BPSK is

$$p = \frac{1}{2}\left(1 - \sqrt{\frac{\overline{\gamma_b}}{1+\overline{\gamma_b}}}\right) \tag{4.22}$$

where $\overline{\gamma_b}$ represents the average signal-to-noise ratio. In general $\overline{\gamma_b} = \frac{E_b}{N_0}\mathbf{E}[\alpha^2]$ where α is the attenuation factor due to fading.

4.3.1 Link Layer Decoding Performance

We use different levels of power for inter-cluster and intra-cluster transmission because the distance between cluster nodes and the cluster head are at most 1/5 of the radio distance for inter-cluster transmission. Let PD represent the difference between the power used by the cluster nodes and the power at the sender node, in dB. We consider two cases where PD=10dB and 20dB, i.e. the cluster nodes use a transmit power that is 10dB and 20dB less than the sender transmit power, respectively. This means the SNR level is at least 4.5dB ($10\log_{10}(250/50)^{3.5} - 20$ dB $= 4.5$ dB) higher than the signal received from the sender. For each power level, the simulation takes 100 runs and finds the average decoded bit-error rate. A (2,1,3) convolutional code is used for code combining with Viterbi decoding at the cluster head. The decoded bit-error rate P_b with weighted code combining at the cluster head is plotted as a function of L in Figure 4.11. The SNR is measured at the receiver, i.e., the cluster head. Therefore the SNR is proportional to the sender transmission power. Changing PD from 10dB to 20dB does not change the overall performance of the code combining technique significantly. The change is negligible when the sender transmits at a considerably high power, e.g., in this simulation when SNR=8dB.

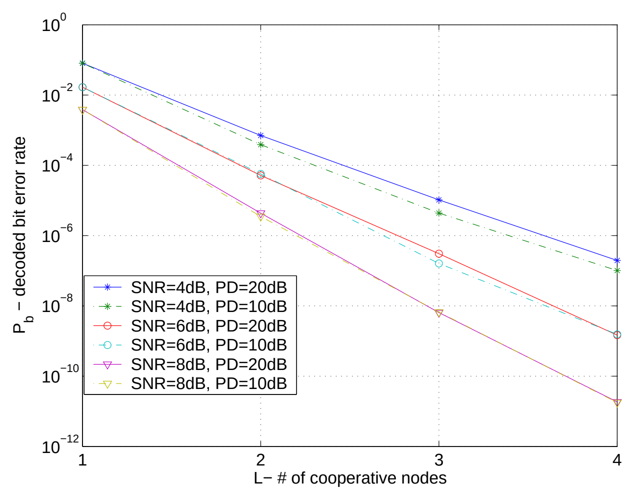


Figure 4.11: Decoded Bit-error Rate P_b vs. Number of Cooperative Nodes L. PD is the Amount of Power Deduction of the Intra-cluster Transmission upon the Inter-cluster Transmission.

We also tried different cluster radii for the simulations. For PD=20dB, we simulated the cluster radii of 50m and 100m. The decoded bit-error rate is plotted in Figure 4.12. A larger cluster radius leads to a worse decoding performance since some cluster nodes may be too far from the sender node. However, it is shown in both Figure 4.11 and Figure 4.12 that the decoded bit error rate decreases sharply when

L increases. A system designer should take this fact into account when deciding about the maximum number of the cooperation nodes.

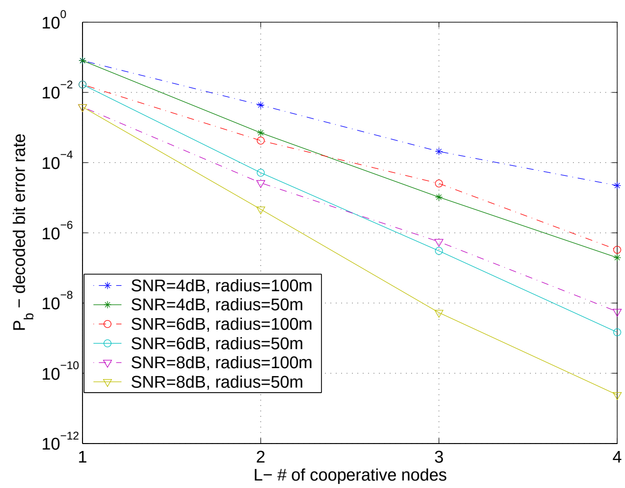


Figure 4.12: Decoded Bit-error Rate P_b vs. Number of Cooperative Nodes L with Different Cluster Radius. Smaller Cluster Radius Has a Better Performance.

To provide a reliable link performance, a very low bit error rate is desired. In another round of simulations, a couple of fixed decoded bit-error rates, 10^{-7}, 10^{-6}, and 10^{-5}, are set to be the objectives. The choice of the desired P_b mainly depends on the frame size. For each random topology, the sender power level is adjusted to achieve the desirable P_b. Cluster nodes use 20dB less power than the sender node (PD=20dB). We plot the required SNR at the cluster head as a function of cluster size to compare the dB gain of the cooperative code combining technique, as shown is Figure 4.13. Note when $L = 1$ it means there is no cooperation. So the difference between the SNR of cooperation and non-cooperation is very similar to the concept of *coding gain*.

To view the coding gain more clearly, we plot the probability of error versus the signal to noise ratio (E_b/N_0 at the cluster head) of different cooperative levels ($L = 1 \sim 4$), shown in Figure 4.14. From this figure, it is shown that the decoding performance of cooperative FEC with code combining ($L > 1$) has a obvious dB gain over that of the non-cooperative FEC ($L = 1$).

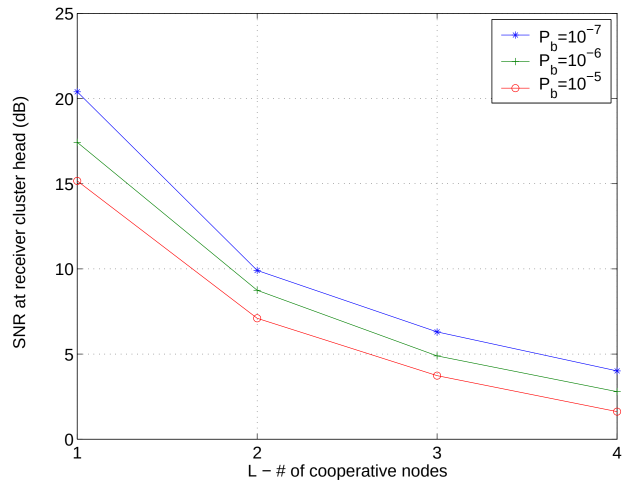


Figure 4.13: SNR vs. Number of Cooperative Nodes L. With a Fixed Objective P_b, the Required SNR Decreases with the Increase of the Cluster Size L.

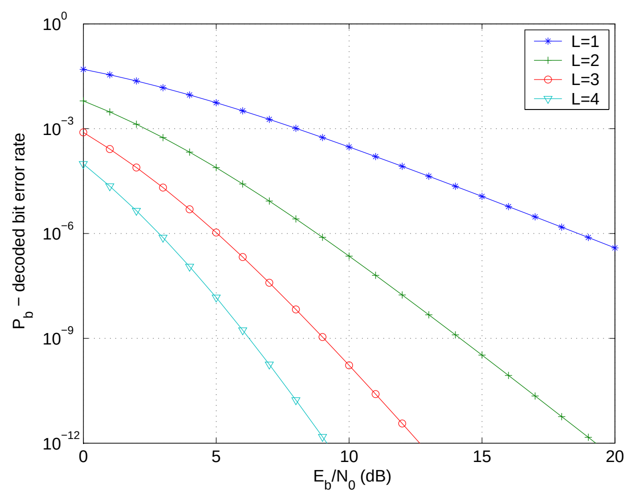


Figure 4.14: Decoded Bit-error Rate P_b vs. SNR (dB) with Different Cluster Size L. The dB Gain with the Use of Cooperation over Ordinary FEC is Substantial.

4.3.2 Energy Consumption

Generally clustering is proposed to solve scalability problem, node mobility problem in wireless ad hoc networks. Clusters can maintain a relatively stable effective topology. The membership in each cluster changes over time in response to node mobility, node failure or new node arrival. Clustering techniques are expected to achieve better scalability since most of the topology changes within a cluster are hidden from the rest of the network. In addition, clustering can be extremely effective in multicast, broadcast, and pear-to-peer communication. Clustering can support data fusion and data dissemination/dissipation. In sensor networks, some clustering protocols are claimed to save energy expenditure. The clustering in our scheme is for the cooperation purpose, yet it can save energy as well.

The cost for the cooperation is the energy consumed at the cluster nodes. To take this into account, we will model the aggregate energy spent in transmitter together with all the cluster nodes for successfully transmitting one bit.

We use a simple energy consumption model to evaluate the total cost of the communication, including the transmitting and the receiving [102]. The average energy consumption of the radio in transmission process can be described by:

$$E_t = P_{tx}T_{tx} + P_{out}T_{tx} \tag{4.23}$$

and the average energy consumption of the receiving system can be expressed as

$$E_r = P_{rx}T_{rx} \tag{4.24}$$

where $P_{tx/rx}$ is the power consumption of the transmitter/receiver, P_{out} is the output transmit power, $T_{tx/rx}$ is the transmit/receive on-time (actual data transmission time). Note that $T_{tx/rx} = l/R$, where l is the packet size and R is the data rate in bits per second. Also note that if R_c is the code rate then the number of information bits in a packet is $l' = lR_c$. To transmit an l'-bit information message, the radio expends $(P_{tx} + P_{out})\frac{l'}{RR_c}$ and to receive l' bit message, the radio expends $P_{rx}\frac{l'}{RR_c}$. The electronics power P_{tx} and P_{rx}, are the amount of power spent in the transmitter electronics circuitry, depending on digital coding, modulation, filtering and spreading of the signal, while P_{out} is the amount of energy spent in the RF amplifiers to counter the propagation loss. Here P_{out} takes into account the constant factor in the path loss term, as well as the antenna gains of the transmitter and the receiver. When receiving a packet, only the receiver circuitry is invoked.

In general, with the assumption on fading channel and the modulation scheme, the probability of error is a function of signal to noise ratio E_b/N_0. After knowing how strong the signal is, then we can convert E_b/N_0 to carrier to noise ration using the equation:

$$\frac{C}{N} = \frac{E_b}{N_0}\frac{f_b}{B_w} \tag{4.25}$$

where f_b is the bit rate, and B_w is the receiver noise bandwidth.

Since we now have the carrier to noise ratio, we can determine the necessary received carrier power after we calculate the receiver noise power. The thermal noise power N_{th} is computed using Boltzmann's equation:

$$N_{th} = kTB \tag{4.26}$$

where k is Boltzmann's constant $= 1.380650 \times 10^{-23}$ J/K, T is the effective temperature in Kelvin, and B is the receiver bandwidth. Therefore, $N_{th} = (1.380650 \times 10^{-23}$ J/K $\times$290K$\times$1 MHz $= 4 \times 10^{-15}$ W $= 4 \times 10^{-12}$ mW = -114dBm.

The receiver has some inherent noise in the amplification and processing of the signal. This is referred to as the receiver noise figure N_{rx}. So the receiver noise level will be:

$$N(\text{dBm}) = N_{th} + N_{rx} \tag{4.27}$$

We can now find the carrier power as $C = C/N \times N$, or in dB $C = C/N + N$. This is how much power the receiver must have at its input. To determine the transmitter amplifier power P_{out}, we must account for the path loss that we are building in to the system.

The log-distance path loss model has been used extensively in the literature [1]. The average large-scale path loss for an arbitrary transmitter-receiver separation is expressed as a function of distance by using a path loss exponent γ, which indicates the rate at which the path loss increases with distance. The value of the propagation loss exponent γ is highly dependent on the surrounding environment (usually between 2 to 4). The average path loss for a T-R separation with distance d can be expressed by

$$\overline{PL}(\text{dB}) = \overline{PL}(d_{ref}) + 10\gamma \log_{10} \frac{d}{d_{ref}} \tag{4.28}$$

where d_{ref} is the close-in reference distance and can be based on a free space assumption from the transmitter to d_{ref}. So, assume no system loss, with unity antenna gain, the path loss for the reference distance is given by

$$\overline{PL}(d_{ref}) = 10\log_{10}\frac{(4\pi)^2}{\lambda^2} \tag{4.29}$$

where λ is the wavelength in meters.

Finally, adding the path loss to the receiver carrier power will give the required transmitter amplifier power:

$$P_{out}(\text{dBm}) = C + \overline{PL} \tag{4.30}$$

In our link layer cooperation scheme, d_0 and d_1 are the average intra-cluster and inter-cluster distance, respectively. Since $d_0 < d_1$ or even $d_0 \ll d_1$, they correspond different level of path loss, leading to different power consumption of radio amplifier. The energy spent for successfully relaying one packet during a single hop is:

$$\begin{aligned} E_c \;=\; & [(P_{tx}+P_{out})\frac{k}{RR_c} + \bar{L}P_{rx}\frac{k}{RR_c} \\ & +(\bar{L}-1)(P_{tx}+P_{out}^{(0)})\frac{k}{RR_c} + (\bar{L}-1)P_{rx}\frac{k}{RR_c} \\ & +(P_{tx}+P_{out})\frac{k'}{R} + P_{rx}\frac{k'}{R}]\bar{T}_c \end{aligned} \tag{4.31}$$

where $\bar{L}$ is the average number of the cooperative nodes in a cluster, R_c is the FEC code rate used for code combining, k' represents the control packet length for ACK, ARQ request, or request intra-cluster cooperation, $\bar{T}_c$ is the average transmission times for a packet, including first transmission and retransmissions. Note $P_{out}^{(0)}$ represents the power requirement for radio amplifier used in intra-cluster transmission.

The expression in first line of the equation indicates the inter-cluster communication (one node sends and $\bar{L}$ cluster nodes receive); the second line indicates intra-cluster communication (all the cluster nodes except the cluster head send and the cluster head receive); and the last represents the ACK or ARQ request.

