LABORATORY FOR
COMPUTER SCIENCE

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

MIT/LCS/TR-670

Decentralized Channel Management in Scalable Multihop Spread-Spectrum Packet Radio Networks

Timothy Jason Shepard

July, 1995

545 TECHNOLOGY SQUARE; CAMBRIDGE, MASSACHUSETTS 02139 (617) 253-5851

ERRATA

Since the thesis was submitted, the following minor errors have been discovered in the document. All of these have been corrected in the version you are reading now. Other than corrections for these errors, this version is identical to the submitted thesis.

Page 13, in fourth paragraph, "This work will assume that the all stations"
should be "This work will assume that all the stations"

Page 23, in second line of caption of Figure 2.1:
reference should be to equation 2.12 instead of 2.11.
(fixed in paper TR)

Page 24, 3 lines up from bottom of page, "of an interfering transmissions."
should be "of an interfering transmission."

Page 26, 13 lines up from bottom of page,
"system wide" should be "system-wide"

Page 27, in footnote number 9,
"the recent past might be good-enough predictor" should be
"the recent past might be a good-enough predictor".

Page 33, 8 lines up from bottom of page,
"schedule will be a divided into equal-length slots ..." should be
"schedule will be divided into equal-length slots ...".

Page 36, wrong type of dash before "see Section 3.1.6"

Page 47, 5 lines up from bottom of page,
In the text "Where x is the top 38 bits ..." the letter x
should be set in italic font. (fixed in paper TR)

page 68, wrong type of dash before "see Chapter 3"

In places where a sentence ends with a capital letter (e.g. from an abbreviation like "dB" or "dBW") immediately before the period, LaTeX was misled into believing that the period did not indicate the end of a sentence. This resulted in too little space set after the period. This occurs twice on page 23, once each on pages 24 and 28, twice on page 29, once on page 40, twice on page 46, thrice on page 47, once on page 48, thrice on page 49, twice on page 50, and once each on pages 52, 56 and 75.

Decentralized Channel Management in Scalable Multihop Spread-Spectrum Packet Radio Networks

by

Timothy Jason Shepard

S.B., Massachusetts Institute of Technology
(1986)

S.M., Massachusetts Institute of Technology
(1990)

E.E., Massachusetts Institute of Technology
(1991)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

July, 1995

Signature of Author ______________________________

Department of Electrical Engineering and Computer Science
July 11, 1995

Certified by ______________________________

Senior Research Scientist David D. Clark
Thesis Supervisor

Accepted by ______________________________

Professor Frederic R. Morgenthaler
Chairman, Departmental Committee on Graduate Students

Decentralized Channel Management
in
Scalable Multihop Spread-Spectrum Packet Radio Networks

by

Timothy Jason Shepard

Submitted to the Department of Electrical Engineering and Computer Science on July 11, 1995 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering and Computer Science

ABSTRACT

This thesis addresses the problems of managing the transmissions of stations in a spread-spectrum packet radio network so that the system can remain effective when scaled to millions of nodes concentrated in a metropolitan area. The principal difficulty in scaling a system of packet radio stations is interference from other stations in the system. Interference comes both from nearby stations and from distant stations. Each nearby interfering station is a particular problem, because a signal received from it may be as strong as or stronger than the desired signal from some other station. Far-off interfering stations are not individually a problem, since each of their signals will be weaker, but the combined effect may be the dominant source of interference.

The thesis begins with an analysis of propagation and interference models. The overall noise level in the system (mainly caused by the many distant stations) is then analyzed, and found to remain manageable even as the system scales to billions of nodes. A scheme for designing a scalable packet radio network is then presented. Included is a method of scheduling packet transmissions to avoid collisions (caused by interference from nearby stations) without the need for global coordination or synchronization. Simulations of a system of one thousand stations are used to verify and illustrate the methods used.

A method of choosing routes (minimum-energy routes) is demonstrated in simulation to produce a fully connected and functional network for one hundred and one thousand randomly placed stations. Unfortunately, congestion as the system scales is unavoidable if the traffic is not limited to some degree of locality. If traffic is limited to a few hops, then for a large system the techniques presented in this thesis are superior to ideal time division multiplexing of a clear channel.

Thesis Supervisor:	David D. Clark
Title:	Senior Research Scientist

Acknowledgments

Many people helped make this possible in many ways. I hesitate to start naming names for fear of forgetting someone.

Most of all, my parents made this possible. Whether it is nature or nurture that more makes the person is an old question, but I have no doubt that the nurture, care, and love given by my parents is more responsible than anything else for making me who I am today. My mother *Carol Shepard* is, in stereotypical Missourian style, always the skeptic when first told something new, and is blessed with a bright and particularly rational mind. (She probably got much of this from her mother, *Lydia Utlaut Meise*, whom I remember giving a skeptical look when told new things.) My father *Ralph Shepard*, a mining engineer by training, gave me a very hands-on and practical view of all things around me, particularly of technological things. When something around the house stopped working, we quickly had the cover off and were poking around inside, even if neither of us had any idea what we would see on the other side of the cover. Often, with the cover removed, the inner-workings were completely revealed and the problem made evident. At other times, when we were otherwise baffled as to the source of the problem, we'd put the cover back on and discover that the problem had vanished. He was always very patient and gave the best answer he could when I asked "Why?" and "How?" questions. (I distinctly remember my first epiphany, sitting at the kitchen table when I was just starting first grade, when my father explained to me how we could count forever by adding a new digit at the left whenever we exhausted all the combinations of the digits in the places we had so far. I think I realized then how slowly a logarithm grows, though I had no idea that it was called "logarithm".) A combination of my mother's healthy skepticism and clear thinking, with my father's faith that we would be able to understand and fix the problem (if we could only figure out how to get the cover off) formed the roots of my desire and ability to think things through and figure them out for myself.

Often I find I do my best thinking when I'm trying to explain to someone what I had thought I had already figured out. Many people listened while I muddled my way through explanations of half-baked ideas relevant to this thesis. They include: *Cora Dancy, Mike Biafore, Mike Lauer, Joe Rushanan, Jeff Schiller, Ed Ajhar, Win Treese, Alan Bawden, Carl Sundberg, Matt Reynolds, Mitchell Charity, Nick Papadakis, Gerry Sussman, Ye Gu, Jeff Van Dyke, Anna Charny, Chris Lefelhocz, Peter Daly, Andreas Gast*, and I'm sure many others whom I've unfortunately failed to remember.

Rainer Gawlick met with me about once a week for much of the past year to discuss the problem of routing in packet radio networks. We made little progress on the problem, or even towards figuring out what problem we were trying to solve, but the discussions helped much to refine my thinking about packet radio networks. We were also joined by *David Karger* for some of these discussions.

Greg Troxel taught me much about how radios work and first sparked my interest in the sport of hidden-transmitter hunting where I learned much practical intuition about radio propagation. He also, by example, encouraged me to believe that it is possible to finish. He read an early draft of the thesis and provided much useful

feedback.

Jim Bales taught me much about physics and math over many "nerd feeds" and also helped me think through some of the noise growth calculations which appear in Chapter 2. He also proofread a late draft of the entire thesis in a 24 hour period and provided me with many marks per page. The thesis is much better shape than it would have been without his heroic effort.

Janey Hoe, *Magdalena Leuca*, and *Garrett Wollman* also helped proofread portions of the thesis or early drafts of various chapters at varying stages of development.

My two committee members *Tom Knight* and *Jerry Saltzer* both provided good suggestions for improving the thesis and provided prompt attention to the thesis when I needed it.

My supervisor *David Clark* provided guidance, support, and the freedom to pursue my own ideas. I'm sure I should have followed his advice more often than I did.

This research was supported in part by the Advanced Research Projects Agency under contract number DABT63–92–C–0002. The author was supported in part by an AT&T-Bell Laboratories Ph.D. Scholarship sponsored by the AT&T Foundation.

Contents

Chapter 1

Introduction

This work addresses issues relevant to realizing a large-scale packet radio network in which the stations are cooperative but at the same time autonomous and not dependent upon any installed infrastructure. The issues addressed include the growth of the noise level as the system scales, the technical challenge of controlling access to the channel without any centralization of control, and strategies for routing. Scalability and decentralization of control are the primary concerns throughout this work.

1.1 Perspective

Telecommunication is achieved by either sending signals through cables or by letting generated signals propagate naturally through space as electromagnetic radiation (e.g. radio waves). Cables can provide seemingly unlimited bandwidth, but require capital investment to obtain right of way and install. Radio waves require no capital investment nor right of way for propagation, but the radio spectrum is heavily regulated by governments and is usually thought to be scarce.

The technologies of both media are advancing. In the case of cables, the amount of bandwidth available through a given cross section of cable increases as the technology improves for modulating and detecting light at high bandwidths (in fiber-optic cables). However, the fixed costs of labor-intensive installation can continue to dominate the costs of communication by cables and keep them expensive despite improvements in technology.

In the case of radios, improvements in technology both improve the efficiency of spectrum use (from better coding and signal processing) and increase the amount of spectrum available for communication (as the construction of electronics to operate at higher radio frequencies becomes practical). Though the radio spectrum is often believed to be scarce, the increasing practical accessibility of higher frequency bands should, for a while, more than offset increasing demand for spectrum. The maximum operating frequency of electronics for computers is currently increasing exponentially as a function of time, with a doubling time of two or three years.[1] The fundamental limits of electronic circuit performance that determine maximum operating frequency

[1] For how many more decades this trend will continue is unclear.

are no different for radio electronics. These high speed electronic circuits are necessary for the practical accessibility of higher frequency radio bands. Understanding this, the believed scarcity of the radio spectrum is a myth.[2]

This work proceeds from the assumption that a significant amount of spectrum will be available for unlicensed operation directly between consumer-owned devices and that it will be possible to build the electronics to operate at the available frequencies. Indeed, in the U.S. there are already frequency bands available for unlicensed use that can support reasonably high data rates.[3] Wireless LANs and point-to-point radio systems (for use between buildings) with maximum ranges of a few kilometers are commercially available today. Users are already buying their own unlicensed radio equipment and installing and using it to connect computer networks across town without having to pay any ongoing fee to a service provider. The customers own the equipment and take responsibility for maintenance. No one has responsibility for coordination between these systems. Observing this developing situation triggered the motivating questions for this work: Can this situation last? What if this growing anarchy of wireless communication really catches on? Will these point-to-point links continue to work if there are millions of them in a metropolitan area? How far can this scale?

As will be seen, if the devices are to communicate over any distance that is large with respect to the average spacing between devices, then the devices will need to cooperate and relay messages for other devices. It will be most useful if these devices can organize themselves into a larger relaying system without any engineering of the deployment of devices. Otherwise, whomever is responsible for engineering and deploying any "backbone" devices will need to be ensured compensation for their efforts. The self-organizing ability does not rule out someone deploying infrastructure capable of communicating with these devices and then only allowing paying customers to benefit from this deployed infrastructure. But, from a technical standpoint, users should be able to make use of the device to communicate with others in the area who have these devices without necessarily needing to contract with some communication provider. By designing the devices to cooperate in forwarding of traffic, the usefulness and value of the devices are enhanced.