If $\bar{L} = 1$, expression for E_c describes the energy consumption for a non-cooperative network, which is exactly the case with conventional hybrid ARQ with the same FEC coding but without cooperation. The communication energy spent in a single hop is for this hybrid ARQ is:

$$E_h = [(P_{tx} + P_{out})\frac{k}{RR_c} + P_{rx}\frac{k}{RR_c} + (P_{tx} + P_{out} + P_{rx})\frac{k'}{R}]\bar{T}_h \tag{4.32}$$

where T_h is the counterparts of T_c, the average transmissions for a packet in a single hop in the non-cooperation scenario. Keep in mind that although expressions are in the similar forms, with or without cooperative FEC, the power levels for meeting the requirement of probability of error are very different.

Table 4.2 gives the parameter settings used in our simulations. The parameters about the energy model are from [102].

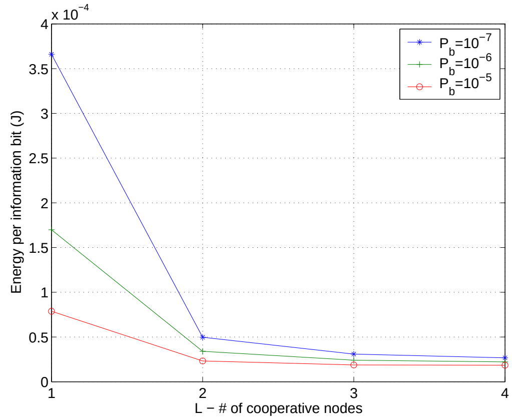
γ	3.5
f_b	2 Mbps
B_w	1 MHz
λ	0.122 m
N_{rx}	10 dB
d_{ref}	1 m
d_1	250 m
d_0	50 m
R_c	1/2
P_{tx}	81 mW
P_{rx}	180 mW
l	500 bytes

Table 4.2: Variables in the Energy Model

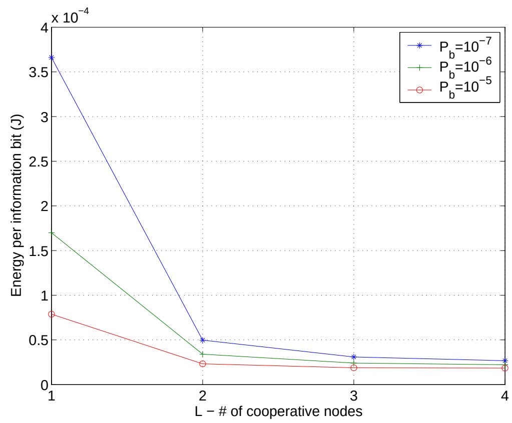
Given the above energy model, we use the topology in Figure 4.10 to simulate the average energy consumption per useful information bit for different number of cooperative nodes. Three plots in Figure 4.15 illustrate the results for different source to destination distances d_1 and cluster radii d_0. Note Figure 4.15(a) and Figure 4.15(b) are almost the same. This shows that cluster radius is not a significant factor as far as energy consumption is concerned. Also note that the scale of Figure 4.15(c) is different from that of the other two. This shows the distance from sender to receiver is the major factor for the power consumption. In Figure 4.15(c), the distance between sender to receiving cluster head is relatively small, thus power used to combat path loss is relatively low. The power consumed in transmit/receive electronics circuitry begins to take more effect. This is why the energy per information bit increases slightly after $L = 2$. Therefore, from the prospective of power efficiency, cluster size L is not necessarily very big. Nevertheless, cooperative FEC offers energy saving over traditional forwarding. In addition, the cooperative FEC is more efficient when long distance transmission is needed.

The above simulations are just some case studies to illustrate how cooperation can increase the decoding performance. If the channel quality is better than the channel used in these simulations, we may choose a code with a higher rate than 1/2 used in the above examples. In fact, such a low code rate as 1/2 will bring too much overhead in ad hoc networks. Obviously codes with lower rates have a better performance in terms of the decoded error rate. Given the desired P_b, and the channel condition, we can choose the appropriate operating point (code rate and cluster size) to meet the needs. A higher rate convolutional code can be achieved using punctured codes, which is a simple operation on a lower rate code without additional complexity. Likewise, the result can be extended to the case with longer distances between the transmitter and the receiver, different node density etc. The cluster size may be adapted to the channel condition and the code rate.

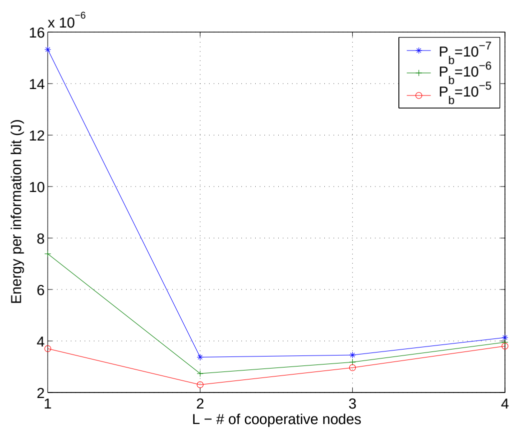
At this point, we intend to ignore some secondary considerations in this thesis. These considerations include the contribution of idle-listening on the RF channel to the energy consumption, and contention resolvent at the cluster head. We regard these issues as secondary because it is well accepted the energy consumed by listening is much less than that by transmission or reception, and contention problem can be solved by intra-cluster transmission MAC, e.g., TDMA or CDMA. It is also worth mentioning that the decoding energy is not covered in our energy consumption consideration. There is a good reason behind this. Cooperative FEC decoding does no more decoding than ordinary FEC decoding. Given our previous argument that the complexity of each code combining decoding is as same as each basic code decoding, the energy spent for cooperative decoding keeps unchanged. As far as we consider scenarios with same FEC codes, the decoding energy expenditure is a



(a) d_1=250m, d_0=50m



(b) d_1=250m, d_0=100m



(c) d_1=100m, d_0=50m

Figure 4.15: Energy per Useful Information Bit vs. Number of Cluster Nodes Given Different Decoding Probability of Error P_b. Note that $L = 1$ Corresponds to Non-cooperative Networks.

constant factor for all the cooperation schemes.

We also leave the higher level design to our future work. The cluster-based routing and forwarding has been extensively studied by researchers. This cooperative FEC can adopt most of the traditional cluster-based routing architectures and build the cooperation based on these routing protocols. Recently Biswas and Morris [103] proposed an opportunistic routing protocol which involves one sender and choice of multiple receivers when forwarding packets. This follows our basic motivation of cooperation - nodes with unreliable links can help each other - and can be integrated to our cooperation scheme.

4.4 Summary

Code combining is a well-established technique used in Hybrid ARQ transmissions. Its use with some specific convolutional codes are simulated by researchers but the generic analytical performance results are never given. In this thesis we analyzed the decoding performance of the cluster-based cooperative networks with the code combining technique. Simulation results from various aspects show this cooperation architecture is effective in improving the link performance and reducing the energy consumption. This result is promising in that the reduced power requirement leads to less interference caused by a transmission, thus can improve the capacity of the wireless networks.

The results in this chapter are under the consideration of a single hop network. Yet they are applicable to a multi-hop network as well. However, more problems will be involved, such as the effect of interference, MAC design, and so on. Our future work will look into the detailed cross layer design of the network, including cooperation-intended cluster-based routing, medium access issues in the intra-cluster communications, network performance from all aspects, and more information theoretic analysis of the coding technique and network capacity.

The vision of our work is to develop enabling core technology for cooperative wireless networks and fuel the interdisciplinary effort which is required to make cooperation at each level a reality. Our results indicate that this approach achieves a quantum leap in the performance/cost trade off. The future focus of our work is on designs which explicitly exploit physical layer, data link layer, and network layer cooperation among nodes.

CHAPTER 5
Capacity Improvement with the Use of Directional Antennas at Transceivers

5.1 Introduction

Ad hoc wireless networks are wireless networks without fixed base stations or any wireline backbone infrastructure. The nodes use peer-to-peer packet transmissions and multi-hop routes to communicate with each other. Throughput capacity is a key characteristic of ad hoc networks. The problem of the capacity of wireless ad hoc networks has been extensively studied. Consider an ad hoc network with n nodes randomly located in a domain of area one square meter. It was shown by Gupta and Kumar in [12] that under a Protocol Model of interference, such a network could provide a per node throughput of $\Theta(\frac{1}{\sqrt{n \log n}})$ bits/sec. It was also shown there that even under the best possible placement of nodes, such a network could not provide a per-node throughput of more than $O(\frac{1}{\sqrt{n}})$ bits/sec. In this case, the total end-to-end capacity is roughly $O(\frac{n}{\sqrt{n}})$, which is $O(\sqrt{n})$.

The key reasons why the overall capacity is reduced are:

a) interference in a zone around the receiver prevents any other node in the zone from receiving data from any transmitter.

b) as the number of hops increases, the "forwarding burden" of nodes increases; i.e., they spend a fraction of their capacity relaying other nodes' traffic rather than their own. Even if the interference zone around receivers is of area 0, due to range limits of any one-hop, the multi-hops necessary for a large network may in general grow as $\sqrt{n}$.

Several works study how these reasons affect the capacity of the network and try to find ways to complement these effects [3, 4, 5, 6, 9, 10]. The major purpose of these works is to discover the possibilities of improving the capacity of the networks from different approaches .

Such research on the capacity of wireless ad hoc networks and the popular IEEE 802.11 protocol typically assume the use of *omnidirectional* antennas at all nodes. An outcome of this assumption is that all nodes lying in the vicinity of a pair of communicating nodes are required to stay silent. However, with *directional* antennas, more than one pair of nodes located in each other's vicinity may potentially communicate simultaneously, depending on the directions of transmission. This can increase spatial reuse of the wireless channel.

In order to evaluate the performance and spatial reuse property of directional antennas, some new MAC protocols are proposed [13, 14, 15, 16, 17, 104, 105]. Most of these works focus on designing MAC protocols to take advantage of the use of directional antennas and show with simulation results the improvement on the performance of wireless ad hoc networks. Recently, some research work has tried to provide a theoretical framework to understand how much capacity improvement can be achieved. It is presented [106] that mutual interference by simultaneous transmissions poses bounds on the amount of capacity gain one can achieve by using directional antennas instead of omnidirectional ones. They calculate interference-based capacity bounds for a generic antenna model as well as a real-world antenna model and analyze how these bounds are affected by important antenna parameters like gain and beam-width. It is found in [8] that an increase of $\Theta(\log^2(n))$ in maximum stable throughput is all that can be achieved by allowing arbitrarily complex signal processing at the transmitters and receivers. They also show that under some condition on the radius of each transmission, the maximum stable throughput could achieve linear in the number of nodes.

In this thesis we introduce an interference model for antennas to analyze the capacity improvement on using directional antennas. We consider two types of networks, *Arbitrary Networks*, where the node locations, destinations of sources, and traffic demands, are all arbitrary, and *Random Networks*, where the nodes and their destinations are randomly chosen. Basically random implies a statistical distribution and arbitrary means the distribution could be arranged to achieve best result in terms of capacity. This thesis also gives a complete analysis, extensive discussions on the achieveability of the capacity bound, and simulation results for validation.

This chapter is organized as follows: First we give a brief description of the antenna background in order to introduce our antenna model in Section 5.2.Then analysis for the capacity improvement for arbitrary network is given in Section 5.3, and for random network in Section 5.4. Section 5.5 extends the results with a hybrid antenna model. In Section 5.6 we present our simulations and results. Last in Section 5.8 we give our concluding remarks.

5.2 Antenna Model

The antenna family has more than twenty types of antennas which can be grouped in variable ways [107]. In the study of wireless networks, the antenna model is often grouped under omnidirectional and directional. Omnidirectional antennas, also known as isotropic antennas, radiate and receive equally well in all directions. This unfocused approach scatters signals, reaching desired users with only a small percentage of the overall energy sent out into the environment.

Given the limitation of omnidirectional antennas, directional antennas are used to overcome this inadequacy. Figure 5.1 shows a common directional pattern in two-dimension [108]. The main lobe is the direction of maximum radiation or reception. In addition to the main lobe, there are also sidelobes and backlobes. These lobes represent lost energy so good antenna designs attempt to minimize them. We always want the main lobe to extend toward a user with a null directed toward a co-channel interferer. In this thesis, the beamwidth refers to HPBW (Half-Power Beamwidth), measured between the -3 dB points, i.e. the points on the main lobe where the signal strength drops off -3 dB (one-half) from the maximum signal point.

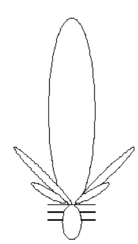


Figure 5.1: Radiation Pattern for Directional Antennas

In physical layer, the parameter to measure the antenna is the gain in each direction. Antenna *gain* is given in units of dBi (dB gain with respect to an isotropic source) or dBd (dB gain with respect to a half-wave dipole). In medium access technology, the transmission or receiving range is always used to describe the distance the antenna can reach. From path loss model [1], it is shown that the distance is an inversely proportional function of the antenna gain, with the exponent factor γ. For this reason, we will only refer *range* as the indication of antenna gain.

In our model, we use a circle to model the omnidirectional antenna, with the only parameter - radius r indicating the transmission or receiving range. We approximate the directional antenna pattern as a circular sector with radius r and angle α or β depending on the mode of the antenna (α for transmitter and β for receiver). We use different parameters for antenna beamwidths of transmitters and receivers because we use a new conditional probability argument to handle the case of directional reception. Radius r represents the transmission range or receiving range according to the antenna mode. The angle of the sector approximates the beamwidth of the antenna pattern. The reasons we simplify the antenna model are as follows. It's difficult to model a real antenna with precise values from main lobes, sidelobes, and backlobes. As we will show in the following sections, the only thing that matters is the interference area of the nodes. Simplifying the shape of antenna pattern will not change this property and using a complex model will not result in a fundamental change in our work on capacity analysis.

5.3 The Transport Capacity for Arbitrary Networks

5.3.1 Interference Model

We build the *sender-based* interference model as follows: the interference zone is defined as the area that a transmission can cover. That is to say, this transmission will interfere with all the nodes except the intended receiver. The reason that we use the *sender-based* interference model instead of the *receiver-based* model in [12] is that this model can result in the same scaling law yet is easier to be extended to analyze directional antennas.

We assume the senders should be set apart in order to avoid collision in the intersection of the transmission zones. This is a conservative assumption of the real case, since things may be better in practice. All the bounds derived from this model hold for this assumption. Also the capacity improvement achieved by using directional antennas is feasible.