1.2 Multihop packet radio networks

A *packet radio network* consists of a number of packet radio stations that communicate with each other. A packet radio network carries messages in packets like a wired packet network (with nodes forwarding each packet separately), but uses radio signals instead of wires to carry the packets between stations. A packet radio network could

[2]The spectrum that will eventually be available for radio communication is not unlimited. The atmosphere is effectively transparent only at frequencies below 10 GHz. Between 10 GHz and 500 GHz the atmospheric attenuation as a function of frequency varies between 0.01 dB/km and 50 dB/km, and below 100 GHz is always less than 1 dB/km except in the O_2 absorption band from 53 GHz to 67 GHz where the attenuation is as high as 16 dB/km. For further information see [Fre87] and [CCI86].

[3]See 47 CFR 15.247 in the U.S. Code of Federal Regulations.

be an extension of a larger packet network in which packets might travel over both wires and radio signals. Each station in a packet radio network includes an embedded computer to perform packet routing and forwarding functions, radio transmitting and receiving equipment, and a port for connecting to a local terminal, computer, or network. In a *multihop* packet radio network each station participates cooperatively in forwarding traffic between other stations, thereby extending the communication range of each station to include all transitively reachable stations.

A fairly general model of packet radio networks will be presented, and then some simplifying assumptions will be made for the remainder of this work. After this model has been presented, the related work will be discussed in terms of this model.

The framework consists of three parts: a model of a signal (used to model the received and transmitted signals), a model of propagation, and a model for which transmissions are successfully received by the intended recipient. The design space within this framework includes controlling what signals are transmitted (including the amount of power), when the transmissions occur, and to some extent the behavior of the receiver.

A signal (transmitted or received) is most completely modeled as a real-valued function of time. The signal transmitted by station i is denoted by $s_i(t)$. The received signal at station i is denoted by $y_i(t)$. At times when a station is not transmitting a message, its output signal is zero. When one station wishes to convey a message to a another, it transmits a signal that influences the signal received by the receiver. The receiver attempts to recover the message from the received signal.[4] The details of designing signals to carry messages and designing receivers to detect messages from received signals are beyond the scope of this work, but we do need to understand what performance can be realistically achieved, and what parameters determine performance.

The two parameters of a transmitted signal that are important for understanding system performance are the transmitted signal's power level and its bandwidth. Both of these parameters are limited by government regulation and by the limitations of the particular transmitter hardware. Transmitter power will be assumed to be controllable. (Bandwidth is most likely fixed at time of manufacture.) Successful reception of the message at the receiver will depend upon receiver performance, the total power level of interfering signals at the receiver, and the received power level of the signal containing the message. The influence transmitted signals have on received signals is determined by *propagation*.

Propagation and noise determine the received signals as a function of the transmitted signals. Assuming linearity and time-invariance, a general model is

$$y_i(t) = n_i(t) + \sum_{j=1}^{M} h_{ij}(t) * s_j(t) \tag{1.1}$$

where M is the number of stations, $h_{ij}(t)$ is the response at station i to an impulse

[4]The word *signal* here refers to the entire received signal. Later this same word is used to refer to just the component of the whole received signal that is due to the transmitted signal from which a message is to be extracted.

in time transmitted by station j (the $h_{ij}(t)$ will be collectively referred to as the propagation matrix H), $n_i(t)$ is the signal due to thermal noise at station i, and the symbol $*$ represents convolution. This model for propagation is not much of a simplification of the real world (the world between antennas is mostly linear and time-invariant). Some other traditional areas of research (described in the related work section) use simpler models for signals and propagation that are incapable of capturing two important aspects of propagation, interference, and detection. These aspects cannot be ignored if large-scale packet radio networks are to be accurately modeled.

Interference may come from a variety of sources, including thermal noise, atmospheric effects, other man-made devices operating at radio frequencies, and even extraterrestrial sources. The only external source of interference included in Eq. 1.1 is the thermal noise. Note that interference due to transmissions from other stations in the packet radio network is included explicitly in the summation. Thermal noise will be shown to be insignificant compared to the effect of interference from other stations in the system designed here (and we will soon ignore it). However, in a system where only one station transmitted at a time (into an otherwise clear channel), the performance would be limited by the thermal noise.

The impulse response $h_{ij}(t)$ is a general model for propagation in that it can represent the strength of the propagation, the propagation delay, and any multi-path propagation. Both propagation delay and multi-path effects will be ignored for the remainder of this work, so $h_{ij}(t)$ will be assumed to be a just a scalar multiple of the unit impulse, $h_{ij} \cdot \delta(t)$, and Eq. 1.1 can be simplified to

$$y_i(t) = n_i(t) + \sum_{j=1}^{M} h_{ij} s_j(t) \tag{1.2}$$

where the h_{ij} are now scalars instead of functions.[5] The propagation model is not complete until the h_{ij} are specified. In the real world, stations will observe the actual propagation between stations that are capable of direct communication. The exact value of propagation between each pair of stations not capable of direct communication will not be important. In this work, propagation will be modeled by setting each h_{ij} proportional to $1/r_{ij}$ where r_{ij} is the distance from station i to station j. This model corresponds to the familiar $1/r^2$ free space loss (in power) for electro-magnetic radiation.[6]

[5] Propagation delay and multi-path propagation are important effects to be considered when actually designing the system, but the effects are not important to this work. If necessary, actual delays could be observed and easily compensated for in the scheduling technique presented in Chapter 3. The successful detection of wide-band spread-spectrum signals (which this work will use) is particularly robust to interference from multi-path propagation. If necessary, a *rake receiver* can be employed to detect and combine the separately arriving copies of a transmitted signal. Rake receivers are described in [SOSL94].

[6] The *power* of a signal captured by an antenna is proportional to the power per unit area of the incident electro-magnetic radiation. The power per unit area falls off as the inverse square of the distance from the source. The voltage measured on the feedline is proportional to the square root of the power on the feedline. See [RR64] Chapter 4.

Actual propagation in most cases will either be nearly equal to the free space propagation (when the antennas are within radio line of sight) or will be attenuated (when there are obstructions). Hence, assuming free space propagation, we will tend to accurately model the strength (at a receiver) of the stronger signals from nearby sources while overestimating the strength of the many weaker signals from more distant transmitters.

Whether or not a given packet transmission will be successfully received in a real network will depend upon many technical details. However, a bound on the performance of the receiver can be derived from Shannon's capacity theorem[7] if we assume that the receiver makes no attempt to model and subtract the interfering signals. This assumption is reasonable given the number of interfering signals expected in a network of many stations. Techniques for multiuser detection that do estimate the interfering signals can surpass the bounds derived from Shannon's capacity theorem, but are only practical when the number of interfering signals is few.[8]

Shannon's theorem bounds the capacity C of a communication channel by a function of the average signal power S, the average interfering noise power N, and the bandwidth W:

$$C \leq W \log_2 \left(1 + \frac{S}{N}\right). \tag{1.3}$$

This bound can be used to provide a model for successful reception of a transmission if we assume that the transmitter and receiver are attempting to optimize the probability of successful reception. Achieving the Shannon bound is not practically possible, but for a reasonable effort, a capacity can be achieved that corresponds to the Shannon bound for a situation a few dB worse in signal-to-noise ratio.

In general, stations might vary the rate at which they communicate depending on the observed interference. This work will assume that all the stations will communicate at some rate that is fixed by the design, and will address what this rate should be.

A packet will be successfully received at a station i from station k if, while it is being received, the received signal-to-noise ratio is at least some small factor, $\alpha > 1$ (and probably around 3), more than the minimum required signal-to-noise ratio, i.e.

$$\frac{S}{N} \geq \alpha \left(2^{\frac{C}{W}} - 1\right). \tag{1.4}$$

C is now not exactly the capacity, but the data rate at which the stations are attempting to communicate. The signal strength S is the power of the signal received at station i from the sending station k (i.e. the power in the signal $h_{ik}s_k(t)$) and N

---

[7] [Sha48]

[8] Verdú in [Ver89] suggests that multiuser detection might be possible when the number of interfering signals does not exceed 10 to 15, and states that the complexity of multiuser detection is exponential in the number of interfering signals. The packet radio networks considered here might nevertheless benefit from receivers that model and subtract only a few of the strongest interfering signals, but consideration of this potential improvement of receiver performance will remain beyond the scope of this work.

is the power contained in the sum of the interfering signals,

$$N = n_i(t) + \sum_{j=1, j\neq k}^{M} h_{ij}(t) s_j(t). \tag{1.5}$$

The power in this signal is the same as the sum of the powers of each of the interfering signals, as we have assumed that the signals are uncorrelated and of zero mean. As will be shown in Chapter 2, in a large system the interference from other stations will dominate any thermal noise, so the thermal noise will now be ignored. Hence the signal-to-noise ratio at a receiver i for the transmission from station k can be computed (for purposes of simulation) from just the powers of the transmitted signals and the h_{ij}'s squared,

$$\left(\frac{S}{N}\right)_{ik} = \frac{h_{ik}^2 P_k}{\left[\sum_{j=1}^{M} h_{ij}^2 P_j\right] - h_{ik}^2 P_k}. \tag{1.6}$$

The P_i's, and hence S and N, are actually all functions of time and vary as stations begin and end transmissions and vary transmitter power levels. The criterion for successful reception of a packet is then that the signal-to-noise ratio be greater than the required minimum for the duration of its reception, which can be determined from the power levels alone.

The signal-to-noise ratios in the system designed here will in general be significantly less than one. Therefore we can approximate $\log_2\left(1 + \frac{S}{N}\right)$ in Eq. 1.3 as $\frac{x}{\ln 2}$ (approximately $1.44x$) so raw throughput is roughly proportional to signal-to-noise ratio.

1.3 Spread spectrum

Spread spectrum[9] will play a key role in this work. This term is used to describe techniques for practically achieving communication by radio when the signal-to-noise ratio is less than one (within the used bandwidth). Techniques of spread-spectrum communication first came of interest as a means of communication with resistance to hostile jamming. More recently, interest in spread-spectrum systems has developed in other application areas, such as cellular telephone systems and wireless local area networks, where spread spectrum is used in place of more traditional multiplexing techniques to achieve multiple access. Spread spectrum, though in some sense theoretically not as efficient as time or frequency division multiplexing, can greatly simplify the problem of managing the channel, especially when the synchronization required to efficiently schedule (in time) or assign channels (in frequency) cannot be achieved in a decentralized system.[10]

The most straightforward spread-spectrum technique for digital communication is *direct-sequence* spread spectrum (or DSSS). DSSS radio systems operate by using a

[9] See any of [Dix84], [SOSL94], [ZP85], [Nic88], [CEMS83] for a more thorough introduction to spread-spectrum radio techniques.

[10] This advantage is mentioned by Gallager in the introduction of [Gal85].

more traditional narrow band modulator and demodulator, but before the modulated signal is transmitted, a fast-running pseudo-random bit sequence is combined with the signal. This pseudo-random bit sequence is known as the *spreading code* and the rate at which it runs is known as the *chip rate*. The chip rate is usually an exact multiple of the data bit rate, yielding an integer chips to bit ratio. The receiver generates the same spreading code synchronously and combines it with the received signal, undoing the spreading that was done in the transmitter. The receiver can then use traditional narrow-band filters and detector to isolate the signal and demodulate it to determine the data bits. The bandwidth occupied by the DSSS signal is roughly equal to the chip rate.

In spread spectrum systems, the *processing gain* is the apparent gain in the desired signal achieved over interfering signals that are uncorrelated with the spreading code. Interfering signals that are uncorrelated with the spreading code will tend to be spread evenly over the band after being combined with the spreading code in the receiver. The narrow-band filter, whose width will be roughly equal to the data bit rate, will filter out all but a fraction of the interfering signal while retaining most of the desired signal (which has been despread to fit within the filter's pass band). The desired signal has apparently been amplified over the interfering signal by a factor equal to the original bandwidth divided by the bandwidth of the narrow-band filter. The processing gain is this factor, and is also equal to the ratio of the chip rate to the data bit rate.

Spread-spectrum radio techniques can be used to build systems that are capable of communication (at rates within the Shannon bound) in channels where the interference has lowered the signal-to-noise ratio to well below one, but there are some practical limits. Some additional signal level, or *headroom* (the constant factor α in the formula above), will be needed over the minimum implied by the Shannon bound. In [SOSL94][11] it is determined that around 5 dB of headroom (after processing gain) is needed to achieve a 10^{-6} bit error probability using a DS/BPSK (binary phase shift keying) radio link with Viterbi decoding in a spread-spectrum multiple access application. (The example system analyzed in [SOSL94] has a processing gain of around 30 dB and could handle a few hundred ($100\sqrt{10} \approx 316$) interfering signals each with power equal to the desired signal.) A bit error rate as low as 10^{-6} is more than sufficient to build a very reliable system with error correcting codes.

There is also a practical limit to the operating frequencies of electronic circuits. Practical RF circuits (circuits designed to operate at radio frequencies to amplify, mix, and filter radio signals) can be straightforwardly built today with maximum operating frequencies of over 5 GHz. But clock frequencies of digital circuits, of the sort needed for detection (e.g. Viterbi) or to generate and control the spreading code in a direct sequence spread-spectrum system, are presently limited to a few hundreds of MHz. If a system is to be low cost and low power (so that it could be operated using small batteries for power) the maximum digital clock frequencies will need to be kept to only a few MHz. Limiting the maximum rate of the spreading code limits the maximum width of a single channel.

[11] Chapter 5, page 1109.

There is another practical limit on digital radio links that has nothing to do with spread spectrum, but is relevant in this work if comparisons are to be made with clear channel systems. In a clear channel, where the transmitter can deliver sufficient power to the receiver to achieve a high signal-to-noise ratio, it is in practice difficult to reliably achieve a bit rate through the channel much greater than a few bits per second per Hertz of channel bandwidth. This limitation is because of the logarithm in the Shannon capacity theorem (Eq. 1.3). Once the signal-to-noise ratio is significantly greater than one, further increases in capacity would require exponential increases in signal-to-noise ratio.

1.4 The challenge

Given the model outlined above, the engineering challenge is to devise a behavior of the stations such that, without the benefit of either planned deployment or external infrastructure, they may organize themselves into an effective communication network. The design space is large; there are many plausible schemes for routing, and scheduling of the transmissions, encoding the data into the channel, controlling power, and managing the queues. The principal challenge that must be confronted by any scheme is self-interference, particularly if the system is to remain functional while scaling to many stations. This work develops an approach to this challenge, with particular attention to issues of scaling and decentralization.

The remaining section of this chapter will discuss the related work. Chapter 2 will outline the approach after presenting an analysis of the how the interference levels grow as the system scales to millions or billions of nodes. A practical method for scheduling transmissions to avoid collisions, without the need for global synchronization, will be presented in Chapter 3. Chapter 4 demonstrates the techniques in simulation and measures the signal-to-noise ratios in simulated networks of 100 and 1,000 randomly-placed stations. Chapter 5 observes some problems with system performance with attention to two issues: congestion and delay. Chapter 6 concludes this work and discusses topics for further research.

1.5 Context

Related work includes work in two traditional areas: packet radio networks and multiple-access communication theory. The prior work can best be understood by attention to how the models used for propagation and successful reception differ in these traditional areas of research.

1.5.1 Multiple-access communication theory

Work in multiple access communication theory usually uses a degenerate propagation matrix H where all rows are identical, which implies that all receivers receive the same signal.[12] Usually the columns are also assumed to be identical, implying that each transmitter is received with identical power. Given this model, the engineering problem is to design the transmitted signals and the detectors to optimize some goal, usually to maximize the communication rate. There are three basic methods of multiplexing: frequency division, time division, and code division.

Frequency division is the most straightforward method of managing the separation of users; different users of the spectrum are isolated from each other in frequency and the receiver can use a band pass filter to separate the desired signal. Frequency-division multiplexing (FDM) is most useful when the relationship between transmitter and receiver is a fairly permanent one. If this relationship is transient, then FDM leaves all the problems of multiple access communication to the problem of transient channel assignment, which is itself a formidable problem. Even if the relationships are fairly stable, long-term assignment to exclusive channels precludes the statistical sharing of the channels, leading to inefficiency if the sources are not continuous.

Time-division multiplexing (TDM) can allow for statistical multiplexing of the traffic and eliminates the problem of transient channel assignment, but introduces the problem of resolving contention for the channel. There are two traditional approaches to resolving the contention: random-access schemes and explicitly-scheduled schemes. Random access schemes (such as ALOHA[13]) can nicely solve the problem in situations where all the stations are attempting to communicate with a single base station (such as in the original ALOHA network), or when all the stations can hear each other equally well (such as on an Ethernet). In more complicated scenarios involving non-uniform propagation and non-centralized traffic, the potential performance of random-access schemes is less well understood. Scheduled schemes are distinguished from random-access schemes in that they require synchronization and system-wide coordination.

Code-division multiplexing is the term used to denote spread-spectrum techniques of multiplexing where the signals are allowed to overlap in time and frequency. Redundancy in the coding is exploited to demultiplex the signals at the receiver.

1.5.2 Packet radio networks

Work that goes under the heading *packet radio networks*[14] can be distinguished from work in multiple access communication theory by the use of propagation models that do not have all receivers receiving the same signal. Often a model known as *transmission radius* is employed.[15] This model is equivalent to setting the terms in the propagation matrix H to either one or zero depending on the distance between the two stations. If the distance between the two stations is less than the transmission radius, then the corresponding entries in the matrix are one. If the distance is greater, the entries are zero. While a zero entry implies that direct communication is not

[12] See [Abr93] and [Gal85] for an introduction to this field.

[13] originally described in [Abr70]

[14] Good overviews of the available literature on packet radio networks can be found in [KGBK78] and in [IEE87].

[15] Examples include [KS87].

possible, it also implies incorrectly that no interference can result (within the model) between the two stations. The usual requirement for successful reception in these models is that exactly one signal be present at the receiver for the duration of the reception of a packet. With these simplifying models, the problem is reduced to a problem of scheduling transmissions so that no overlap occurs at an intended receiver.

Unfortunately, this transmission radius model neglects interference effects from remote stations. Chapter 2 will show that the cumulative effect of interference from remote stations in a large system of stations is significant. Thus the binary model of propagation and the simplified model for successful reception are insufficient for examining the performance of a packet radio network as it scales to a large number of stations.

In this work a more realistic view of radio propagation will be maintained, and the combined effect of distant interference will not be neglected. This work is the first known work in designing packet radio network systems that: Uses a more realistic propagation model; Examines closely issues of scaling; and Produces a scheme for channel access that requires neither centralization of control nor hop-by-hop receiver feedback to avoid packet loss due to collisions.

Many sentiments underlying this work are expressed in [Kar91]. This work goes further by analyzing the growth in noise as the system scales and by presenting a system design. The system designed here eliminates the problem of packet loss due to collisions, a problem identified in [Kar91] as important to the efficient operation of packet radio networks. [Kar91] expresses a widely held view by saying, "Schemes that rely on receiver feedback (as opposed to channel sensing at the transmitter) to avoid collisions would seem to be the only practical approach to this problem." The method of preventing collisions presented in Chapter 3 requires neither receiver feedback nor channel sensing at the transmitter, and yet eliminates collisions without requiring anything other than pairwise coordination.

Chapter 2

Design Strategy

The principal challenge in designing a large packet radio system is managing and coping with the interference that comes from within the packet radio system itself. How this challenge is approached depends on how interference and propagation are modeled. Often interference is viewed as an all-or-nothing phenomenon. Some models for propagation limit the potential sources of interference to only a few nearby neighbors (e.g. those within a transmission radius).[1] In other models, all stations can hear and receive interference from all other stations equally well.[2] This work adopts a more complex model. We note that interference is a quantifiable phenomenon and is measured by the resulting signal-to-noise ratios at which packets are received. This work will view the interference as resulting from two additive components: (1) the interference resulting from the large number (possibly millions) of distant stations in the general area, and (2) particularly strong interference from individual nearby stations. These two sources of interference will be approached separately. In this chapter we will approach the first component by determining its magnitude and then controlling its level through power control and routing (we will see that we cannot, in general, eliminate it and must just cope with it as best we can). In the next chapter we will present a way of scheduling transmissions in a purely local manner to avoid interference from nearby transmissions (including the receiving station's own transmitter).

One approach often used when interference is limited is to try to schedule transmissions so that each packet can be transfered without experiencing any interference from any other transmissions.[3] For our completely filled-in propagation matrix, this approach would require coordination between all stations participating in the system and exclusive one-at-a-time use of the channel. This coordination would be challenging if there are many (millions of) stations. If the propagation matrix is not completely filled-in (because some stations can neither interfere with reception at, nor communicate directly with, some other stations) then the coordination problem

[1] e.g. see [KS87] and [TK84]

[2] e.g. see [Gal85]

[3] e.g. see [BG87] Chapter 4

may become even more complicated.[4] In either case, if the successful transfer of a packet requires that millions of stations refrain from interfering with its reception, then a single failure (to refrain from interfering) might disable the whole system.