First let's consider the simplest case whose transmission and reception are both omnidirectional. Suppose node Tx_1 and node Tx_2 are transmitting over the mth subchannel, where their transmission ranges are r_1 and r_2, respectively. Then these two transmitters should be at least at a distance $(1+\Delta)(r_1+r_2)$ to avoid collision from each other, which is:

$$|Tx_1 - Tx_2| \geq (1+\Delta)(r_1+r_2) \tag{5.1}$$

The quantity $\Delta > 0$ models situations where a guard zone is specified by the protocol to prevent a neighboring node from transmitting on the same subchannel at the same time. We require the guard zone $\Delta > 0$. The Interference Model is illustrated in Figure 5.2.

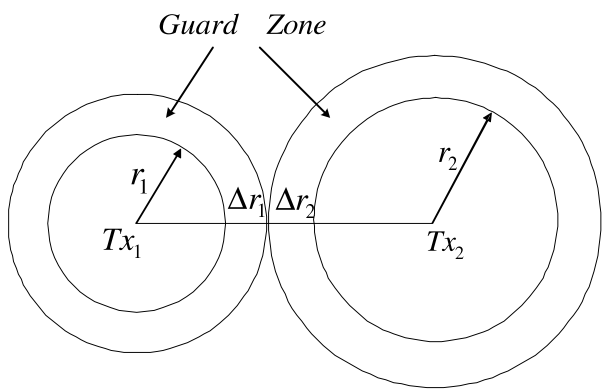


Figure 5.2: *Sender-based* Interference Model

We consider the setting on a planar disk of unit area. Consider the following assumptions:

(A1) There are n nodes arbitrarily located in a disk of unit area on the plane. These nodes are immobile.

(A2) The network transports $\lambda n T$ bits over T seconds.

(A3) The average distance between the source and destination of a bit is $\bar{L}$. So together with (A2), this implies that a transport capacity of $\lambda n \bar{L}$ bit-meters per second is achieved.

(A4) Each node can transmit over any subset of M subchannels with capacities W_m bits per second, $1 \leq m \leq M$, where $\sum_{m=1}^{M} W_m = W$.
(A5) Transmissions are slotted into synchronized slots of length τ seconds.

5.3.2 Omnidirectional Antennas

Now we get to the proof part for the upper bound on transport capacity using omnidirectional antennas. We adopt the reasoning introduced in [12] to get the upper bound for transport capacity of the network.

Consider bit b, where $1 \leq b \leq \lambda nT$. Let us suppose that it moves from its origin to its destination in a sequence of $h(b)$ hops, where the hth hop traverses a distance of r_b^h. Let L_b be the distance between the source and destination of bit b. Since the total traversed distance via multiple hops is always less or equal to the straight line connecting the source and destination, then from (A3)

$$\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h(b)} r_b^h \geq \sum_{b=1}^{\lambda nT} L_b = \lambda nT\bar{L} \tag{5.2}$$

Note now in any slot at most $n/2$ nodes can transmit. Hence for any subchannel m and any slot s

$$\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h(b)} 1(\text{The } h\text{th hop of bit } b \text{ is over subchannel } m \text{ in slot } s) \leq \frac{W_m \tau n}{2} \tag{5.3}$$

Summing over the subchannels and the slots, and noting that there can be no more than $\frac{T}{\tau}$ slots in T seconds, yields

$$\sum_{b=1}^{\lambda nT} h(b) \leq \frac{WTn}{2} \tag{5.4}$$

Let $H = \sum_{b=1}^{\lambda nT} h(b)$ for future reference.

From the Interference Model introduced above, disks of radius $(1 + \Delta)$ times the lengths of hops centered at the transmitters over the same subchannel in the same slot are essentially disjoint. Ignoring edge effects, all these disks are within the domain. Since at most $W_m\tau$ bits can be carried in slot s from a receiver to a transmitter over the mth subchannel, we have

$$\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h(b)} 1(\text{The } h\text{th hop of bit } b \text{ is over subchannel} \\ m \text{ in slot } s)\pi(1+\Delta)^2(r_b^h)^2 \leq W_m\tau \cdot 1 \tag{5.5}$$

Here the unit in each side is bit-meter^2. Summing over the subchannels and the slots gives

$$\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h(b)} \pi(1+\Delta)^2(r_b^h)^2 \leq WT \tag{5.6}$$

This can be rewritten as

$$\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h(b)} \frac{1}{H}(r_b^h)^2 \leq \frac{WT}{\pi(1+\Delta)^2 H} \tag{5.7}$$

Note now that the quadratic function is convex. Hence

$$(\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h(b)} \frac{1}{H} r_b^h)^2 \leq \sum_{b=1}^{\lambda nT}\sum_{h=1}^{h(b)} \frac{1}{H}(r_b^h)^2 \tag{5.8}$$

Combining (5.7) and (5.8) yields

$$\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h(b)} r_b^h \leq \sqrt{\frac{WTH}{\pi(1+\Delta)^2}} \tag{5.9}$$

Now substituting (5.2) in (5.9) gives

$$\lambda nT\bar{L} \leq \sqrt{\frac{WTH}{\pi(1+\Delta)^2}} \tag{5.10}$$

Substituting (5.4) in (5.10) yields the result:

$$\lambda n\bar{L} \leq \frac{1}{\sqrt{2\pi}}\frac{1}{(1+\Delta)} W\sqrt{n} \text{ bit-meters per second.} \tag{5.11}$$

The transport capacity using omnidirectional antennas based on our interference model is $T_{OO} = \frac{1}{\sqrt{2\pi}}\frac{1}{(1+\Delta)}W\sqrt{n}$ bit-meters per second. Note that $\bar{L}$ is the average distance between the source and the destination in a unit area domain. It can be seen as a constant. So the throughput capacity is always proportional to the transport capacity. Also, it is more meaningful to compare the per-node throughput than the aggregate throughput. For this reason, *throughput capacity* is always used to refer *per-node throughput* in the later part of this chapter.

5.3.3 Directional Transmission and Omnidirectional Reception

If we let the sender be directional, the transmission pattern is no longer a circle with uniform directivity in every direction. The transmission pattern for a directional antenna has a main lobe pointing to the receiver, as well as several side lobes which have much less power in those directions. In this thesis, as mentioned before, we think of the directional antenna model as a sector characterized by (a) the transmission/reception range r and (b) the beamwidth α for transmission or β for reception mode.

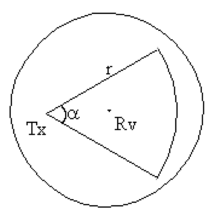


Figure 5.3: *Sender-based* Interference Model for Directional Transmission

For the sender-based interference zone defined in previous section, the zone becomes a sector with radius r and angle α, shown in Figure 5.3. Thus, the disks in the unit area domain become sectors with radius $(1+\Delta)$ (considering the guard zone) times the lengths of hops centered at the transmitters and the beamwidth α (assume all the transmitters have the same beamwidth). These sectors centered at the transmitters over the same subchannel in the same slot are essentially disjoint. Similar to the omnidirectional antenna case (except that α is introduced), the interference zone area is calculated as $\frac{\alpha}{2\pi}\pi(1+\Delta)^2r^2$. Following the same procedure above ((5.2) to (5.11)), we change only (5.5) such that:

$$\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h(b)} 1(\text{The } h\text{th hop of bit } b \text{ is over subchannel } m \text{ in slot } s) \frac{\alpha}{2\pi} \pi (1+\Delta)^2 (r_b^h)^2 \leq W_m \tau \tag{5.12}$$

So the transport capacity becomes:

$$T_{DO} = \frac{1}{\sqrt{\alpha}} \frac{1}{(1+\Delta)} W\sqrt{n} \text{ bit-meters per second.} \tag{5.13}$$

Compared the result in the previous section, the directional transmission scales the capacity by $\sqrt{\frac{2\pi}{\alpha}}$.

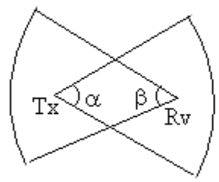
5.3.4 Directional Transmission and Directional Reception

What will happen if the receiver antennas are also directional? Intuitively, the result should be more optimistic and the following analysis confirms that it is. Let's consider the sender-based interference zone. Unlike the omnidirectional reception, not all the receivers in this zone will be interfered with. Therefore the transmission zones (interference zone) are not necessarily disjoint. We propose the following modification to the interference zone concept. We introduce a new conditional probability argument for this case. This conditional probability is defined to be the probability that a specific receiver will experience interference given it is in the transmission zone. Assume all the receivers have the same antenna characteristics and can point in any direction with equal probability. The probability that the antenna pattern of a receiver will cover the transmitter is $\frac{\beta}{2\pi}$. So the conditional probability of interference for a receiver within the transmission zone is $\frac{\beta}{2\pi}$, demonstrated in Figure 5.4.

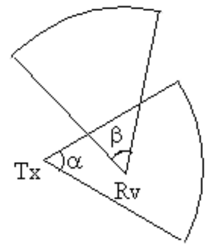
On average, there are $\frac{\beta}{2\pi}$ proportion of the number of receivers inside the transmission zone will get interfered with. Thus the *conditional* interference zone area is:

$$\frac{\beta}{2\pi}[\pi(1+\Delta)^2(r_b^h)^2\frac{\alpha}{2\pi}] = \frac{\alpha\beta(1+\Delta)^2(r_b^h)^2}{4\pi} \tag{5.14}$$

In the left side of the equation, $\frac{\beta}{2\pi}$ is the conditional probability. The part inside



(a) Interference



(b) No Interference

Figure 5.4: Interference Model for Directional Antennas

the square brackets is the area of the sector (transmission area). Changing the inequality (5.5) with this conditional interference zone area we get:

$$\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h(b)} 1(\text{The } h\text{th hop of bit } b \text{ is over subchannel } m \text{ in slot } s) \frac{\alpha\beta(1+\Delta)^2 (r_b^h)^2}{4\pi} \leq W_m \tau \tag{5.15}$$

Following the steps from (5.6) to (5.5), we get the transport capacity using directional antennas at both transmitter and receiver's sides:

$$T_{DD} = \sqrt{\frac{2\pi}{\alpha\beta}} \frac{1}{(1+\Delta)} W\sqrt{n} \text{ bit-meters per second.} \tag{5.16}$$

Compare this capacity with what we get for the omnidirectional case in (5.5), the capacity gain is $\frac{2\pi}{\sqrt{\alpha\beta}}$, which is more than the gain using directional antennas only as transmitters.

We have skipped the omnidirectional transmission and directional reception case before this subsection, because the derivation of case directional transmission and reception has already included that of this case. The sender-based interference zone is a circle and the receivers are considered as sectors. The conditional interference area is $\frac{\beta}{2\pi}$ times the area of the disk, so the capacity gain is $\sqrt{\frac{2\pi}{\beta}}$.

If we consider antenna beamwidth as constant independent of n, the results are still under the ambit of the results obtained in [12] and are not really improvements in the sense of improving the order of the capacity. However, as the angles α and β both become smaller, the transport capacity increases. The network transport capacity is $O(W\sqrt{\frac{n}{\alpha\beta}})$. Then each node will obtain a throughput capacity of $O(\frac{W}{\sqrt{n\alpha\beta}})$ bit-meters per second. When the beamwidth approaches 0, the wireless networks can be seen as the wired links.

In the argument of the directional reception case, probabilistic analysis of the interference region is used to evaluate the average capacity of the networks based on the placement of nodes optimized by omnidirectional antennas. This is because that the location of the nodes, source and destination are arbitrary, but not the direction of the antennas. So the results show a further improvement in capacity of ad hoc wireless networks using directional antennas. The results should not be considered as an *upper* bounds on capacity, but rather as *lower* bounds on the potential capacity *improvement* by using directional antennas.

From the view of scalability of networks, if the product of beamwidths $\alpha\beta$ decreases asymptotically as fast as $\frac{1}{n}$, the per node capacity will be unrelated with the number of nodes in the domain; thus, the maximum throughput capacity can scale and becomes a constant which is no longer $O(\frac{1}{\sqrt{n}})$, even though this result is based on the conservative interference model. So in this sense there is no limitation for the sender-based model we use.

5.3.5 Achievability of the Capacity Improvement

5.3.5.1 Will the capacity grow arbitrarily high?

A question may arise when we look at the result of the capacity of the directional antennas, for example, capacity improvement indicated in (5.16). What will happen when the beamwidth α or β approaches to zero? Will the capacity grow arbitrarily high? The answer is no.

When the antenna beamwidth reaches a specific threshold, such that the transmission range conducted by all the transmitters just cover the whole domain, the per-node throughput will achieve a constant number related to W. This constant should be less or equal to $W/2$, because the wireless network link is half-duplex. When the angles of antenna beam get even smaller, the capacity will not increase any more. When we go through the inequalities from (5.10) to (5.5), equality can be achieved under some conditions for each inequality. (For example, for inequality (5.10), the equality can be achieved if and only if the routes for each source-destination pair are along a straight line.) So the capacity improvement for directional antennas is feasible. Now look at (5.12) and (5.15) where antenna angles are involved. When the angle α or β is too small, the aggregate transmission range conducted by all the transmitters will not cover the whole domain. Thus equality cannot hold due to the small area of the interference zone. That is to say, the interference has been fully reduced and we cannot get any improvement by narrowing the beam of the antennas.

Figure 5.5 presents a numerical example to illustrate the scaling law of the network capacity (disregarding the constant factor) when the antenna beamwidth is in different orders of the number of nodes n in network. Assuming $\alpha = \beta$, when α decreases slower than $\frac{1}{\sqrt{n}}$, the scaling law in (5.16) holds. The faster α decreases, the slower the per-node throughput decreases with the increase of the number of nodes in the domain. When α decreases as fast as $\frac{1}{\sqrt{n}}$, the throughput will be $O(W)$. That is the best result we can get with directional antennas; there is no change even if α decreases faster than $\frac{1}{\sqrt{n}}$. In this case, (5.16) no longer holds. The positive aspect for this is that the antenna beamwidths are not necessarily near zero to get high throughput. The threshold of α or β is a function of n.