A second approach is to require only local coordination. Distributed systems that have a large degree of interdependence are generally not as robust as systems with a lesser degree of interdependence. But if the degree of interdependence is inherent to the problem to be solved, as it might seem to be in the case of a large packet radio network, the designer might have little choice. Fortunately we can adequately treat the interference from distant transmissions as noise and use spread-spectrum radio techniques to communicate in the presence of this noise.

In this chapter we will examine this second approach and outline a design strategy based upon it. First, the level of interference that must be tolerated will be assessed, and then a strategy will be presented for ensuring that all packet transmissions can be received with sufficient signal-to-noise ratio without requiring global coordination or synchronization to control access to the channel. In the next chapter we will develop a method of scheduling to avoid interference from nearby transmissions without increasing interdependence.

2.1 Noise levels in a large system

Spread-spectrum techniques of modulation and detection provide an ability to communicate in the presence of interference. But spread-spectrum methods are not cure-alls as the achievable data rates are bounded by the Shannon limit. In the presence of high levels of interference, signal-to-noise ratios will be reduced, and hence the rate of communication will be reduced. If the signal-to-noise ratios sink as the system scales, then the communication rates must sink as well. (The relationship between signal-to-noise ratio and the Shannon bound on communication rate is essentially linear when the signal-to-noise ratio is significantly less than one.) This section will examine how the signal-to-noise ratios decline as the system scales.

Pessimism must be a guide when attempting to evaluate the level of interference in a system of stations that will be deployed without the chance to engineer the number or the placement of stations. But in this case, too much pessimism leads to the conclusion that the whole scheme is unworkable. We will first examine a model with a bit too much pessimism, and then refine it slightly to return to a workable level of noise.

[4] Two stations that are incapable of direct communication may have to coordinate their use of the channel to avoid causing interference with receptions at some other point in the network. The coordination is made difficult by their inability to communicate directly.

Assume that stations are distributed at some average density ρ throughout the infinite plane, and that each station is operating its transmitter at unit power output and at duty cycle η. Then the power radiated per unit area in the plane is (on average) $\eta\rho$. For a receiver located in the plane, the power level received Γ relative to the power received from a station at a distance of one characteristic length $R_0 = \rho^{-\frac{1}{2}}$, can be computed by integrating:

$$\Gamma = \int_{R_0}^{\infty} \frac{1}{r^2}\eta\rho 2\pi r \, dr. \tag{2.1}$$

Unfortunately, with the infinite bound, the integral diverges. Hence, for a receiver located in an infinite plane with a uniform and finite density of transmitters, the received power level would be infinite. The signal-to-noise ratio would be zero regardless of the source of the signal, so no communication would be possible.[5] There are a number of ways out of this conundrum. The key is to notice that the integral just barely diverges. For example, the slightest bit of atmospheric attenuation, which would introduce an $e^{-\alpha r}$ factor to the integrand, would make the integral converge to a constant. Nor does the integral diverge if we integrate out to some reasonable bound, stopping short of infinity.

Fortunately, we do not live on an infinite flat earth. In UHF and higher bands, only stations that are not hidden over the horizon can contribute to the interference at a receiver.[6] Hence the population of stations that are able to interfere with a given receiver will be limited to those in the same geographic region. If the earth's surface were perfectly spherical and all antennas were at the same height, then this region would be the interior of a circle. This limit on propagation *is* well modeled by a "transmission radius", but it is not a radius that can be engineered into the station or controlled as the system operates, but rather a consequence of station placement and surrounding terrain. A metropolitan area on flat terrain (or nestled in a bowl-shaped valley) may have all stations within direct line-of-sight propagation, hence the circle could cover at least an entire metropolitan area. The model for propagation is then $\frac{1}{r^2}$ within a circle encompassing a metropolitan area, with no interference from any stations outside the circle.

The growth in the overall level of interference as the system grows in number (and density) can now be estimated. Assume that M interfering stations are distributed randomly within a circle of radius R, and that stations outside the circle can be

[5]This observation is similar to a troublesome answer to the question "Why is the sky dark at night?" (Olbers' paradox). If we assume an infinitely large and infinitely old universe, a constant density of galaxies in the universe, and that each galaxy radiates a given amount of power, and then perform a similar integration (this time in three dimensions) we can conclude that an infinite amount of power should be impinging upon our eye when we look up at the sky. Presumably we have made an assumption that is not true for the universe in which we reside.

[6]The term *horizon* here means something slightly different at radio frequencies than it does at light frequencies. At radio frequencies there is a significant amount of propagation over the visual horizon, but a horizon-like attenuation of signals occurs nevertheless at a distance further out. The distance at which this attenuation occurs is sometimes called the *radio horizon* and is often modeled as if it behaved like a visual horizon of an earth with the radius increased to $\frac{4}{3}$ of the actual earth's radius. For more see [RR64] Chapter 4.

ignored. The average density ρ is then $\frac{M}{\pi R^2}$. As the number of stations M increases, so does the density. The distance to nearest neighbors also decreases, remaining proportional to the distance $R_0 = \rho^{-\frac{1}{2}}$. The signal level S from such a nearest neighbor transmitting with unit power would be

$$\begin{aligned} S &= \frac{\alpha}{{R_0}^2} && (2.2)\\ &= \frac{\alpha}{\left(\frac{1}{\sqrt{\rho}}\right)^2} && (2.3)\\ &= \alpha\rho && (2.4)\end{aligned}$$

where α depends on the antennas and wavelength used. The total power of interfering signals, N, ignoring the contribution from local interference inside the circle[7] of radius $R_0 = \rho^{-\frac{1}{2}}$ (and ignoring other sources of noise) can be calculated as

$$\begin{aligned} N &= \int_{R_0}^{R} \alpha\frac{1}{r^2}\eta\rho 2\pi r\, dr && (2.5)\\ &= \alpha\eta\rho 2\pi \ln r]_{R_0}^{R} && (2.6)\\ &= \alpha\eta\rho 2\pi \left[\ln R - \ln R_0\right] && (2.7)\\ &= \alpha\eta\rho 2\pi \ln\frac{R}{R_0} && (2.8)\\ &= \alpha\eta\rho 2\pi \ln\sqrt{\frac{M}{\pi}} && (2.9)\\ &= \alpha\eta\rho\pi \ln\frac{M}{\pi}. && (2.10)\end{aligned}$$

So the signal-to-noise ratio (SNR) is

$$\begin{aligned} \frac{S}{N} &= \frac{\alpha\rho}{\alpha\eta\rho\pi \ln\frac{M}{\pi}} && (2.11)\\ &= \frac{1}{\eta\pi\ln\frac{M}{\pi}}. && (2.12)\end{aligned}$$

Thus the expected signal-to-noise ratio of a signal from one of the nearest neighbors depends only on $\ln M$ (the log of the total number of stations) and η (the duty cycle). Figure 2.1 shows a plot of the log of the signal-to-noise ratio as a function of the base ten log of the number of stations. The signal-to-noise ratio falls very slowly, approaching -20 dB for $\eta = 1$ as the number of stations approaches 10^{12}. This observation is encouraging. The signal-to-noise ratio of a neighbor's transmission falls slowly even as the number of stations grows exponentially (even with $\eta = 1$, it does not reach -21 dB until 10^{18} stations).

[7] We cannot just simply drop the lower bound of this integral to zero for then the integral would blow up. But interference from local sources will be managed separately and explicitly later. Choosing a lower bound of $\rho^{-\frac{1}{2}}$ is reasonable (as stations closer than this distance are clearly local) and convenient (because it makes the algebra work out nicely).

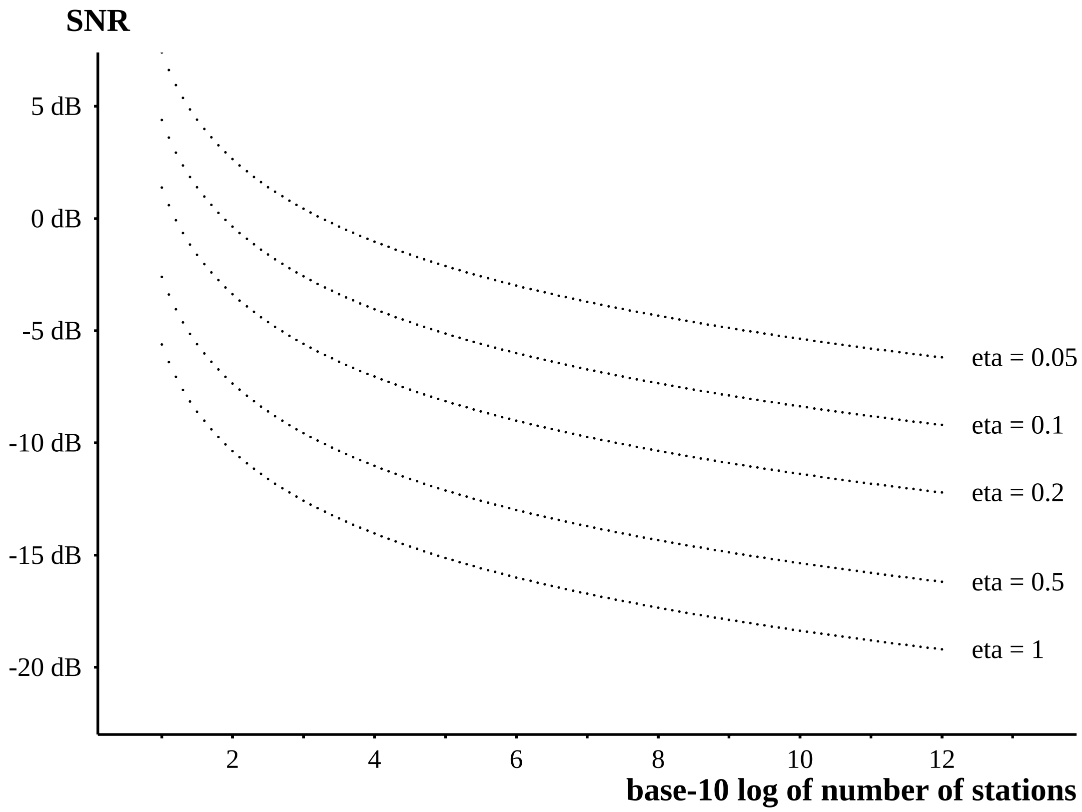

Figure 2.1: Decline of the signal-to-noise ratios as M, the number of stations, grows (Eq. 2.12). Each member of the family of curves is for a different value of the duty cycle, η.

According to this model, for almost any imaginably large population of stations in a packet radio network, direct communication at a definite rate with nearby neighbors (neighbors nearer than $\rho^{-\frac{1}{2}}$) should remain possible, provided that the stations can cope with signal-to-noise ratios of around -20 dB. Indeed, by Shannon's capacity theorem, $C = W \log\left(1 + \frac{S}{N}\right)$, even with a signal-to-noise ratio of one part in one hundred, the theoretical communication capacity remains non-zero. In this case, $C = W \log_2(1.01)$, thus $\frac{C}{W} = 0.014$, or theoretical capacity of approximately 14 bits per second per kilohertz of channel bandwidth.

Thus far, these calculations have assumed $\eta = 1$. Stations will have to spend at least some of the time listening. For more reasonable values of η, the noise levels are improved. At an average duty cycle of one quarter, $\eta = 0.25$, the signal-to-noise ratio is better by a factor of four, or $+6$ dB. The resulting signal-to-noise ratio of around -14 dB yields a theoretical capacity of around 56 bits per second per kilohertz of channel bandwidth, but only when the station is transmitting. There is no gain in throughput by reducing the transmit duty cycle in a large noisy system. Halving the duty cycle increases the average signal-to-noise ratio by a factor of two, which improves the data rate (while transmitting) by approximately a factor of two, but would result in no net gain in performance since the transmitters would then be operating for only half of the original amount of time.

What about neighbors that are not so near? The placement of stations or the routing algorithm might require direct communication between stations that are more than $\rho^{-\frac{1}{2}}$ distance apart. Free space radio propagation falls off by a factor of four, or -6 dB, for each doubling in distance, so we can expect that a station at a distance of $2\rho^{-\frac{1}{2}}$ will be heard with a signal-to-noise ratio reduced by a factor of four, or -6 dB. Another factor of two in distance would be another -6 dB. Each 6 dB reduction in signal-to-noise ratio reduces achievable throughput by a factor of four. Thus, in large scale packet radio networks, direct communication (at a reasonable rate) will be possible only with nearby neighbors.

While this model gives us an estimate of the overall noise levels, the exact value of the signal-to-noise ratio will depend on the details of station placement and transmission control. There remains the problem of managing the transmissions on an individual and local basis. For example, interference from a very near station might amount to much stronger interference than the aggregate interference produced by distant stations, but since a nearby station is local, the interference can be managed locally.

2.2 Collisions

In more conventional models of packet radio networks (those involving a hard transmission radius and simple success-if-exclusive criterion for successful reception) the term *collision* is often used to describe how packets are lost due to interference. In the more sophisticated model used in this work (where the criterion for success is sufficiency of the signal-to-nose ratio at the receiver) the term collision may be misleading in that it suggests an overly simple interpretation of the interaction between packets at receivers. In the model in this work, whether or not a packet is received successfully depends on more than just the number of simultaneous signals at a receiver. Nevertheless, a taxonomy of collision types will help us to understand local interference, even in the context of our more complete model.

If a collision occurs, then it must fall into one of the following three cases (see Figure 2.2):

1. Collisions due to the transmission of another packet from a station not involved in the exchange of the dropped packet, which is not addressed to the station receiving the dropped packet.

2. Collisions due to multiple stations attempting to send packets simultaneously to a single station.

3. Collisions due to a packet arriving at a station while another packet is being transmitted by the receiving station.

This enumeration covers all possible cases of an interfering transmission. If the interfering transmission does not involve the receiving station, either as a receiver or transmitter, then it is a type 1 collision. If it does involve the receiving station as

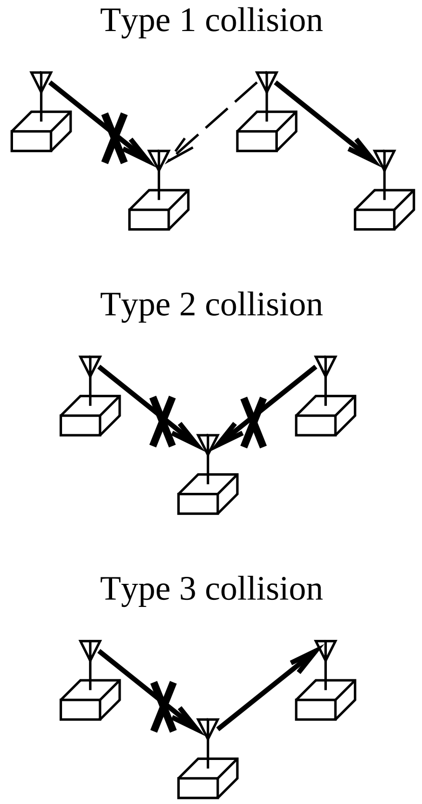

Figure 2.2: Examples of each type of collision. The X indicates the lost packet.

the intended target of the interfering transmission, then it is a type 2 collision. If it involves the receiving station as the sender of the interfering transmission, then it is a type 3 collision. Multiple collision types may occur simultaneously in more complicated situations.

Our use of spread spectrum can eliminate most packet loss due to type 1 collisions. If a nearby interfering station is transmitting, and the receiver is already prepared to cope with a signal-to-noise ratio of $\frac{1}{100}$ due to the (potential) overall level of noise, then in order to significantly increase the level of interference, a nearby station would have to be very near indeed. Even a station at one fourth of the $\rho^{-\frac{1}{2}}$ distance would have only a small effect on the total amount of interference.[8] So, in most cases, this type of collision is not a problem. When stations are so close that type 1 interference is a problem, then it really is a local problem, and the stations involved must cooperate and each must refrain from transmitting in a manner that interferes excessively with the receptions at its neighbor. A method of achieving this coordination in a decentralized manner will be presented in Chapter 3.

Type 2 collisions are very similar to type 1 collisions. The only difference is that the intended receiver of both transmissions is the same station. These can be eliminated by enabling stations to receive multiple transmissions in parallel. With

[8] A low-power signal added to a high-power signal yields a signal with power level not much different than that of the original high-power signal.

spread-spectrum radio receivers, elimination of packet loss due to this type of collision requires only multiple tracking and despreading channels. A multiplicity of despreading channels is already a common feature of existing spread-spectrum receivers. For example, GPS (Global Positioning System) receivers often have six or twelve despreading channels. With a sufficient number of despreading channels, packet loss due to type 2 collisions can be eliminated. The number of despreading channels needed will depend on the details of the routing and transmission control schemes used, but in any case, it should not be larger than the number of neighbors that might communicate directly with the station. This number should be small, since, as we have already seen, only nearby stations will be capable of direct communication over the din of background noise. (A routing strategy that will be presented at the end of this chapter was used in a number of simulations of randomly placed stations and the number of routing neighbors never exceeded eight.)

Type 3 collisions are a more difficult problem. The interference from a transmitter located with a receiver will be so powerful that no feasible amount of processing gain (even when combined with the isolation provided by the antenna duplexer) can achieve reception while the local transmitter is operating. But like the nearby case of Type 1 collisions, packet loss due to Type 3 collision is a local problem, and need not be solved globally. It is sufficient to ensure that the local transmitter does not operate at times when other stations might send a packet to the local receiver. A method of achieving this coordination in a decentralized manner will be presented in Chapter 3.

2.3 Design strategy

A design strategy for a viable packet radio network can now be devised. The first section of this chapter has shown that if there are many (millions of) stations in an area, the stations will be immersed in a din of interference. Nevertheless, by using spread-spectrum radio techniques with a moderately high processing gain (in the range of 20 dB to 25 dB) stations will be able to communicate directly with nearby neighbors (those stations within a distance of approximately $2\rho^{-\frac{1}{2}}$) even as the system scales to large numbers of stations. By using spread spectrum, the interference from distant stations can be treated as random noise, and no system-wide coordination is needed to manage the use of the channel.

By cooperatively forwarding packets, the stations may organize themselves into a fully connected network to allow communication beyond the immediate neighbors. Whether or not the network is fully connected will depend on if there are enough stations to blanket the area, if the stations are distributed uniformly enough, and what distance can be covered in one hop. The analysis that produced Figure 2.1 assumed that the neighbor was $\rho^{-\frac{1}{2}}$ distance away and that the interfering transmissions were at the same power level and were evenly distributed throughout the region. But the design will need to accommodate communication between neighbors that are not exactly at this distance, and will need to cope with varying densities.

If the stations are randomly and independently distributed in the plane at density ρ, and if $\rho^{-\frac{1}{2}}$ is the maximum distance that can be covered in a single hop, then the expected number of neighbors that a station will have is the expected number of stations inside a circle of radius $\rho^{-\frac{1}{2}}$, which is $\rho\pi\left(\rho^{-\frac{1}{2}}\right)^2 = \pi$. That the expected number is only π suggests that $\rho^{-\frac{1}{2}}$ may not be far enough to ensure connectivity. No claim about connectivity can be made without knowledge of the particular geometry, but it is reasonable to expect that variations in density will at some stations require reaching farther than to just three expected stations. Doubling the distance to $2\rho^{-\frac{1}{2}}$ (by increasing the processing gain by 6 dB or a factor of four) should suffice in most situations. Assuming again uniform distribution, the expected number of reachable stations would then be 4π. This doubling of range comes at the expense of a factor of four in raw throughput since the processing gain has been increased, and any further increase in range (by increased tolerance to interference) would impact throughput. (A doubling in range would quadruple the noise-to-signal ratios, reducing raw throughput by a factor of four since achievable throughput depends linearly on signal-to-noise ratio in a noisy system.) With the signal-to-noise ratios for stations at $\rho^{-\frac{1}{2}}$ distance in the -10 dB to -15 dB range for reasonable duty cycles, the need to budget around 5 dB of headroom for successful detection in the receiver, and the need for an additional 6 dB margin for more distant stations, the proper amount of processing gain is determined to lie in the range of 20 to 25 dB.

2.3.1 Power control

In the analysis so far, all transmissions were assumed to be at the same power level. In cases where stations are closer than maximum range, transmitting at full power is excessive. A more sensible approach is to control the power transmitted. If the stations are controlling power but are still transmitting with the same average power density as before, then the analysis of average signal-to-noise ratio remains the same. But by reducing power in situations where lower power levels can still deliver a sufficient signal-to-noise ratio at the intended receiver, interference to other stations can be reduced, increasing the signal-to-noise ratios in receivers at other stations.

The ratio of the average noise power level to the average signal level should not be affected by power control. This criterion strongly suggests a power control algorithm: *transmit with sufficient power to deliver a constant pre-determined amount of power to the intended receiver.*[9] The choice of pre-determined power level is not critical, because increasing or decreasing it will just slide all power levels in the system up or down, maintaining the same ratios everywhere, including the received signal-to-noise ratios. By fixing the received power level, the variance in signal-to-noise ratio can be reduced.

This power control algorithm is also appealing for another reason: as different areas in a network may vary in density, stations will automatically compensate by controlling power levels to deliver the same amount of power to the intended receiver.

[9]An even better idea might be to transmit with sufficient power sufficient to just achieve the necessary signal-to-noise ratio. Achieving that would require knowing what the noise levels at the receiver will be, but the recent past might be a good-enough predictor of the future noise levels. This idea will not be explored further here.