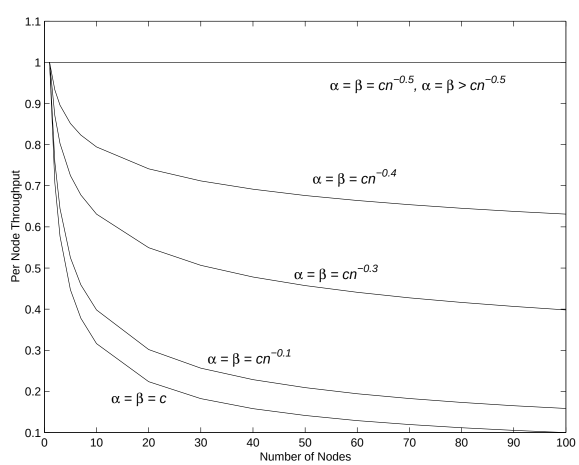


Figure 5.5: Per-node Throughput as a Function of n

5.3.5.2 The important role of transmission range

Recall the two reasons we mentioned before that why the network capacity cannot scale linearly: interference and hops. The small beamwidth can be used to reduce interference, and the number of hops can be kept down from $O(\sqrt{n})$ by increasing transmission range, which is one of the advantages of directional antennas. In our analysis for arbitrary networks, the use of optimal transmission range is assumed. In order to get the nondecreasing throughput when n increases, the transmission range should be kept large enough to reach the destination in hops of $O(1)$. This is consistent with result in [8] that the transmission range should be a constant independent of n in order to get linear maximum stable throughput in the number of nodes, although their main claim doesn't lie on this. The reason that constant transmission range is undesirable in [8] is that constant transmission range is impractical. On the contrary, we show that it is achievable, for constant transmission range doesn't necessarily require constant transmission power. Here we take a few words to state that in fact antenna property can make this requirement on transmission range met.

The antenna model we use so far is in 2 dimension. To analyze the relation between beamwidth and the antenna gain, we define the antenna as like a cone in a 3-dimensional view, as seen in Figure 5.6. Considering a sphere of radius R, the surface area A on the sphere for a beamwidth α can be approximated as a circle of radius $R\tan\frac{\alpha}{2}$. Let S be the surface area of the sphere, and P the emanated power. By definition, antenna gain, also known as *directivity*, D can be given by the following:

$$D = \frac{P/A}{P/S} = \frac{S}{A} = \frac{4\pi R^2}{\pi R^2 \tan^2\frac{a}{2}} = \frac{4}{\tan^2\frac{a}{2}} \tag{5.17}$$

According to radio propagation equation [1] we have:

$$P_{RV}(dB) = P_{TX}(dB) + K + G_{RV} + G_{TX} - 10\gamma\log_{10} d \tag{5.18}$$

Where G_{RV} and G_{TX} are the receiver and transmitter antenna gains in dB. P_{RV} and P_{TX} are the received and transmit power respectively in dBm. K is a constant dependent on the environment and wavelength and γ is the path loss exponent. The

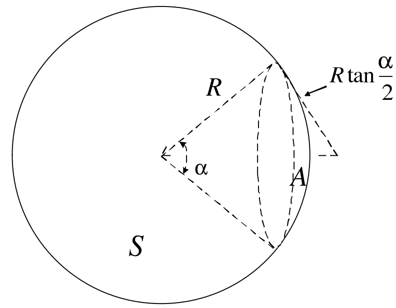


Figure 5.6: A 3-Dimensional View of Directional Antenna Pattern

path loss exponent is dependent on the environment. For free space, γ is 2 and for mobile environment γ is usually 4 (two ray ground reflection). $\gamma = 4$ is used in this chapter.

Transmission distance (range) $r \sim O(\sqrt[4]{D}) = O(\sqrt{\frac{2}{\tan\frac{\alpha}{2}}}) = O(\frac{1}{\sqrt{\alpha}})$ when $\alpha \to 0$. Recall that α need to be $O(\frac{1}{n})$ to get non-varnishing per-node throughput (refer to equation (5.13)). Thus $r \sim O(\frac{1}{\sqrt{\alpha}}) = O(\sqrt{n})$, which means transmission range will naturally increase as fast as $\sqrt{n}$ when beamwidth decrease as fast as $\frac{1}{\sqrt{n}}$. Note here the transmission range calculated is in the unit of reality case. When converted to the unit domain area (normalized) case, r becomes $O(1)$, which is exactly what we want. Similar result can be obtained for other antenna modes.

The interference model we use is a protocol model in which Δ reflects the fact that the interference range is always larger than the transmission range. The results for the transport capacity support the intuitive expectation that larger guard zone (Δ) leads to lower capacity. Shorter transmission range (short hop routing) is favored traditionally to reduce energy consumption and interference. The statement about energy consumption is true in concept, since power attenuates exponentially with distance. Yet it does not consider the the energy consumption in the transmit/receive electronics circuitry. Moreover when short hop forwarding is used, there is routing overhead, end-to-end delay, end-to-end reliability, and so on. In regard to the interference caused by longer transmission distance, the use of directional antennas can change the spatial reuse characteristics, for transmission coverage does not inherently cause interference. Besides, narrow beam has a small coverage even it reaches a long distance.

Figure 5.7 represents an example to illustrate how the capacity is improved when directional antennas are used at both transmitter and receiver side. The dotted circles are transmitters and the solid circles are receivers. The interference is eliminated as far as no receiver is inside other's transmission range. Even when a receiver is inside an unwanted transmission range, it can adjust its receiving pattern to avoid the interference. Also the transmission range should be large enough to reduce the node forwarding burden. When other technique is used to get the antenna pattern to a straight line, the two dimension interference area becomes a one dimension line, and the throughput problem can be analyzed as the one dimensional capacity problem. This gives an extreme case for a theoretical analysis and may have some limits in practice. However, for a fixed n or fixed area domain, this result can give guidelines and implications on the design of the networks.

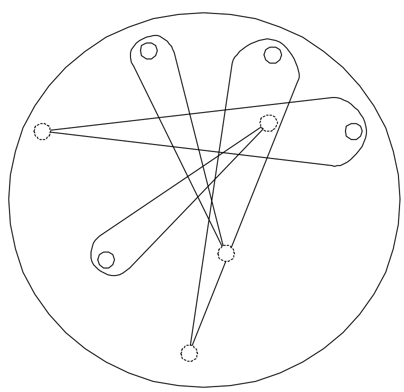


Figure 5.7: An Example of Using Directional Transmission and Directional Reception

5.4 The Throughput Capacity for Random Networks

Though the setting of the problem of random networks is very different from that of arbitrary networks, the analysis of capacity improvement by using directional antennas is very similar.

We use the important concept of Voronoi tessellation whose definition is: the partitioning of a plane with n points into n convex polygons such that each polygon contains exactly one point and every point in a given polygon is closer to its central point than to any central point of other polygons. An example of Voronoi tessellation is shown in Figure 5.8.

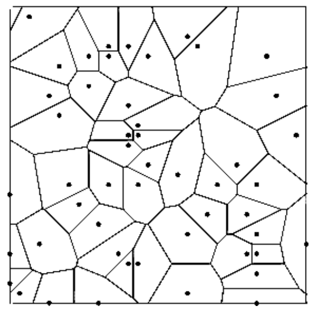


Figure 5.8: Voronoi Tessellation

We briefly repeat the steps in [12] for completeness and omit the proof for most of the lemmas.

Definition: Adjacent Cells: Two cells are called *adjacent*, if they share a common point.

Lemma 1 *We can construct a Voronoi tessellation V_n in relation to the number of nodes n and the locations of nodes. In this tessellation:*

(V1) *Every Voronoi cell contains a disk of area $100 \log n/n$. Let*

$$\rho(n) := \text{radius of a disk of area} \frac{100 \log n}{n} \tag{5.19}$$

(V2) *Every Voronoi cell is contained in a disk of radius $2\rho(n)$.*

(V3) *Each Voronoi cell contains at least one node.*

(V4) *We can choose the range $r(n)$ of each transmission such that*

$$r(n) = 8\rho(n)$$

This range allows direct communication within a cell and between adjacent cells. Every node in a cell is within this distance $r(n)$ from every other node in its own cell or adjacent cell.

We assume that power control is not used and the beam of each directional antenna can be steered to it's intended sender or receiver. So the transmission range needed for connectivity is the same as that of omnidirectional antennas. Thus the Voronoi tessellation V_n in Lemma 1 is also usable for directional antennas.

Definition: Interfering Neighbors: Two cells are said to be *interfering neighbors* if there is a point in one cell which is within a distance $2(1+\Delta)r(n)$ of some point in the other cell.

One important property of the constructed Voronoi tessellation is that the number of interfering neighbors of a cell is *uniformly bounded.* This property is the key factor for analyzing the capacity gain of using directional antennas.

Lemma 2 *For the case that all the antennas are omnidirectional, every cell in V_n has no more than c_1 interfering neighbors. c_1 depends only on Δ and grows no faster than linearly in $(1+\Delta)^2$.*

Proof: Let V be a Voronoi cell. If V' is an interfering neighboring Voronoi cell, there must be two points, one in V and the other in V', which are no more than $2(1+\Delta)r(n)$ units apart. From (V2), the diameter of a cell is bounded by $4\rho(n)$. Hence V', and similarly every other interfering neighbor in the Interference Model, must be contained within a common large disk D of radius $6\rho + 2(1+\Delta)r(n)$.

Such a disk D cannot contain more than $c_2 = \frac{(6\rho+2(1+\Delta)r(n))^2}{\rho^2(n)} = (22+16\Delta)^2 \sim O((1+\Delta)^2)$ disks of radius $\rho(n)$. By (V2), there can therefore be no more than $c_1 = c_2 - 1$ cells within D. This therefore is an upper bound on the number of interfering neighbors of the cell V.

How can we extend this lemma to the case that antennas are directional? Considering the interference zone introduced in section 5.3, the same concept can be utilized here. First we must consider how to assess the number of interfering neighbors when the transmission antennas are directional. Whatever shape the transmission pattern is, from the perspective of interference and spatial reuse, the aspect that matters is the area of the interference zone. The nodes are randomly, i.e., independently and uniformly, located in the domain. So the number of interference neighbors is proportional to the area of the interference zone.

Like the antenna model we used to analyze arbitrary networks, we also assume the antennas are sectorized with the same parameters used in Section 5.3. So the area of interference zone for directional transmission is $\frac{\alpha}{2\pi}\pi r^2(n)$. So, virtually the interfering neighbors should not exceed $c_3 = \frac{\alpha}{2\pi}c_1$.

Similarly, when omnidirectional transmission and directional reception are used, the area of conditional interference zone derived using probability becomes $\frac{\beta}{2\pi}\pi r^2(n)$. Correspondingly the number of interfering neighbors is bounded by $c_4 = \frac{\beta}{2\pi}c_1$. Again, when antenna transmission and reception are both directional, the area of conditional interference zone is $\frac{\alpha}{2\pi}\frac{\beta}{2\pi}\pi r^2(n)$ and the number of interfering neighbors is no more than $c_5 = \frac{\alpha\beta}{4\pi^2}c_1$.

Lemma 3 *Let c_6 be the number of interfering neighbors for each cell in tessellation V_n. There is a schedule for transmitting packets such that in every $(1 + c_6)$ slots, each cell in the V_n gets one slot in which to transmit, and such that all transmissions are successfully received within the transmission and reception coverage.*

Each node wishes to communicate with the node nearest to a randomly chosen location. Routing strategy is to choose the routes of packets to approximate the straight-line which is connecting the source and destination. So the routes actually are the cells the straight-line intersects.

Lemma 4 *There is a $\delta'(n) \to 0$ such that*

$$\mathbf{P}(\sup_{V \in V_n}(\textit{number of lines } L_i \textit{ intersection } V) \leq c_7\sqrt{n \log n}) \geq 1 - \delta'(n)$$

Note the final destination forwarding inside a cell is at most one hop away. The traffic is relayed by the cells intersected by the routing straight-line. Hence the traffic handled by a cell is proportional to the number of lines passing through it. Since each line carries traffic of rate $\lambda(n)$ bits pe second, the following bound can be obtained.

Lemma 5 *There is a $\delta'(n) \to 0$ such that*

$$\mathbf{P}(\sup_{V \in V_n}(\textit{Traffic needing to be carried by cell } V) \leq c_7\lambda(n)\sqrt{n \log n}) \geq 1 - \delta'(n)$$

From Lemma 3 we know that there exists a schedule for transmitting packets such that in every $(1+c_6)$ slots, each cell in tessellation V_n gets one slot to transmit. Thus the rate at which each cell gets to transmit is $W/(1+c_6)$ bits per second.

On the other hand, the rate at which each cell needs to transmit is less than $c_7\lambda(n)\sqrt{n\log n}$ with high probability. This rate can be accommodated by all cells if it is less than the rate available, i.e., if

$$c_7\lambda(n)\sqrt{n\log n} \leq \frac{W}{1+c_6} \tag{5.20}$$

So we have proved the following theorem, noting the linear growth of c_1 in $(1+\Delta)^2$ in Lemma 2.

Theorem 2 *For random networks, there is a deterministic constant $c > 0$ not depending on n, Δ, or W, such that*

$$\lambda(n) = \begin{cases} \frac{cW}{(1+\Delta)^2\sqrt{n\log n}}, & \textit{Omni Tx Omni Rv}, \\ \frac{2\pi}{\alpha}\frac{cW}{(1+\Delta)^2\sqrt{n\log n}}, & \textit{Dir Tx Omni Rv}, \\ \frac{2\pi}{\beta}\frac{cW}{(1+\Delta)^2\sqrt{n\log n}}, & \textit{Omni Tx Dir Rv}, \\ \frac{4\pi^2}{\alpha\beta}\frac{cW}{(1+\Delta)^2\sqrt{n\log n}}, & \textit{Dir Tx Dir Rv}. \end{cases}$$

bit per second is feasible with high probability. Here we use the easy-to-read abbreviation: Omni = omnidirectional; Dir = directional; Tx = transmission mode; Rv = reception mode.

Comparing the different results among each antenna mode, we can clearly see the throughput gain factor when directional antennas are used. We get an ideal throughput gain of $\frac{4\pi^2}{\alpha\beta}$ using directional transmission and reception.

For random networks, unlike arbitrary networks, the per-node throughput will *not* be a constant with the increasing of the network size if minimum power levels are chosen for each directional antenna. The multi-hop burden still exists in that the source and the destination may be far away from each other. In general directional antennas have the potential to reduce the multi-hop problem by per-node power control. Ideally the transmission range may be far enough to reach any node the sender wants to communicate with. So only when the antenna beamwidths are small enough and the transmission range is far enough, can the throughput capacity be a constant not exceeding $W/2$.

Note even the gain factor of capacity by using directional antennas is larger for random networks than for arbitrary networks, the scaling term determines the absolute capacity of arbitrary networks and random networks.

5.5 Hybrid Antenna Model

5.5.1 Interference Area

To achieve a better model of real directional antennas, now we use an antenna model whose beamforming patterns are a mix of omnidirectional and directional antenna models. We define *hybrid antenna model* as an antenna model whose main lobe is characterized as a sector and whose sidelobes and backlobes together form a circle. Shown in Figure 5.9, the antenna pattern has a gain value g_m for the main lobe of beamwidth α. It also has a sidelobe of gain g_s of beamwidth $(2\pi - \alpha)$. (Use β if it is a receiver). In the 3-dimension space, the radiation pattern is like a conical main lobe plus a bulb-shaped sidelobe at the base of the cone.

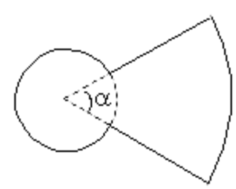


Figure 5.9: Radiation Pattern of Hybrid Antenna Model

In our work, the transmission range and the beamwidth are determined by the network. So g_m is also determined since it is dependent on the transmission range and vice versa. To estimate what is the radius of the sidelobe given the gain and the beamwidth of the main lobe, first we use the result from [16]. Given g_m and α, we have

$$g_s = \frac{\eta D - g_m}{D - 1} \tag{5.21}$$

where $D = \frac{4}{tan^2 \frac{\alpha}{2}}$ is the directivity of the antenna, and η is the efficiency of the antenna which accounts for losses.

The radius (transmission range) of the sidelobe can be calculated using the path loss equation 5.18.

From the path loss equation, we can also estimate the ratio of the radius of the sidelobe to that of the main lobe. For example, if $g_s/g_m = x$, G_{TX} becomes $G_{TX} + 10\log_{10} x$. To maintain equality, d changes to $dx^{\frac{1}{\gamma}} = d\sqrt[4]{x}$. So we can get the transmission range of the sidelobe given the transmission range and the beamwidth of the main lobe. Let's further define the radius of the sidelobe of the transmitter as $s_\alpha \cdot r$ and the radius of the sidelobe of the receiver as $s_\beta \cdot r$.

Now we come to the calculation of the area of the interference zone for each combination of antenna modes. Define the interference zone area of the omnidirectional antennas as $A_{OO} = \pi r^2$.

For directional transmission and omnidirectional reception, the interference zone area is the area of the hybrid antenna radiation pattern, which is calculated as:

$$A_{DO} = \frac{\alpha}{2} r^2 + \frac{2\pi - \alpha}{2} s_\alpha^2 r^2 = A_{OO} \frac{\alpha + (2\pi - \alpha) s_\alpha^2}{2\pi} \tag{5.22}$$

For omnidirectional transmission and directional reception, the conditional interference area is the transmission area multiplied by the probability that the nodes inside the transmission range will experience interference. This conditonal interference area is calculated as:

$$\begin{aligned} A_{OD} &= A_{OO}(\mathbf{P}\{|Tx - Rv| \leq s_\beta r\} + \mathbf{P}\{|Tx - Rv| > s_\beta r\} \cdot \mathbf{P}\{\text{main lobe of Rv} \\ &\quad \text{pointing to } Tx\}) = A_{OO}(s_\beta^2 + (1 - s_\beta^2)\frac{\beta}{2\pi}) \end{aligned} \tag{5.23}$$

For directional transmission and directional reception, the case is more complicated. We divide the transmission area into two parts: one is a small circle with radius $s_\alpha r$, and the other is an annulus sector with radii $s_\alpha r$ and r, shown in Figure 5.10. We then separate the conditional probability into two parts according to where the receiver resides. Since r is fixed, for the same total energy of the directional antenna, the larger the beamwidth, the smaller the gain of the sidelobe. So first consider $\alpha > \beta$, then $s_\alpha < s_\beta$. In this case, if the receiver is inside the transmitter's small circle, it will get interfered since the sidelobes/backlobes of the receiver will cover the transmitter (note $s_\alpha < s_\beta$). Otherwise if the receiver is inside the annulus sector, there are two possibilities for this receiver to get interfered: sender and receiver are so close that the receiver's sidelobes/backlobes cover the sender; they are not close enough but the receiver's main lobe points to the sender. Accordingly the conditional interference area is:

$$\begin{aligned} A_{DD} &= \pi s_\alpha^2 r^2 + \frac{\alpha}{2}(1-s_\alpha^2)r^2(\mathbf{P}\{|Tx-Rv| \leq_\beta r|C_1\} + \mathbf{P}\{|Tx-Rv| > s_\beta r|C_1\}\frac{\beta}{2\pi}) \\ &= \pi s_\alpha^2 r^2 + \frac{\alpha}{2}(1-s_\alpha^2)r^2(\frac{s_\beta^2 - s_\alpha^2}{1-s_\alpha^2} + \frac{1-s_\beta^2}{1-s_\alpha^2}\cdot\frac{\beta}{2\pi}) \\ &= A_{OO}(s_\alpha^2 + \frac{\alpha}{2\pi}(s_\beta^2 - s_\alpha^2) + \frac{\alpha\beta}{4\pi^2}(1-s_\beta^2)) \end{aligned} \tag{5.24}$$

where C_1 is the condition statement "Rv is inside the annulus sector". Note $\frac{\alpha}{2}(1-s_\alpha^2)r^2$ is the area of the annulus sector.

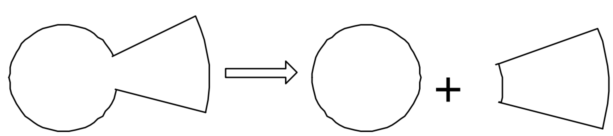


Figure 5.10: Partition the Radiation Pattern

Then consider the opposite $\alpha < \beta$, i.e. $s_\alpha > s_\beta$. If the receiver resides in the small circle of the transmission coverage, similarly, the sidelobes/backlobes of the receiver may or may not cover the transmitter, depending on how close they are. If the receiver is in the annulus sector, the probability of interference is $\beta/2\pi$, the probability the main lobe of receiver pointing to the sender. Consequently the conditional interference area is:

$$\begin{aligned} A_{DD} &= \pi s_\alpha^2 r^2(\mathbf{P}\{|Tx-Rv| \leq s_\beta r|C_2\} + \mathbf{P}\{|Tx-Rv| > s_\beta r|C_2\}\frac{\beta}{2\pi}) + \frac{\alpha}{2}(1-s_\alpha^2)r^2\frac{\beta}{2\pi} \\ &= \pi s_\alpha^2 r^2(\frac{s_\beta^2}{s_\alpha^2} + (1-\frac{s_\beta^2}{s_\alpha^2})\frac{\beta}{2\pi}) + \frac{\alpha}{2}(1-s_\alpha^2)r^2\frac{\beta}{2\pi} \\ &= A_{OO}(s_\beta^2 + \frac{\beta}{2\pi}(s_\alpha^2 - s_\beta^2) + \frac{\alpha\beta}{4\pi^2}(1-s_\alpha^2)) \end{aligned} \tag{5.25}$$

where C_2 is the condition statement "Rv is inside the small circle".

For easy reference, we combine results (5.24) and (5.25) into one:

$$A_{DD} = A_{OO}(\min(s_\alpha^2, s_\beta^2) + \frac{\max(\alpha,\beta)}{2\pi}|s_\alpha^2 - s_\beta^2| + \frac{\alpha\beta}{4\pi^2}(1-\max(s_\alpha^2, s_\beta^2))) \quad (5.26)$$

regardless of the relationship between α and β.

The special case where $\alpha = \beta$ lets the conditional interference area become $A_{DD} = A_{OO}(s_\alpha^2 + \frac{\alpha^2}{4\pi^2}(1-s_\alpha^2))$.

5.5.2 Arbitrary Networks

With the interference area given for each antenna mode taking the effect of the sidelobes and backlobes into consideration, we may go back to the equations in Section 5.3 on how the transport capacity is given. Obviously no change will be made when all nodes use omnidirectional antennas. Consider the directional transmission and omnidirectional reception, then (5.5) becomes:

$$\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h(b)} \pi(1+\Delta)^2(r_b^h)^2\frac{\alpha+(2\pi-\alpha)s_\alpha^2}{2\pi} = \sum_{b=1}^{\lambda nT}\sum_{h=1}^{h(b)} \pi(1+\Delta)^2(r_b^h)^2\frac{A_{DO}}{A_{OO}} \leq W_m\tau \quad (5.27)$$

Likewise we have similar inequalities for the other combinations of antenna modes. Follow the procedure in Section 5.3 will give the transport capacity of the arbitrary networks for all cases:

$$\lambda n\bar{L} \leq \begin{cases} \frac{1}{\sqrt{2\pi}}\frac{1}{(1+\Delta)}W\sqrt{n}, & \text{Omni Tx Omni Rv,} \\ \frac{1}{\sqrt{2\pi}}\sqrt{\frac{A_{OO}}{A_{DO}}}\frac{1}{(1+\Delta)}W\sqrt{n}, & \text{Dir Tx Omni Rv,} \\ \frac{1}{\sqrt{2\pi}}\sqrt{\frac{A_{OO}}{A_{OD}}}\frac{1}{(1+\Delta)}W\sqrt{n}, & \text{Omni Tx Dir Rv,} \\ \frac{1}{\sqrt{2\pi}}\sqrt{\frac{A_{OO}}{A_{DD}}}\frac{1}{(1+\Delta)}W\sqrt{n}, & \text{Dir Tx Dir Rv} \end{cases}$$

bit-meters per second.

Here the gain factor for each case can be derived from (5.22), (5.23), and (5.26). This result is based on the hybrid antenna model, which considers the effects of sidelobes and backlobes of directional antennas.

5.5.3 Random Networks

The hybrid antenna radiation pattern can be applied to random networks as well. Using the concept of interfering neighbors, Theorem 2 may be extended to Theorem 3 representing the hybrid antenna model.

Theorem 3 *For random networks, there is a deterministic constant $c > 0$ not depending on n, Δ, or W, such that*

$$\lambda(n) = \begin{cases} \frac{cW}{(1+\Delta)^2\sqrt{n\log n}}, & \textit{Omni Tx Omni Rv}, \\ \frac{A_{OO}}{A_{DO}}\frac{cW}{(1+\Delta)^2\sqrt{n\log n}}, & \textit{Dir Tx Omni Rv}, \\ \frac{A_{OO}}{A_{OD}}\frac{cW}{(1+\Delta)^2\sqrt{n\log n}}, & \textit{Omni Tx Dir Rv}, \\ \frac{A_{OO}}{A_{DD}}\frac{cW}{(1+\Delta)^2\sqrt{n\log n}}, & \textit{Dir Tx Dir Rv}. \end{cases}$$

bit per second is feasible with high probability.

Note that A_{DO}, A_{OD}, and A_{DD} are all monotone increasing functions of s_α and s_β. This shows that smaller antenna sidelobes result in higher throughput gain.

5.6 Simulation Validation

Numerous works have been conducted on the performance of the wireless networks using directional antennas. However, few tackle the problem of how the scaling behavior changes with the deployment of directional antennas instead of omnidirectional ones. The challenge lies in several bases: the big gap between the theoretical conclusions and the experimental results due to the inefficient protocol design, overhead etc., the difficulty in building a simulation platform for directional antennas, and the lack of a universal protocol stack for directional antennas. In this section, we would like to introduce a simple simulation mechanism to validate our analytical results on the capacity scaling of the networks using directional antennas.

We use end-to-end per-node throughput in our analysis for random network. This result comes from the assumption on the straight line routing and the optimal global scheduling. To avoid the effect of the routing on the throughput, we choose to evaluate the aggregate one-hop throughput instead of the end-to-end average per-node throughput. In this case, we are only interested in the aggregate one-hop throughput in a MAC view, where the MAC scheduler tries to achieve the maximum transmissions.

Here is the random network set up: n nodes are randomly distributed in a unit area disk, as in Figure 5.11. Since our interest is in studying the highest possible raw throughput of the MAC layer, we assume no protocol-specific MAC overheads such as RTS/CTS and ACK frames, and no backoff windows used in the contention resolution phase. All the nodes in a network, whether equipped with omnidirectional or directional antenna, have the same transmission range. As a MAC layer simulation, we focus on the single hop traffic, and assumes no statistical dependency between the traffic originating at different nodes. The MAC layer manages traffic on a stand-alone basis in the sense that it does not depend on routing or queueing strategies.

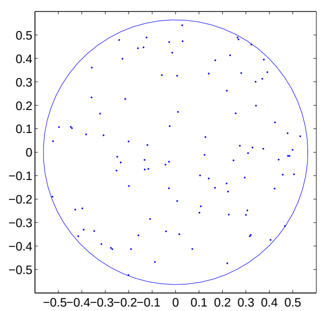


Figure 5.11: A Random Network Topology: 100 Nodes in a Unit Area Disk

While how to give a distributed algorithm to optimize the scheduling is an open question, we propose a simple distributed MAC scheduling to approximate the global optimization: the MAC scheduler randomly chooses a node in the network and thus the node can transmit at its maximal rate to a neighbor in its transmission range, the scheduler then randomly selects another node as a potential sender and check if this transmission would interferes any of the scheduled transmissions, if so, do not send, otherwise allow this transmission and move on until all the nodes in the network are checked. The simultaneous transmission pairs are then counted in this snapshot of the MAC time slot. Note this algorithm is not a MAC protocol to serve the application requirements, instead its purpose is to statistically approach the single-hop capacity of the network.