The power density then remains constant: if the density in some area is quadrupled, the distance to neighbors is cut in half, so power levels can be cut by a quarter, maintaining constant power density as the station density varies. Therefore the analysis from the first section of this chapter remains applicable even to networks employing this method of power control.

2.3.2 Minimum-energy routes

We already know that packets traveling more than $2\rho^{-\frac{1}{2}}$ must be routed through intermediate stations. When there is a candidate intermediate station, and an option exists of either sending the packet directly or through the intermediate station, which should be done? In some sense, with power control, this choice will always exist, for if we choose to send the packet directly, we can increase the transmitted power to deliver the proper amount of power to the intended receiver. But if stations routinely did this to communicate directly with distant stations (stations significantly farther than $2\rho^{-\frac{1}{2}}$ where ρ is in this case the density in the immediate area), then we would be violating a crucial assumption of the earlier analysis: that the power density is constant and roughly $\rho\eta$. Violating this assumption in this way would significantly reduce the signal-to-noise ratios. Such high-power transmissions would also cause a high level of interference to the (presumably numerous) neighbors close to the transmitter. The criteria used to determine routes will need to prefer the short hops, which produce less interference, and avoid skipping over intermediate stations.

In an actual network, the stations may not know where they are geometrically, but they will be able to observe the path gains between themselves and construct entries in the propagation matrix H for the hops that are usable. A criterion for selecting routes that is directly determinable from the propagation matrix would be particularly convenient. A routing criterion that is directly determinable from the propagation matrix and that seems to meet our needs is *minimum-energy routing*.[10] Consider the following scenario:

Station A wishes to send a packet to station C. Station B is a candidate intermediate station. Using minimum-energy routing, station B should be used as an intermediate hop if it reduces the total amount of interference to a distant station D *caused by the movement of this packet*. If B is used as an intermediate hop, the duration of the interference (to station D) caused by this packet will be doubled, but the power levels

[10]The idea of minimizing energy is mentioned in [Kar91] and credited there to David Mills.

of the two transmissions may be much less than the single hop transmission. They would be less by as much as a factor of four if station B is exactly centered between stations A and C. Taking this intermediate hop would reduce the level of interference by as much as a factor of four at station D. Then the total energy (power integrated by time) of the interference to station D (or any other distant station) caused by this packet will be reduced by a factor of two.

Geometrically, with $\frac{1}{r^2}$ free-space propagation, minimum-energy routing will always take the intermediate hop if it lies within the circle with diameter $\overline{AC}$, so it will choose routes that respect the local density and will not skip over intermediate hops. By using minimum energy routing, the interference from the ensemble effect of many packets traversing the network is kept to a minimum, enabling as much raw data throughput as possible across each local hop. There are trade-offs. For example, this approach does not minimize latency. The multitude of store-and-forward delays incurred by always taking intermediate hops will adversely affect delay. This routing method may be inappropriate if delay is the overriding concern.

Minimum-energy routes are straightforward to compute. The common algorithms for computing min-cost paths in networks[11] can be used to find the least-cost paths in the propagation matrix H, where the costs are the reciprocal of the path gains. (The reciprocal of the path gain is proportional to the power that would be used with power control.) The algorithm is also easy to distribute. Each station need only remember the next hop for each potential destination and the total energy along that route to the destination. Hop-by-hop routing is possible since, at each station, each transit packet will be routed as if it had originated at the transit station. In other words, a minimum-energy route from A to C that goes through B will use the same route from B to C as any other route that goes through B to get to C.

In a somewhat contrived example, minimum-energy routes will provide optimal throughput. Assume that a number of stations in a large network have a steady supply of traffic, each for a nearby station, and that in each case there is an otherwise idle station between the source and destination of the traffic. In each case, routing through the intermediate station is a lower energy path than taking a direct hop to the destination. If there is otherwise no contention for transmitter time (the intermediate hops are otherwise idle), and assuming that the times when a station may transmit and receive are fixed ahead of time (as they will be in the scheduling method presented in the next chapter) then the achievable throughput will depend only upon the bandwidth and the signal-to-noise ratio.[12] Since signal-to-noise ratio is improved everywhere if every packet is sent through the intermediate lower-energy route, throughput is maximized by minimizing the energy.

In practice other considerations may be in play, such as delay performance and contention for transmitter time at intermediate stations. With more complicated metrics of performance, minimum-energy routing might not be optimal. However in

[11] For example, the Distributed Asynchronous Bellman-Ford Algorithm is described in [BG87].

[12] In practice, stations might be manufactured to use a fixed amount of processing gain, in which case the throughput would not incrementally depend on the signal-to-noise ratio. Here we are assuming that stations will make the most of the signal-to-noise ratio they are experiencing and communicate as fast as possible.

Chapter 5 we will see a demonstration where an attempt to improve on minimum-energy routes leads to drastically worse system performance.

2.4 Summary

A functional packet radio network of millions of stations scattered throughout a metropolitan area, with no central control or coordination and with no planned deployment of infrastructure, appears to be feasible by employing spread spectrum with a processing gain in the region of 20 to 25 dB, and if packets are routed over short hops without skipping over potential intermediate stations. Minimum energy routing is a straightforward and viable way of determining routes. Issues unresolved in this work thus far include avoiding packet loss due to local interference, and overall system performance. These are the topics of the next chapters.

Chapter 3

Managing Local Interference

Chapter 2 presented an analysis of the overall noise levels in a system of packet radio stations, and presented a design strategy for enabling such a system to remain effective while scaling to a large number of nodes. While Chapter 2 dealt with the issues of global interference, there still remains the problem of preventing destructive local interference. That is the topic of this chapter.

Even if the level of background interference is at an acceptable level, interference from a source located near a receiver may alone raise the level of interference to a destructive level. The simplest and most important case of interference would be a station's transmitter interfering with the reception of a packet by its own receiver.

Traditional schemes for packet radio networks involve either a global scheduling scheme[1], or they allow collisions and use hop-by-hop acknowledgments and retransmissions to recover from lost packets. While either of these schemes may be used to cope with packet loss due to local interference, neither of these schemes is acceptable for our purposes. Packets lost due to collisions need to be retransmitted, which wastes resources. If the per-hop packet loss rate is at all noticeable, hop-by-hop acknowledgments are needed, otherwise the end-to-end packet success rate would approach zero for paths consisting of many hops. Hop-by-hop acknowledgments waste resources, since end-to-end acknowledgments are already likely to be used by higher level protocols. Global scheduling can be used to prevent all packet loss due to collisions, but is unworkable if the system is to scale to millions of stations.

What is needed is a method for preventing local interference that can ensure that transmitted packets are not interfered with at the intended receiver, without requiring global coordination or synchronization. Such a method is presented in this chapter. First, a method of avoiding interference from just the receiving station's own transmitter will be presented, and then the method will be extended to avoid interference from all of the nearby stations that matter.

[1]See the Packet Radio Networks section in Chapter 4 of [BG87].

3.1 The scheduling method

When a packet is to be sent to another station, it must be sent at a time when its reception will not be prevented by the level of interference. If we are only concerned about interference from the receiving station's own transmitter, then meeting this constraint can be easily accomplished: the sending station only needs to know at what times the receiving station may be transmitting. If the receiver's schedule is known by the sending station, then the sending station can choose to send the packet at some other time. In order to make its schedule known to its neighbors, a station (the intended receiver) needs to schedule the times that it may be transmitting and inform its neighbors what those times will be.

Global clock synchronization is not required. Only the ability to relate one station's clock[2] with another's is required. This ability can be accomplished if stations occasionally rendezvous and exchange clock readings. Differences between clocks and small differences in clock rates can be mutually modeled, and the resulting models, along with the published transmit schedules (each reckoned by the publishing station's clock), can be used by neighbors to predict when a station will be transmitting.[3]

The method is as follows. Each station will independently produce and publish a schedule for itself. A schedule divides time into receive windows and transmit windows for a station. The schedule published by a station is a commitment by that station to listen (refrain from transmitting) at particular times in the future (during the receive windows). A station with a packet to be sent to another station will compare its own schedule with the receiving station's schedule and send the packet during a time when one of its own transmit windows overlaps with a receive window of the receiving station enough to handle the packet length. Each station only needs to be aware of the schedules of the immediate neighbors to which it might be directly sending packets.

The schedules must be devised so that the overlaps will exist for all pairs of stations that may communicate directly. Simple periodic schedules will not do. If two stations using simple periodic schedules were to happen to be running at the same phase, then communication between them would not be possible. (Choosing clock values to avoid unfortunate phase offsets is not possible if the stations are to produce the schedules independently without any global coordination.) This problem is solved by using schedules that are random or pseudo-random. If each station independently chooses a random schedule, then these schedules will allow many opportunities to communicate between any pair of stations.

---

[2]The term *clock* as used in this work does not imply knowledge of what time it is. Here clock just means something that advances at some known rate.

[3]See Troxel [Tro94] for an example of how the drift of a clock driven by a quartz oscillator can be modeled from historical data and for a demonstration of how the model can then be used to accurately predict future drift. Extending his techniques to enable mutual modeling between two drifting clocks is straightforward: His model for offset is a quadratic function of time; since the difference between two quadratics is also a quadratic, no change should be needed in his basic method.

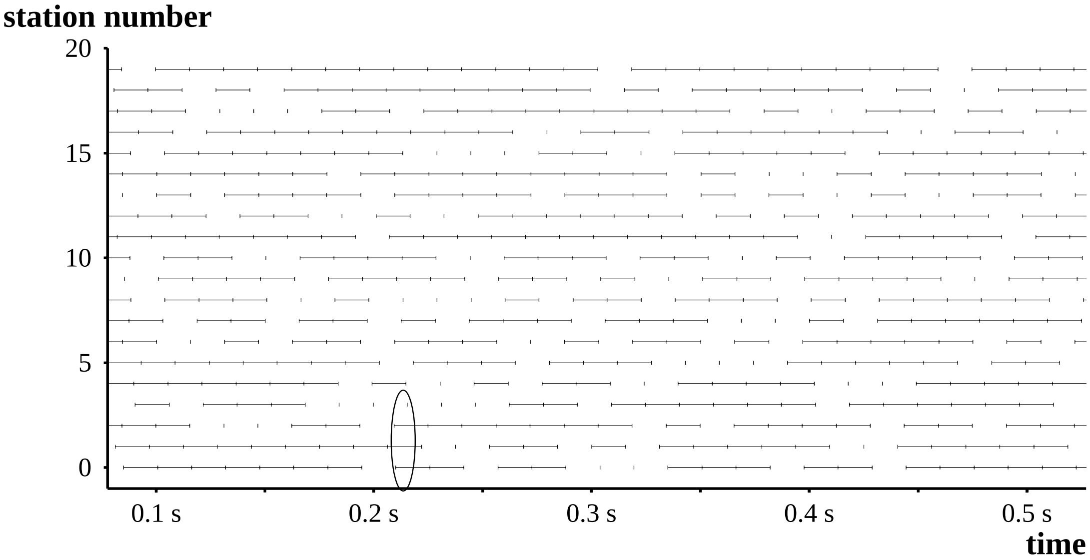

Figure 3.1: A sample pseudo-random schedule for 20 stations. The line is drawn in slots where the station is scheduled to transmit, and the line is omitted in slots where the station is scheduled to listen. The receive duty cycle (the average fraction of slots that are scheduled for listening) here is 0.3. To send a packet from one station to another, it must be scheduled to fit in a period of time when the sending station is allowed to transmit and when the receiving station is listening. For example, at the circled time above station zero could not send to station 1 or station 2, but could transmit to station 3.