Monte Carlo simulations are performed by MATLAB to analyze the random network of nodes from 100 to 1000. The transmission range should scale as $\Theta(\frac{\log n}{n})$ to guarantee the connectivity of the network [109]. The exact number for it depends on the network size. We choose $0.56\frac{\log n}{n}$ as our rule of thumb data for networks with 100 to 1000 nodes. Since the transmitting nodes use the full bandwidth, the number of simultaneous transmission pairs is proportional to the aggregate one-hop throughput of the network. Due to the huge amount of data obtained and limited space, we only present the results for cases where both transmitters and receivers are omnidirectional or both are directional with the same antenna beamwidth ($\alpha = \beta$). The antenna model used is the simple sector-based model. All the results are averaged among 100 runs of different node distributions. Figure 5.12 plots the capacity as a function of network size n, fitted by $O(\sqrt{\frac{n}{\log n}})$, the predicted aggregate capacity. The error bars show the data within 95% confidence interval around the mean value, the pointed data.

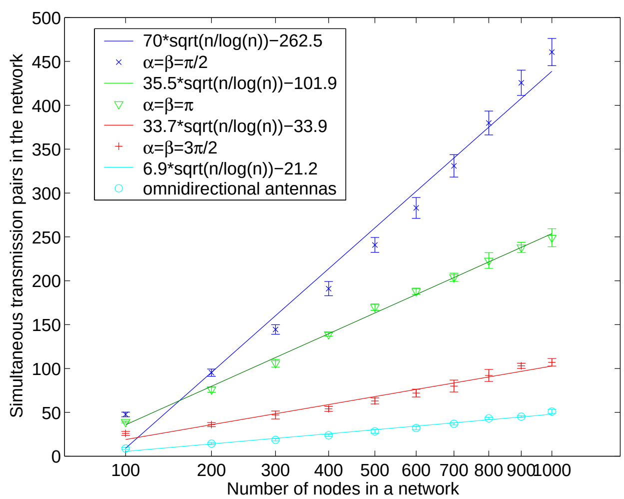


Figure 5.12: Number of Simultaneous Transmission Pairs vs. Number of Nodes in a Network, Fitted by $O(\sqrt{\frac{n}{\log n}})$. X-axis Uses a $\sqrt{\frac{n}{\log n}}$ Scale. The Error Bars Show the 95% Confidence Interval.

The lowest line is the fitted network aggregate capacity scaling law of network with omnidirectional antennas. The samples agree with the line quite well, which shows that our simplification in simulation and the MAC algorithm does not affect the correctness of the results in terms of the capacity scaling property. The other three lines are fitted for the samples from networks with directional antennas with different beamwidths. With fixed beamwidth, the capacity of the network obeys the scale rule $O(\sqrt{\frac{n}{\log n}})$ given in Section 5.4. The different slopes of the lines indicate that the throughput gain is a function of antenna beamwidth.

In order to look at how the beamwidth affects the capacity, we plot the results from a different dimension. From Theorem 1, we know that the capacity increase at the order $O(\frac{1}{\alpha\beta})$ when n is fixed. Since $\alpha = \beta$ in simulations, the capacity should scale as $O(\frac{1}{\alpha^2})$. Figure 5.13 illustrates the capacity from a different view, the number of simultaneous transmission pairs as a function of antenna beamwidth α, fitted by $O(\frac{1}{\alpha^2})$. The error bars show the data within 95% confidence interval around the mean value, the pointed data. Three fitted lines are for different network size $n = 100$, 500 and 1000, respectively. These lines are fitted using only data for beamwidth less than or equal to $\frac{5\pi}{6}$. If beamwidth gets smaller, the network will enter the saturation state, and the capacity cannot increase anymore. This feature has been discussed before in Section 5.3.5. We are aware that if network size gets larger than 1000, some of the parameters involved will change. For instance, when the beamwidth is about $\frac{\pi}{12}$ the network approaches to the saturation state, when networks with 10000 nodes are simulated. While there is no clear definition on how large the network should be to meet the scaling property, we choose the range of network size mainly for practical purpose. Simulation results show that the results match the theorem pretty well, supporting the conclusion on the capacity improvement with the use of directional antennas.

5.7 Applications of Directional Antennas

In this section, we would like to touch upon some application issues associated with the use of smart directional antennas. The protocol or design we propose here is only some simple sketch and hypothetical. We include this part as some foundation that may advise future implementation.

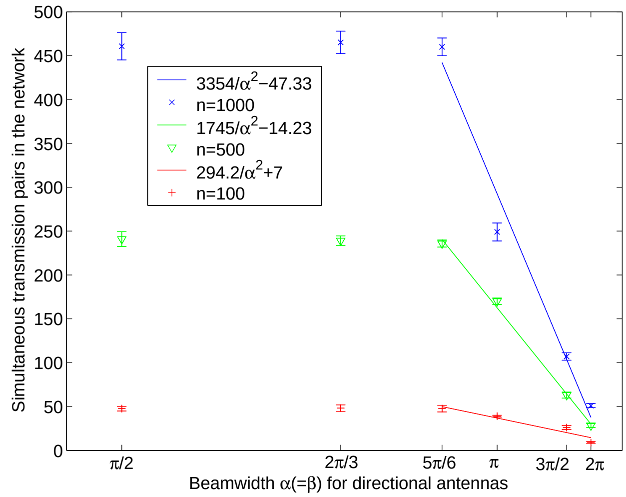


Figure 5.13: Number of Simultaneous Transmission Pairs vs. Antenna Beamwidth α, Fitted by $O(\frac{1}{\alpha^2})$. X-axis Uses a $\frac{1}{\alpha^2}$ Scale. The Error Bars Show the 95% Confidence Interval. The Isolated Points Indicate the Saturation of the Network Throughput.

5.7.1 Localization

One common problem for ad hoc networks is localization for each node's neighbors. The use of directional antennas in extending coverage range and capacity of wireless networks dictates the employment of novel media access control protocols, with which for example mobile nodes provides access to other nodes by learning their locations. Steerable antenna can potentially provide the advantage for this purpose of positioning. The access point or a mobile node can obtain information about the location of another node by using beamforming and steering method to acquire the spatial signature of each user. The objective of the localization protocol is to locate all users as fast as possible. Such protocols allow rapid media access and can be embedded in existing MAC protocols.

Switched Beam, steerable or windowing directional antenna can potentially provide the advantage for positioning. The positioning capability depends on the speed the beam can be switched or steered. The access point or a mobile node can obtain information about the location of another node by using the following method. It first sends a message starting from the first beam. Then it waits for the node in this direction to respond. Upon receiving the message, a node responds by sending back an acknowledgement. The first node then stores the location (or beam direction) of that node. After that or a timeout indicating there is no node residing in this direction, the antenna steers to the next beam, and likewise to cover the whole 360 degree. This steering/response method is used to acquire the spatial signature of each user.

The access point or mobile node sequentially steers the beam towards different directions, so that the entire space is covered. The objective of the localization protocol is to locate all users as fast as possible.

So far we suppose acknowledgement in each directional beam is contention free. We need to consider the situation that there are more than one node in some directions. In this case, some backoff mechanism can be used. The same directional beam of first node may have to send the message multiple times to locate all the neighbors. The space is successively scanned by a beam until all the neighbors are located and the procedure is repeated for all users equipped with directional antenna.

5.7.2 MAC Protocol for Ad Hoc Networks

Directional antennas can provide higher gain, and reduce interference by directing beamforms towards a desired direction and nulls towards undesired interferers. However, they also pose challenges in the design of medium access control (MAC) protocols. Recently, attention has been focused on the possibility of using directional antennas or multiple input multiple output (MIMO) links for medium access control in multi-hop networks [13, 14, 15, 16, 17, 104, 105, 110]. In principle, many of the proposed protocols are similar to IEEE 802.11, carefully adapted for use over directional antennas.

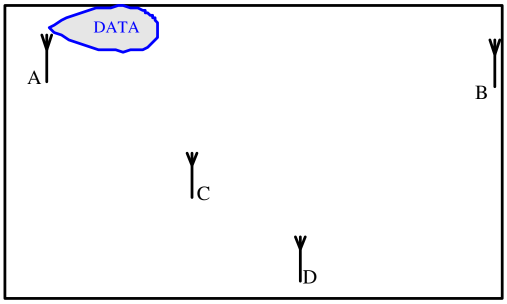
Due to the large amount of work done in the area of single beam directional antennas, here we only discuss some new features of the single-band multi-beam directional antennas in the MAC design. The property of adjustable directional beams of highly advanced antenna, like the plasma windowing antennas [111], gives high flexibility when MAC layer protocols are considered. With the activation or deactivation of plasma tubes at each direction, nodes can easily block or allow transmission/reception to/from any directions. This gives us virtually any combination of multi-beams with any angle. The following is a sketch of our proposed MAC protocol for ad hoc networks with the use of beamforming antennas. This protocol is a CSMA/CA and DCF (Distributed Coordination Function) based MAC protocol. We mainly talk about the omnidirectional or directional RTS/CTS exchange and ignore other parts, such as backoff mechanism and contention window etc.

In our proposed MAC protocol sketch, all RTS/CTS are sent using as many beams as possible. That is to say, the RTS/CTS are omnidirectional except that some beams are blocked because of the reception of previous RTS/CTS. The data and ACK are sent with the appropriate directional beam. The direction or location information is pre-configured in each node or stored in the CTS packet. One localization method is introduced in previous subsection. As shown in Figure 5.14(a), assume after successful RTS/CTS handshake, node A transmits data to node B using the beam pointing to B. In the mean time, all the nodes in the vicinity of node A and B will block the directions pointing to node A and node B and update the NAV (Network Allocation Vector) to the proper value. Now node C wants to talk to node D. Node C sends out RTS packet. The directions pointing to A and B are blocked, so the antenna pattern of node C is near omnidirectional with two direction nulled, see Figure 5.14(b). Node D replies with CTS packet. Similarly, the directions pointing to A and B are blocked, shown in Figure 5.14(c). In Figure 5.14(d), node C starts to send data to node D with a single beam towards node D.

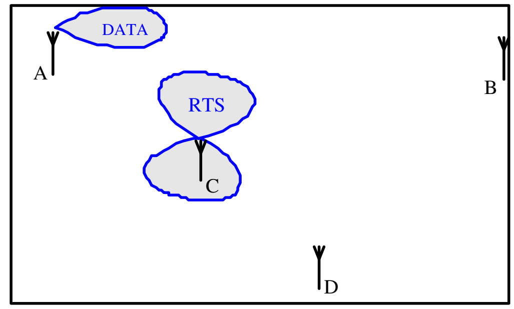
5.7.3 Directional Antenna for a Cluster-based Cooperative Network

In a cluster-based cooperative network, the multi-beam directional pattern can efficiently use multicast instead of broadcast to transmit to the relay cluster nodes. Therefore, the interference caused by the transmitter is greatly reduced and the spatial reuse is improved. Similarly, the receiver may choose the multiple beams to receive from the relay cluster and minimize the impact of overhearing unwanted signals. This property is illustrated in Figure 5.15.

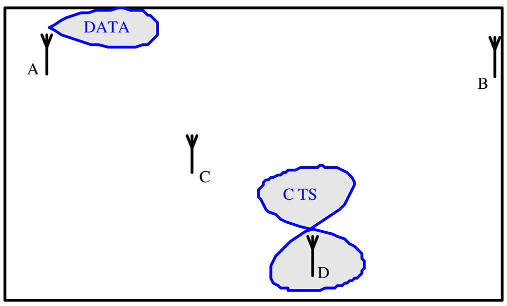
By using multi-beam directional antennas, a node may be able to selectively receive signals only from certain desired directions. This enables the receiver node



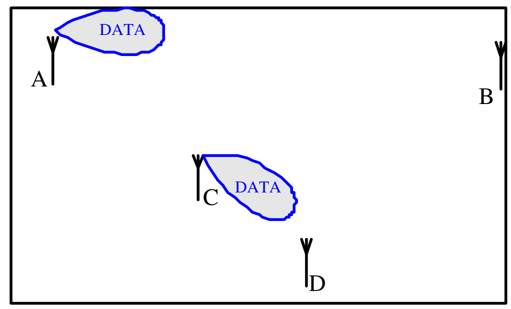
(a) Node A transmits data to node B using the beam pointing to B. Node C wants to talk to node D.



(b) Node C sends out RTS packet. The directions pointing to A and B are blocked, so the antenna pattern of node C is near omni-directional with two directions deactivated.



(c) Node D replies with CTS packet. Similarly, the directions pointing to A and B are blocked.



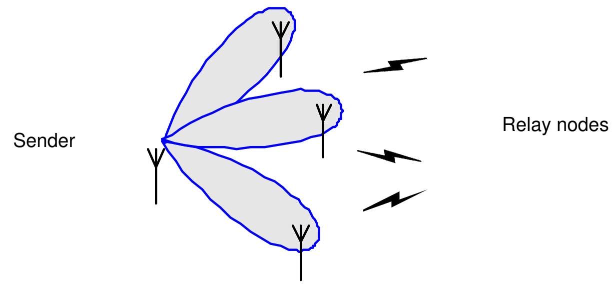
(d) Node C starts to send data to node D with a single beam pointing to node D.

Figure 5.14: A MAC Protocol Sketch Using Directional Antennas

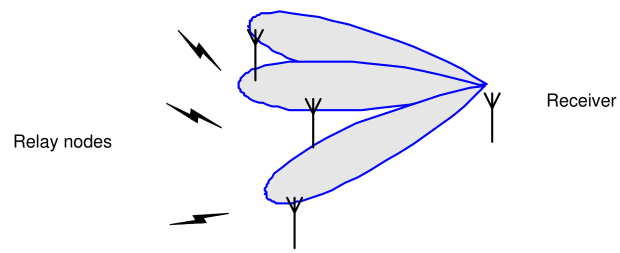
to avoid interference that comes from unwanted directions, thereby increasing the signal to interference and noise ratio (SINR).

Solutions which optimize along the many dimensions of possible smart antenna techniques and approaches also include [112]: architectures with multiple antennas at the base station, at the client device, or both (a.k.a. MIMO); designs targeting a variety of combinations of coherent gains, diversity gains, active interference cancellation/rejection, multiples of capacity through spatial channels; the full range of processing alternatives, from beam switching to adaptive beam-steering, active nulling, and SDMA; successful implementations for both TDD and FDD systems, for TDMA, CDMA, and OFDMA; and pursuit of equipment for wide-area wireless new standard like 802.16.