3.1.1 Unaligned slots

One possible method of implementing the pseudo random schedules is presented here. Implementing the scheduling method requires a method of generating the random schedules and a method of communicating the schedules and the clock readings to neighbors. If each station's clock is set differently, then the stations can all use a single schedule, each reckoned by its own clock. With all stations using the same schedule, then only the clock readings need to be communicated between stations. Time can then be divided into equal size slots, again reckoned independently by each station's clock, and each slot designated to be either a transmit or receive slot.

The schedule (singular since all may be the same) is a binary-valued function of a clock reading that divides time into transmit windows and receive windows. The schedule will be divided into equal-length slots of length T_{slot}, with all times in a slot sharing the same value (transmitting or receiving). Whether a particular slot is for transmitting or receiving can be determined by using a hash function to hash the value of time at the beginning of the slot. If the hash value is less than a threshold, then the slot is a receive slot, otherwise, it is a transmit slot. The threshold is selected to achieve the desired duty cycle. The receive duty cycle, p, is the probability that a slot is a receive slot.

The slots are unaligned, as is shown in Figure 3.1. Between stations there is no

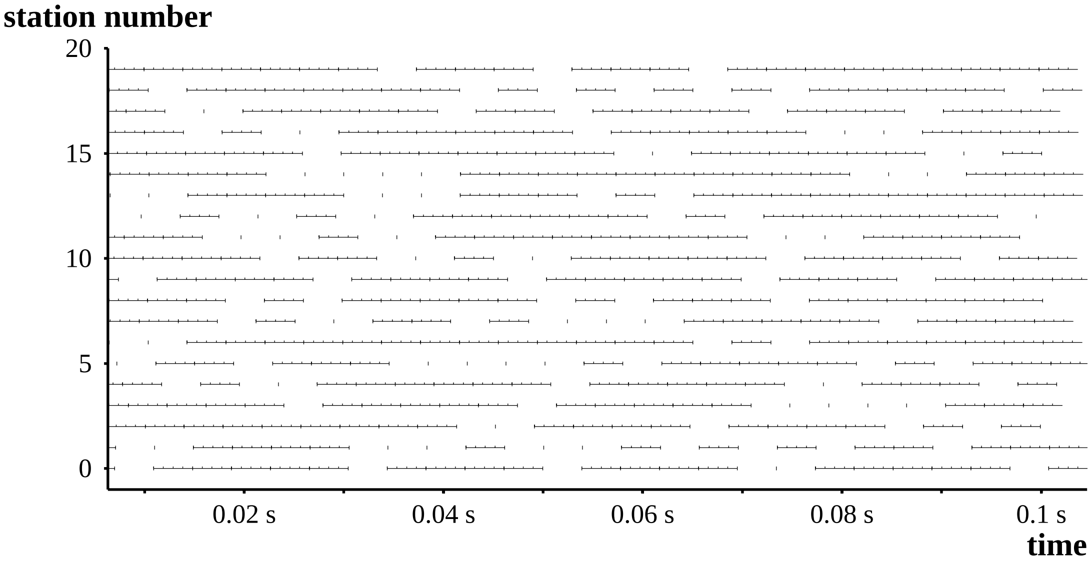

Figure 3.2: A sample pseudo-random schedule for 20 stations showing transmit slots divided into subslots. In this figure each transmit slot is divided into four subslots.

synchronization of clocks and the slot boundaries at a station are determined by the station's own clock. A slot at a station will overlap with parts of two slots of a neighbor, unless that neighbor's clock happens to be exactly aligned. The amount of overlap (or phase difference) between two stations is random and remains constant except for any drift caused by differences in clock rates.

Each station needs to set its clock in a way that ensures that it is set differently than each of its neighbors with which it will be directly exchanging packets. If the clocks were not set differently, then the identical schedules would prevent communication between the two stations. Clocks with only a small difference (of less than one slot time) would not have the full expected amount of time available between them to communicate as their transmit schedules would be somewhat correlated. But if there is at least one slot's time difference between the two clocks, then the schedules will be uncorrelated and we can treat each station's schedule as random and independent.

A simple way to set the clocks so that they are different is to set them independently to a random value. The probability that a station's clock may by chance be set to a value that is close enough to the value of neighbor's clock to cause trouble can be made arbitrarily small by increasing the number of significant high-order bits in the clock. Each additional high-order bit added and initialized randomly will reduce the probability of such an unfortunate coincidence by a factor of two.

3.1.2 Subslots

A station with a packet to send to a neighbor will need to fit the packet into its own schedule and into the schedule of the receiving station. One way of fitting the packets into schedules will be presented here.

A packet must be fit into a period of time when the sending station is in a transmit window and the receiving station is in a receive window. Some packets may require more than just one slot at one of the two stations (depending on the size of the packet in time and the phase offset between the schedules of the two stations). Packets longer than a single slot time (T_{slot}) would need to occupy time in more than one slot at each station, and hence would be more difficult to schedule. Packets smaller than $T_{\text{slot}}/2$ could always be scheduled so that they only involve one slot at each station. But if all packets were that large, than as much as half of the time available to send to a particular station might be in unusable fragments of time too small to hold a packet that large. Even if packets are smaller than half of a slot time, efficiently packing variable-length packets into the schedule may be a difficult problem, particularly if the packets are for a number of different neighbors.

The maximum throughput we can hope to achieve (with ideal packing) is determined by the receive duty cycle p. We assume the same value for p is used by all stations. For a given p, the total fraction of time that communication is possible from a station, A, to a particular neighbor B, is $p(1-p)$. In the limit of small packets, $p(1-p)$ is the maximum possible throughput (relative to the raw data rate at which stations communicate) from A to B.

Here we will simplify the packing problem by limiting the maximum size of a packet to a small fraction of a slot. We will divide each transmit slot into a whole number of fixed-size subslots, each just large enough to carry a maximum sized packet. We will send at most one packet in each subslot. (Or alternatively we could say that all packets are of fixed size equal to the subslot size.) Now when scheduling a packet we can just consider one-by-one each transmit subslot and check for an acceptable state at the receiver for the duration of the subslot.[4]

Since we have divided each transmit slot into m equally sized subslots and only consider sending one packet in each subslot, the maximum throughput is determined by the number of usable subslots. A set of conditions sufficient to ensure that transmission in a particular subslot at station A will be successfully received at station B is:

- The subslot does not straddle a slot boundary at station B. This occurs with probability $\frac{m-1}{m}$.
- The subslot begins in a receive slot at station B. This occurs with probability p.
- The subslot is in a transmit slot at station A. This occurs with probability $(1-p)$.

Since each condition is independent, $\frac{m-1}{m}p(1-p)$ is the fraction of subslots, or time, in which all three constraints are met. Therefore a throughput of at least $\frac{m-1}{m}p(1-p)$ can be achieved to a single neighbor. This bound is a slight underestimate because we have removed from consideration all subslots that straddle the boundary between

[4]In some systems (depending on the parameters) the method presented here will result in very small packets that may be unnecessarily wasteful of bandwidth (since the per-packet overhead will be amortized over very little payload). Nevertheless, the approach presented will serve to illustrate the general performance characteristics of the system. In a low bandwidth system, the small fixed-size packets would be very much like cells in an ATM network. Larger packets of varying size that are typically found in computer networks would need to be fragmented into cells with small headers and then reassembled at the other end, much like how ATM networks will carry Internet IP packets.

two consecutive receive slots at the receiver. Also, in the case where a station has a number of neighbors, it may be able send packets of length $\frac{T_{\text{slot}}}{m}$ more than $\frac{m-1}{m}$ of the total time when communication is possible because some of those subslots that straddle the slot boundary of one neighbor may be usable to send to some other neighbor.

There are a few tradeoffs to be considered when selecting the size of slots and number of subslots. Smaller slots will reduce the scheduling delay by increasing the frequency of slots (which will reduce the expected amount of time until a suitable slot arrives—see Section 3.1.6). There will need to be at least a few subslots in each slot to get the throughput to a single neighbor up to a reasonable value (as the maximum throughput to a single neighbor with m subslots per slot is $\frac{m-1}{m}p(1-p)$). However, there will be some amount of per-packet overhead associated with each transmission. Furthermore, the spread-spectrum radios will need some amount of overhead time to synchronize the receiver's despreading code at the beginning of each subslot. There will also need to be some amount of per-packet overhead to carry information needed by the network for routing and control. Because of this overhead, in the limit of very small packets throughput would go to zero (as all of the channel time would then be used for overhead).

3.1.3 Multiple neighbor throughput analysis

A station with multiple neighbors will be able to transmit to some neighbor more than $p(1-p)$ of the time because at some of the times when one of its neighbors is not in a receive window, other neighbors will be. In particular, the fraction of the time when none of a set of k neighbors is listening is $(1-p)^k$ and the fraction of time in which this happens while the station is in transmit mode is $(1-p)^{k+1}$. The fraction of time that a station is in transmit mode in the first place is just $(1-p)$. Subtracting from this the amount of this time when no one is listening yields an expression, as a function of p, for Θ, the total throughput out of a station with k neighbors. The maximum total throughput of a station (assuming it has traffic for all of its neighbors) is thus:

$$\Theta(p,k) = (1-p) - (1-p)^{k+1} \tag{3.1}$$

The receive duty cycle, p, should be set to maximize $\Theta(p,k)$ The particular maximum will depend on the number of neighbors, k, but it is possible to choose a single value for p so that the throughput is near maximum for a range of k's. Figure 3.3 shows a plot of $\Theta(p,k)$ as a function of p for various values of k. (For any particular k, $\Theta(p,k)$ is maximized when $p = 1-(k+1)^{\frac{1}{-k}}$.) As can be seen in the figure, choosing a value for p near 0.3 will provide nearly optimal throughput for a wide range of values for k.