At the same time, unfortunately, directional transmissions cause some problems in ad hoc networks, such like hidden terminal problem, deafness. These prob-



(a) Sender uses multiple antenna beams to transmit information to the relay cluster nodes



(b) Receiver uses multiple antenna receiving beams to receive from the relay cluster nodes

Figure 5.15: Multi-beam Antennas in Cooperation to Achieve Transmit Diversity or Receive Diversity

lems are important when designing MAC protocols for directional antennas and they have been studied extensively by researchers [15, 113, 114].

5.8 Summary

Use of directional antenna in the context of ad hoc wireless networks can largely reduce radio interference, thereby improving the utilization of wireless medium. We study the capacity for different combination of antenna modes. The approach to get the capacity is mainly based on [12], but the interference model used is different.

Our work is focused on discovering the lower bounds in capacity *improvement* that directional antennas can provide relative to the traditional omnidirectional antennas. For instance, in arbitrary networks, with the reduction of the transmission area and the reduced probability of two neighbors pointing to each other, the capacity of networks using directional antennas will be improved by a factor of $\frac{2\pi}{\sqrt{\alpha\beta}}$. Here α and β are the beamwidths of transmission and receiving directional antennas, respectively. If the beamwidths of transmitting and receiving antennas decrease asymptotically as fast as $\frac{1}{\sqrt{n}}$, the capacity stays constant.

For random network, due to the reduction of interfering neighbors, the capacity with the use of directional antennas can achieve a gain as large as $\frac{4\pi^2}{\alpha\beta}$. The use of directional antennas can take advantage of decreasing both interference (local) and multi-hop relay burden (global) through the coordination of the transmission power and antenna directivity.

To model the sidelobes and backlobes of real directional antennas, a hybrid antenna model is used. By calculating the area of the interference zone, we get the capacity improvement gain factors for different transmission/reception modes. Simulation evaluation presented gives support for our analytical model and results on capacity improvement.

CHAPTER 6
Conclusions and Future Work

6.1 Conclusions

A lot of research has conducted the study on capacity of wireless networks, which for example includes capacity analysis for random networks, with or without mobile nodes, throughput and delay trade-off analysis. Although these works are diverse in terms of approaches and objectives, it is widely accepted that wireless ad hoc networks, especially multi-hop networks are under physical constraint of wireless nature, such as coarse communication environment, interference of neighboring nodes etc. Consequently the performance problem is of great importance and needs different approach than that is used in conventional wired networks.

To this end, we intend to find better traffic-carrying capabilities of wireless ad hoc networks, with interdisciplinary study on wireless networking and communication. The theoretical investigation brings forward good design principles, therefore a part of the thesis is devoted to the design of wireless ad hoc networks based on these principles.

For data communications in wireless channels, the dynamic channel fluctuations often cause high frame-error rates. Link layer error control innovations will directly have impact on the network performance, such as throughput and end-to-end delay. In a multi-hop network, the throughput is essentially important for real-time applications due to their high bit rate requirement. Generally protocol design of a practical system introduces a significant gap between the performance and the actual capacity of the network. We then investigate the performance of multimedia applications in wireless networks and the impact of various error control protocols. In particular, we propose a two stage error control scheme that improves the effective throughout of wireless networks. We apply error control to the packet header and packet load separately. The network intermediate nodes either use header FEC or header CRC checksum to successfully transport the packets from the source to the destination. Only at the destination, the error of the load is corrected. We compare the proposed schemes with 802.11 protocol through extensive multi-hop simulations. Specifically, average throughput, end-to-end latency, and video PSNR results are analyzed. The performance comparison between each scheme is discussed in detail. It is shown that header error protection strategy can effectively increase the throughput and the video performance, via both theoretical analysis and simulation results.

We also propose a system design of cooperative networks. This cooperation design principle is intended to give innovations at various layers as to improve overall system throughput and system reliability. Beside proposing new cooperation strategies in physical, link, and network layer, we are currently devoted to detailed design on the link layer cooperation. Our model and analysis propose implications toward the future link layer design. Under a cluster-based network design, code combining is used together with FEC to improve the link layer reliability. This approach is different from how code combining is used in the conventional hybrid ARQ, which is in a sequential way. In this thesis we analyze the decoding performance of the cluster-based cooperative networks with a code combining technique. The analytical results and the simulations show that with the cooperation of nodes in a clustering network, the link reliability will be greatly improved with the same power consumption. Equivalently, this can be viewed as the same link performance with a power saving and lower interference. This result is promising in that the reduced power requirement leads to less interference caused by a transmission, thus can improve the capacity of the wireless networks.

The problem of exploiting the benefits of using directional antennas to improve the channel utilization of ad hoc networks is non-trivial. In this thesis we investigate the capacity of ad hoc wireless networks using directional antennas. In this work, we consider arbitrary networks and random networks where nodes are assumed to be static. In arbitrary networks, due to the reduction of the interference area, the capacity gain is proven to be $\sqrt{\frac{2\pi}{\alpha}}$ when using directional transmission and omni reception. Because of the reduced probability of two neighbors pointing to each other, the capacity gain is $\sqrt{\frac{2\pi}{\beta}}$ when omni transmission and directional reception are used. α and β are the beam width of the directional transmission and receiving pattern, respectively. Although these two expressions look similar, the proof technique is different. By taking advantage of the above two approaches, the capacity gain is $\frac{2\pi}{\sqrt{\alpha\beta}}$ when both transmission and reception are directional. For random networks, interfering neighbors are reduced due to the decrease of interference area when directional antennas are used for transmission and/or reception. The throughput improvement factor is $\frac{2\pi}{\alpha}$, $\frac{2\pi}{\beta}$ and $\frac{4\pi^2}{\alpha\beta}$ for directional transmission/omni reception, omni transmission/directional reception, and directional transmission/directional reception, respectively. We have also analyzed hybrid beamform patterns that are a mix of omnidirectional/directional and a better model of real directional antennas.

Besides simulation results for the network with the deployment of directional antennas, we propose some thoughts on how to make use of the flexible beamforming, multi-beam antenna techniques to help design the MAC protocol, and help the node cooperation as well.

We conclude that by looking at some various novel aspects of the wireless networks and wireless communications, like head error protection, node cooperation, and the deployment of directional antennas, the capacity and performance improvement of the network can be made a reality.

6.2 Future Work

Our preliminary results about cooperation indicate that this approach achieves a quantum leap in the performance/cost trade off. The future focus of our work is on designs which explicitly exploit physical layer, data link layer, and network layer cooperation among nodes. In this thesis, we consider cooperation in each layer separately. Depending on the condition of the channel, which is of a time-varying nature in wireless networks, one of these cooperation methods may prove to be better suited than the other methods. Here there is a trade-off between complexity and higher quality of transmission. Thus, it is desirable to have the system adaptively change from one method to another. Our future research will address the problem of developing criteria and algorithms to implement this adaptability.

Cooperative networks challenge many "proven" approaches to wireless network design (cross layer designs, heterogeneous or homogeneous nodes, cooperation, decentralized organization, energy awareness, etc.). Power saving can be achieved via cooperation in various layers. The consideration of the performance in terms of energy consumption and node mobility management also quantifies the importance of decisions on other layers, notably the clustering, routing and power control layers, and highlights the interplay between protocols spanning different layers, thus motivating a cross-layer design, i.e., a design that spans multiple layers (clustering, routing, power control, link layer error control) of the protocol stack.

Specifically, in the cooperative network paradigm, design and implementation of auto-configuration, cooperative cluster-based routing and forwarding, can be integrated with link layer cooperative FEC to complete the system design. It would be interesting to carry out some real applications over the cooperative network. We have given the theoretical results for the capacity of the network with directional antennas. More experimental works are expected to prove the existing results.

BIBLIOGRAPHY

[1] Theodore S. Rappaport. *Wireless Comunications: Principles and Practice, 2nd Ed.* Prentice Hall PTR, Upper Saddle River, NJ, 2002.

[2] Ajay Chandra V. Gummalla and John O. Limb. Wireless medium access control protocols. *IEEE Communications Surveys*, Second Quarter 2000.

[3] J. Li, C. Blake, D. De Couto, H. Lee, and R. Morris. Capacity of ad hoc wireless networks. In *ACM MobiCom'01*, July 2001.

[4] M. Grossglauser and D. Tse. Mobility increases the capacity of ad hoc wireless networks. In *IEEE INFOCOM'01*, April 2001.

[5] M. Gastpar and M. Vetterli. On the capacity of wireless networks: The relay case. In *IEEE INFOCOM'02*, June 2002.

[6] B. Liu, Z. Liu, and D. Towsley. On the capacity of hybrid wireless networks. In *IEEE INFOCOM'03*, March 2003.

[7] Su Yi, Yong Pei, and Shivkumar Kalyanaraman. On the capacity improvement of ad hoc wireless networks using directional antennas. In *ACM SIGMOBILE International Symposium on Mobile Ad Hoc Networking & Computing (MOBIHOC)*, pages 108–116, Annapolis, MD, 2003.

[8] Christina Peraki and Sergio Servetto. On the maximum stable throughput problem in random networks with directional antennas. In *ACM MobiHoc'03*, June 2003.

[9] U. C. Kozat and L. Tassiulas. Throughput capacity of random ad hoc networks with infrastructure support. In *Proc. ACM MobiCom*, pages 55–65, Sep 2003.

[10] Rohit Negi and Arjunan Rajeswaran. Capacity of power constrained ad-hoc networks. In *IEEE INFOCOM'04*, Mar 2004.

[11] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah. Throughput-delay thade-off in wireless networks. In *IEEE INFOCOM'04*, Mar 2004.

[12] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, IT-46(2):388–404, March 2000.

[13] Asis Nasipuri, Shengchun Ye, and Robert E. Hiromoto. A MAC protocol for mobile ad hoc networks using directional antennas. In *Proceedings of the*

IEEE Wireless Communications and Networking Conference (WCNC 2000), September 2000.

[14] Y. Ko, V. Shankarkumar, and N. H. Vaidya. Medium access control protocols using directional antennas in ad hoc networks. In *IEEE INFOCOM'2000*, March 2000.

[15] R. R. Choudhury, X. Yang, R. Ramanathan, and Nitin Vaidya. Using directional antennas for medium access control in ad hoc networks. In *ACM MobiCom'02*, September 2002.

[16] R. Ramanathan. On the performance of ad hoc networks using beamforming antennas. In *ACM MobiHoc'01*, October 2001.

[17] L. Bao and J.J.Garcia-Luna-Aceves. Transmission scheduling in ad hoc networks with directional antennas. In *ACM MobiCom'02*, September 2002.

[18] S. Lin, D. J. Costello Jr., and M. J. Miller. Automatic-repeat-request error control schemes. *IEEE Communications Magazine*, 22(12):5–16, December 1984.

[19] Shu Lin and Daniel J. Costello. *Error Control Coding, 2nd Ed.* Pearson Education, 2004.

[20] Romano Fantacci and Massimo Scardi. Performance evaluation of preemptive polling schemes and ARQ techniques for indoor wireless networks. *IEEE Transactions on Vehicular Technology*, 45(2):248–257, May 1996.

[21] Stephen B. Wicker. *Error Control Systems for Digital Communication and Storage*. Prentice Hall, Upper Saddle River, NJ, 1995.

[22] Shu Lin and Jr. Daniel J. Costello. *Error control coding: fundamentals and applications*. Prentice-Hall, Englewood Cliffs, N.J., 1983.

[23] Jr. Daniel J. Costello, Joachim Hagenauer, Hideki Imai, and Stephen B. Wicker. Applications of error-control coding. *IEEE Trans on Information Theory*, 44(6):2531–2560, October 1998.

[24] John Byers, Michael Luby, Michael Mitzenmacher, and Ashu Rege. A digital fountain approach to reliable distribution of bulk data. In *proceedings of ACM SIGCOMM '98*. ACM, September 1998.

[25] T. Kostas and S. Jordan. Packet erasure FEC on ARQ protocols. In *SPIE ITCom Internet Performance and Control of Network Systems*, Boston, Massachusetts, July 2002.

[26] Luigi Rizzo. Effective erasure codes for reliable computer communication protocols. *ACM Computer Communication Review*, pages 24–36, Apr 1997.

[27] J.M.Wozencraft and M.Horstein. Coding for two-way channels. In *Proc. 4th London Symp. On Information Theory*. C.Cherry, Ed., 1961.

[28] S. Kallel. Analysis of type II hybrid ARQ schemes with code combining. *IEEE Trans on Communication*, 38(8), August 1990.

[29] S. Kallel. Analysis of memory and incremental redundancy ARQ schemes over a nonstationary channel. *IEEE Trans on Communication*, 40(9):1474–1480, September 1992.

[30] M. B. Pursley and S. D. Sandberg. Incremental-redundancy transmission for meteor-burst communications. *IEEE Trans on Communication*, 39(5):689–702, May 1991.

[31] D. M. Mandelbaum. An adaptive-feedback coding scheme using incremental redundancy. *IEEE Trans on Information Theory*, 20(3):388–389, September 1974.

[32] S. Lin and P. S. Yu. A hybrid ARQ scheme with parity retransmission for error control of satellite channels. *IEEE Trans on Communication*, 30(7):1701–1719, July 1982.

[33] K. R. Narayanan and G. L. Stuber. A novel ARQ technique using the Turbo coding principle. *Communication Letters*, 1(2):49–51, March 1997.

[34] W.-C. Chan, E. Geraniotis, and V. D. Nguyen. An adaptive hybrid FEC/ARQ protocol using Turbo codes. In *Proceedings of the 6th IEEE International Conference on Universal Personal Communications Record*, volume 2, pages 541–545, October 1997.

[35] P.L. Hiew, M. Zukerman, and M. Gitlits. WATM operation optimization based on effect of FEC code rate and ARQ retransmission. In *Proceedings of the IEEE VTC 98*, pages 2542–2546, May 1998.

[36] D. N. Rowitch and L. B. Milstein. On the performance of hybrid FEC/ARQ systems using rate compatible punctured Turbo (RCPT) codes. *IEEE Trans on Communication*, 48(6):948–959, June 2000.

[37] C. Schuler and M. Mateescu. Performance evaluation of ARQ protocols for realtime services in IEEE 802.11 and wireless ATM. In *4th ACTS Mobile Communications Summit*, Sorrento, Italy, June 1999.

[38] D. Chase. Code combining–a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets. *IEEE Trans on Communications*, 33(5):385–393, May 1985.

[39] J. Hagenauer. Rate-compatible punctured convolutional codes (RCPC codes) and their applications. *IEEE Trans on Communication*, 36(4):389–400, April 1988.

[40] S. B. Wicker and M. D. Bartz. Type-II hybrid-ARQ protocols using punctured MDS codes. *IEEE Trans on Communication*, 42(2/3/4):1431–1440, February/March/April 1994.

[41] X. Wang and M. T. Orchard. On reducing the rate of retransmission in time-varying channels. *IEEE Trans on Communication*, 51(6):900–910, June 2003.

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

[43] U.Black. *ATM Foundation for Broadband Networks, vol. 1, 2nd Ed.* Prentice Hall PTR, Upper Saddle River, NJ, 1999.

[44] David A. Eckhardt and Peter Steenkiste. Improving wireless LAN performance via adaptive local error control. In *Proc. ICNP'98*, 1998.

[45] I. Chakeres, H. Dong, E. M. Belding-Royer, A. Gersho, and J. D. Gibson. Allowing errors in speech over wireless LANs. In *Proceedings of the 4th Workshop on Applications and Services in Wireless Networks (ASWN)*, Boston, MA, August 2004.

[46] A. Servetti and J. C. De Martin. Link-level unequal error detection for speech transmission over 802.11 wireless networks. In *Special Workshop in Maui (SWIM), Lectures by Masters in Speech Processing*, Maui, HI, January 2004.

[47] S. Krishnamachari, M. Van Der Schaar, S. Choi, and X. Xu. Video streaming over wireless LANs: A cross-layer approach. In *Proc. Packet Video Workshop*, Nantes, France, April 2003.

[48] Yufeng Shan. Cross-layer techniques for adaptive video streaming over wireless networks. *EURASIP Journal on Applied Signal Processing*, 2005(2):220–228, February 2005.

[49] Q. Li and M. Van Der Schaar. Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation. *IEEE Transactions on Multimedia*, 6(2):278–290, April 2004.

[50] H. Zheng and J. Boyce. An improved UDP protocol for video transmission over internet-to-wireless networks. *IEEE Trans. on Multimedia*, 3(3):356–365, September 2001.

[51] L.-A. Larzon, M. Degermark, S. Pink, L.-E. Jonsson, and G. Fairhurst. The lightweight user datagram protocol (UDP-Lite). RFC 3828, 2004.

[52] G. Ding, H. Ghafoor, and B. Bhargava. Error resilient video transmission over wireless networks. In *Proc. IEEE Workshop on Software Technologies for Future Embedded Systems*, pages 31–34, Hakodate, Hokkaido, Japan, May 2003.

[53] D. Eckhardt and P. Steenkiste. Measurement and analysis of the error characteristics of an in building wireless network. In *Proceedings of the ACM SIGCOMM '96 Symposium on Communications Architectures and Protocols*, pages 243–254, August 1996.

[54] E.N.Gilbert. Capacity of a burst-noise channel. *Bell Syst. Tech. J.*, 39:1253–1265, September 1960.

[55] E.O.Elliot. Estimates of error rates for codes on burst-noise channels. *Bell Syst. Tech. J.*, 42:1977–1997, September 1963.

[56] M. Zorzi, R. R. Rao, and L. B. Milstein. On the accuracy of a first-order Markov model for data transmission on fading channels. In *Proceedings of ICUPC*, pages 211–215, Tokyo, Japan, November 1995.

[57] M. Zorzi, R. R. Rao, and L. B. Milstein. ARQ error control for fading mobile radio channels. *IEEE Transactions on Vehicular Technology*, 46(2):445–455, May 1997.

[58] The network simulator - ns-2. http://www.isi.edu/nsnam/ns/.

[59] Piyush Gupta, Robert Gray, and P. R. Kumar. An experimental scaling law for ad hoc networks. http://decision.csl.uiuc.edu/ prkumar/html-files/ps-files/exp.pdf, 2001.

[60] A.N. Netravali and B.G. Haskell. *Digital Pictures: Representation, Compression, and Standards (2nd Ed).* Plenum Press, New York, NY, 1995.

[61] Yufeng Shan, Su Yi, Shivkumar Kalyanaraman, and John.W. Woods. Two-stage FEC scheme for scalable video transmission over wireless networks. In *SPIE Communications/ITCom, Multimedia Systems and Applications*, Boston, MA, October 2005.

[62] B. Azimi-Sadjadi and A. Mercado. Diversity gain for cooperating nodes in multi-hop wireless networks. In *Proceedings of IEEE Vehicular Technology Conference 2004-Fall*, Los Angeles, CA, Sep 2004.

[63] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity–part I: System description. *IEEE Transactions on Communications*, 51(11), November 2003.

[64] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity–part II: Implementation aspects and performance analysis. *IEEE Transactions on Communications*, 51(11), November 2003.

[65] P. Anghel and G. Leus and M. Kaveh. Distributed Space-Time Coding in Cooperative Networks. *5th Nordic Signal Processing Symposium*, October 2002.

[66] J. N. Laneman and G. W. Wornell. Exploiting Distributed Spatial Diversity in Wireless Networks. *in Proc. Allerton Conf. Commun., Contr., and Computing*, October 2000.

[67] J. N. Laneman and G. W. Wornell. Distributed Space-Time Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks. *IEEE Trans. Inform. Theory*, 49(10), October 2003.

[68] J. N. Laneman and D. N. C. Tse and G. W. Wornell. Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior. *IEEE Trans. Inform. Theory*, 50(12), Dec 2003.

[69] T. Hunter and A. Nosratinia. Cooperation Diversity Through Coding. *Proc. International Symposium on Information Theory*, June 2002.

[70] A. Stefanov and E. Erkip. Cooperative space-time coding for wireless networks. *In Proceedings of IEEE Information Theory Workshop*, April 2003.

[71] A Scaglione and Y. Hong. Opportunistic large arrays: Cooperative transmission in wireless multihop ad hoc networks to reach far distances. *IEEE Transactions on Signal Processing*, 51(8):2082–2092, August 2003.

[72] A. Mercado and B. Azimi-Sadjadi. Power efficient link for multi-hop wireless networks. In *41st Annual Allerton Conference on Communication, Control, and Computing*, Oct 2003.

[73] Su Yi, Babak Azimi-Sadjadi, Shivkumar Kalyanaraman, and Vijaynarayanan Subramanian. Error control code combining techniques in cluster-based cooperative wireless networks. In *IEEE International Conference on Communications (ICC)*, Seoul, Korea, 2005.

[74] E.M. Royer and C.-K. Toh. A review of current routing protocols for ad hoc mobile wireless networks. *IEEE Personal Communications*, pages 46–55, Apr 1999.

[75] S.R. Das, C.E. Perkins, and E.M. Royer. Performance comparison of two on-demand routing protocols for ad hoc networks. In *IEEE INFOCOM 2000*, pages 3–12, Tel Aviv, Israel, March 2000.

[76] T. Camp, J. Boleng, B. Williams, L. Wilcox, and W. Navidi. Performance comparison of two location based routing protocols for ad hoc networks. In *IEEE INFOCOM 2002*, pages 1678–1687, New York, NY, June 2002.

[77] C.A. Santivez, B. McDonald, I. Stavrakakis, and R. Ramanathan. On the scalability of ad hoc routing protocols. In *IEEE INFOCOM 2002*, pages 1688–1697, New York, NY, June 2002.

[78] G. Finn. Routing and addressing problems in large metropolitan-scale internetworks. Technical Report ISI/RR-87-180, USC/ISI, March 1987.

[79] B. Chen and R. Morris. L+: Scalable landmark routing and address lookup for multi-hop wireless networks. Technical Report MIT-LCS-TR-837, MIT, MA, March 2002.

[80] B. Karp and H.T. Kung. GPSR: Greedy perimeter stateless routing for wireless networks. In *Proceedings of the Sixth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 243–254, Boston, MA, August 2000.

[81] J. Li, J. Jannotti, D.S.J. De Couto, D. Karger, and R. Morris. A scalable location service for geographic ad-hoc routing. In *Proceedings of the Sixth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 120–130, New York, NY, August 2000.

[82] L. Blazevic, S. Giordano, and J.-Y. Le Boudec. Anchored path discovery in terminode routing. In *Proceedings of the Second IFIP-TC6 Networking Conference (Networking 2002)*, pages 141–153, Pisa, May 2002.

[83] L. Blazevic, J-Y. Le Boudec, and S. Giordano. A scalable routing scheme for self-organized terminode network. In *Communication Networks and Distributed systems modelling and Simulation conference (CNDS)*, San Antonio, Texas, January 2002.

[84] D. Niculescu and B. Nath. Routing on a curve. In *Proceedings of Workshop on Hot Topics in Networks (HOTNETS-I)*, Princeton, NJ, October 2002.

[85] M. Yuksel, R. Pradhan, and S. Kalyanaraman. An implementation framework for trajectory-based routing in ad-hoc networks.

[86] D. B. Johnson and D. A. Maltz. Dynamic source routing in ad-hoc wireless networks. In T. Imielinski and H. Korth, editors, *Mobile Computing*, chapter 5. Kluwer Academic Publishers, 1996.

[87] M. Jiang, J. Li, and Y. C. Tay. Cluster based routing protocol (CBRP). Internet draft, August 1999.

[88] A. Iwata, C. Chiang, G. Pei, M. Gerla, , and T. Chen. Scalable routing strategies for ad hoc wireless networks. *IEEE J. Select. Areas Commun.*, 17(8):1369–1379, Aug 1999.

[89] J. Moy. OSPF version 2. *IETF Request for Comments (RFC)*, 1247, July 1991.

[90] D. Niculescu and B. Nath. Trajectory based forwarding and its applications. In *Proceedings of ACM MOBICOM*, San Diego, CA, September 2003.

[91] D. Niculescu and B. Nath. Ad hoc positioning system (APS). In *Proceedings of IEEE GLOBECOM*, pages 2926–2931, San Antonio, TX, November 2001.

[92] P. Bahl and V. N. Padmanabhan. RADAR: An in-building RF based user location and tracking system. In *Proceedings of IEEE INFOCOM*, pages 775–784, Tel Aviv, Israel, March 2000.

[93] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The cricket location-support system. In *Proceedings of ACM MOBICOM*, pages 32–43, Boston, MA, August 2000.

[94] N. Bulusu, J. Heidemann, and D. Estrin. GPS-less low cost outdoor localization for very small devices. *IEEE Personal Communications*, 7(5):28–34, October 2000.

[95] Convergence S. Capkun, M. Hamdi, and J.-P. Hubaux. GPS-free positioning in mobile ad-hoc networks. In *Proceedings of Hawaii International Conference on System Sciences*, pages 3481–3490, Maui, HW, January 2001.

[96] R. Iyengar and B. Sikdar. Scalable and distributed GPS free positioning for sensor networks. In *Proceedings of IEEE ICC*, pages 338–342, Anchorage, AK, May 2003.

[97] Wendi B. Heinzelman, Anantha P. Chandrakasan, and Hari Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans on Wireless Communications*, 1(4):660–670, Oct 2002.

[98] B. A. Harvey and S. B. Wicker. Packet combining systems based on the Viterbi decoder. *IEEE Trans on Communication*, 42(2/3/4):1544–1557, February/March/April 1994.

[99] P. Sindhu. Retransmission error control with memory. *IEEE Trans on Communication*, 25(5):423–429, May 1977.

[100] G. Benelli. An ARQ scheme with memory and soft error detectors. *IEEE Trans on Communication*, 33(3):285–288, March 1985.

[101] John G. Proakis. *Digital Comunications 4th Ed.* McGraw-Hill, New York, NY, 2001.

[102] Eugene Shih, Seong-Hwan Cho, Nathan Ickes, Rex Min, Amit Sinha, Alice Wang, and Anantha Chandrakasan. Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pages 272–287, Rome, Italy, Jul 2001.

[103] Sanjit Biswas and Robert Morris. ExOR: Opportunistic multi-hop routing for wireless networks. In *Proceedings of ACM SIGCOMM'05*, pages 133–143. ACM, August 2005.

[104] Romit Roy Choudhury, Xue Yang, Ram Ramanathan, and Nitin H. Vaidya. On designing MAC protocols for wireless networks using directional antennas. *IEEE Transactions of Mobile Computing (TMC)*, 2005.

[105] I. Koutsopoulos, T. Ren, and L. Tassiulas. Efficient media access protocols for wireless LANs with smart antennas. In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, 2003.

[106] A.Spyropoulos and C. S. Raghavendra. Capacity bounds for ad-hoc networks using directional antennas. In *Proc. of IEEE 2003 International Conference on Communications (ICC)*, May 2003.

[107] John D. Kraus and Ronald J. Marhefka. *Antennas: for All Applications, 3rd Ed.* McGraw-Hill, New York, 2002.

[108] J. J. Carr. Directional or omnidirectional antenna? Http://www.dxing.com/tnotes/tnote01.pdf.

[109] Piyush Gupta and P. R. Kumar. Critical power for asymptotic connectivity in wireless networks. In W.M. McEneany, G. Yin, and Q. Zhang, editors, *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*, pages 547–566. Birkhauser, Boston, MA, 1998.

[110] Karthikeyan Sundaresan, Raghupathy Sivakumar, Mary Ann Ingram, and Tae-Young Chang. Medium access control in ad-hoc networks with MIMO links: Optimization considerations and algorithms. *IEEE Transactions on Mobile Computing*, 3(4):350–365, October 2004.

[111] http://www.haleakala-research.com/.

[112] IntelliCell smart antenna solutions. http://www.arraycomm.com/serve.php?page=IntelliCell.

[113] Thanasis Korakis, Gentian Jakllari, and Leandros Tassiulas. A MAC protocol for full exploitation of directional antennas in ad-hoc wireless networks. In *ACM SIGMOBILE International Symposium on Mobile Ad Hoc Networking & Computing (MOBIHOC)*, pages 98–107, Annapolis, MD, 2003.

[114] Romit Roy Choudhury and Nitin H. Vaidya. Deafness: A MAC problem in ad hoc networks when using directional antennas. In *12th IEEE International Conference on Network Protocols (ICNP)*, pages 283–292, 2004.