A more sophisticated system may have each station choose a different value for p after the number of neighbors is known (as a result of the routing computation). This scheme would considerably complicate the analysis because now the throughput would depend not only on a station's value of p and the number of neighbors, but also the value of p chosen by each of the neighbors. From the point of view of any

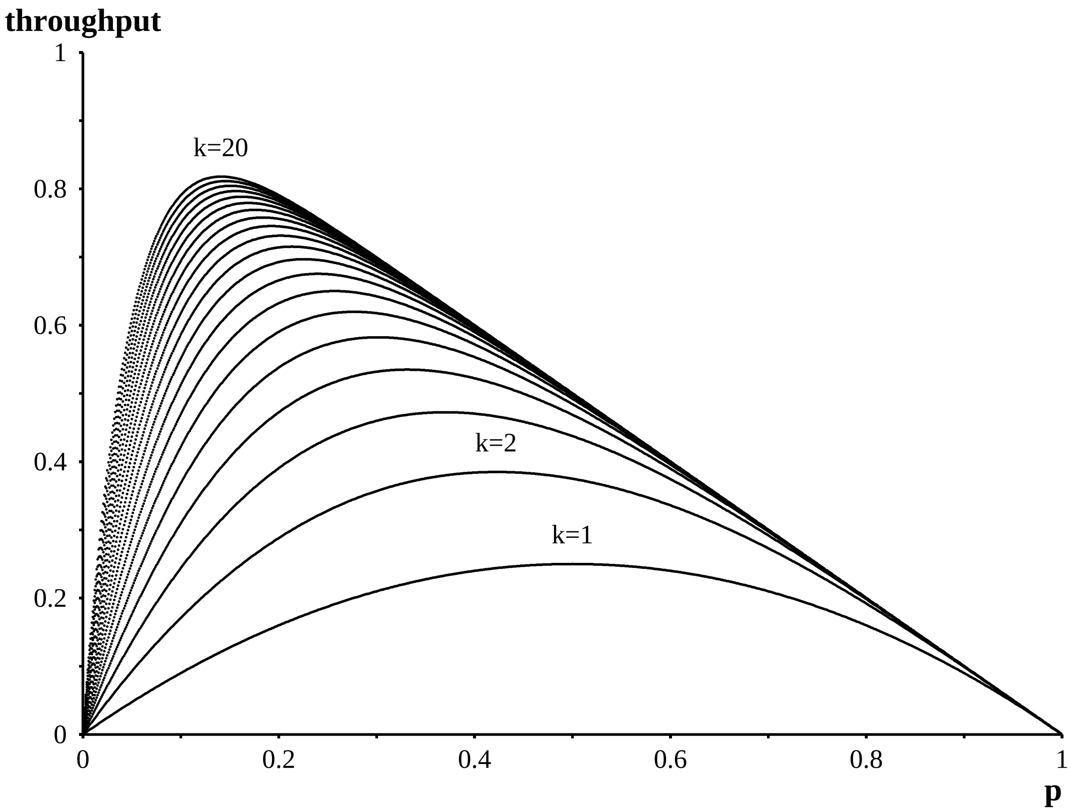

Figure 3.3: Variation of throughput, $(1-p)-(1-p)^{k+1}$, with receive duty cycle, p, for the number of neighbors k ranging from 1 to 20. Note that p near 0.3 produces good throughput for a wide range of k.

single station, throughput would then be best when $p = 0$ (always transmitting) and all neighbors choose $p = 1$ (always listening). Any attempt at optimization would have to be global. In this work, all stations will be assumed to use an identical value for p near 0.3.

3.1.4 Feasible throughputs

The analysis of throughput above assumes that we have plenty of traffic for each neighbor and that we want to maximize the total throughput out of a station (regardless of throughput to each destination). In an actual network, the demand for traffic out of a node may not be distributed evenly enough to make use of all opportunities to send to a neighbor. For example, if all of the traffic is for one single neighbor, then the throughput is at most $(1-p)-(1-p)^2$, regardless of the number of neighbors the sender has.

The additional constraints are that for each subset of k' neighbors, the total throughput to that subset of neighbors cannot exceed $(1-p)-(1-p)^{k'+1}$. There are $2^k - 1$ of these constraints, one for each non-empty subset of the k neighbors, which must be satisfied for a given set of flow rates to be feasible.

The choice of the receive duty cycle p is not influenced much by this observation.

If a station with k neighbors only has traffic for some of them, then the optimal choice for p is just the same as the optimal choice for p if there were fewer neighbors. Figure 3.3 already shows that 0.3 is reasonable value for p for a wide range of k's.

3.1.5 Scheduling the packets

Even though a given set of rates for traffic to various neighbors may be feasible (by the criteria above), a station will need to schedule the packets into the appropriate kinds of subslots in the appropriate ratios to achieve the particular rates. There will be three different kinds of subslots into which a packet may be scheduled:

1. subslots that can be used to send to neighbor A, but not to neighbor B,
2. subslots that can be used to send to neighbor B, but not to neighbor A,
3. subslots that can be used to send to either neighbor A or neighbor B.

If one destination is less popular than the other, then its packets should not be allowed to fill up all the sublots of the third kind. The subslots of the third kind will need to be allocated more heavily to traffic for the more popular station. For example, a station with two neighbors with p equal to $1/2$ should be able to handle a total demand for service of $3/8$ to these two neighbors.[5] The most asymmetric distribution of traffic that could be carried would be an outgoing rate of $2/8$ to one neighbor and $1/8$ to the other neighbor. Each of the three kinds of subslots will occur with frequency $1/8$. In order to support a traffic load with an outgoing rates of $2/8$ to neighbor A and $1/8$ to neighbor B all of the third kind of slot will need to be used for traffic to neighbor A. In this case, all packets for destination B will have to be confined to subslots of the second kind.

Fortunately, a simple-minded and oblivious method of scheduling the packets seems reasonable. (A non-oblivious scheme that tried to ensure that the packets are assigned to the correct kind of subslot would be complex. It would need to be aware in advance of the flow rates to the the various destinations and would have to base its scheduling decisions on all of this information.) By booking each packet in to the first available subslot at the moment it arrives, the bookings will tend to assign packets to the correct sort of slot. In the example above, a few packets for neighbor B may at first be booked into subslots of the third kind. Packets for neighbor A will also be booked into subslots of the third kind, filling slots into the future. As subslots of the third kind get filled up (by the excess demand), packets for station B will tend to always find an earlier booking in a subslot of the second kind.

There are a number of different schemes that could be used to schedule the slots. For example, separate queues could be maintained for each next-hop and queues serviced as the slots come along. The selection of queue to service could be made sensitive to the lengths of the various queues, possibly favoring longer queues over

[5]The value of $1/2$ for p was selected here just to make the illustration easy. For two neighbors and $p = 1/3$, the two illustrative throughput values would be $2/9$ and $4/27$ for a total throughput of $10/27$.

shorter queues in slots for which there is a choice to which neighbor to send. Further examination of these issue will not be presented here.

3.1.6 Delay

The delay in this system comes in two components. As in any store-and-forward packet network, a packet will experience some amount of delay on each hop due to the fact that it cannot begin transmission to the next station until it has been completely received. With the scheduling scheme presented here, there will be some amount of additional delay due to the need to wait for a suitable slot to arrive.

Precisely determining the delays due to this scheduling scheme is complicated. A simpler situation that can be used to model the delays due to the actual method is to consider the delays when there is no interference from other traffic and to consider transmitting packets only in the first subslot of each of the transmitting station's slots.[6] A packet will wait until one of these first subslots overlaps with a receive slot at the receiving station.

When there is very little traffic in the network, the time at which a packet will be sent (in the actual scheduling method) will not be affected by the presence of other traffic in the net, hence it will be sent as soon as a suitable subslot occurs. The suitability of a subslot will be identical to the suitability of other subslots that are part of the same slots at both the transmitter and receiver. Solo packets will be sent in the earliest of such a string of suitable subslots. A string of suitable subslots will begin at either the beginning of the transmitter's scheduling slot, or at the beginning of the receiver's scheduling slot. Hence, by assuming (in the simplified model) that packets will only be sent in the initial subslot of a transmit slot at the transmitting station, we will overestimate the delay since we are not considering all possible times when the packet might be sent. (In Chapter 5 we will see that we have not overestimated the delays observed in simulation by more than one slot time.)

The simple model is easily analyzed as a Bernoulli process. The beginning of each transmit slot corresponds to an independent Bernoulli trial with a probability of success of $p(1-p)$ (where p is the receive duty cycle). The expected number of trials to a success in a Bernoulli process is the reciprocal of the probability of success in a trial. Hence the expected number of slots until the packet can be sent is $\frac{1}{p(1-p)}$. The expected delay is minimized at $p = 0.5$ where is has a value of 4.0 (slots). For $p = 0.3$, the expected delay is 4.76 slot times.

[6] We are assuming here that the first subslot of each of the transmitting stations' slots does not straddle a slot transition at the receiving station. This assumption does not significantly damage the accuracy of this analysis. The additional delays in the actual scheduling method would be shifted by only one subslot in those cases where both of the receiver's slots (which overlap in part the transmitter's first subslot) were not listening slots.

3.2 Respecting neighbors' receive windows

The first section of this chapter has developed a method of avoiding interference from a receiving station's own transmitter. This section will extend this method to avoid all significant interference from local sources.

Interference from a nearby station's transmitter may be a problem if it is used to transmit at high power (to deliver a packet to a distant station). It would be a problem if the nearby station's transmitter delivers an interfering signal with power sufficient to significantly lower the signal-to-noise ratio of packet receptions. Whether the effect is significant or not will depend on how much processing gain the stations are using (which determines their tolerance to interference).

The power levels of signals are commonly discussed in terms of decibels, a logarithm of the power level. But here we are concerned with what the effect of an additional interfering signal (with some power level) is on the overall level of interference, which may already be quite high. The power levels add, but not the logarithms of the power levels. For example, if two signals, one at a power level of 20 dB and the other at a power level of 10 dB are added, the power level of the resulting signal is at 20.4 dB, which is a barely significant change. In order for the addition of a weak signal to increase the overall level of interference by more than 1 dB its power level must be at least one fourth the power level of the overall interference. One decibel, which is about a 25 percent change, is a reasonable threshold for significance. While we can not strictly budget the additional level of interference we may tolerate from each nearby neighbor independently (as two additional sources of interference can combine to produce an even greater level of extra interference), we can hold each such potential additional sources of interference to a maximum increase of 1 dB in total interference and budget a few decibels of additional headroom. It would then take more than four simultaneous high-power transmissions (each contributing just under the 1 dB threshold) from nearby neighbors to have more than a 3 dB effect on the overall level of interference.

Only in infrequent circumstances will a neighbor's transmission increase the level of interference by more than 1 dB. In Chapter 2 it was shown that the background level of noise may be roughly a factor of 100 greater (up by 20 dB) than the level of individual signals received from nearby stations. Stations already must cope with this level of interference. In order for an interfering station to significantly increase (by more than 1 dB) the total amount of interference, it would have to deliver (to the interfere-ee) more than 20 times (or 13 dB more) the amount of power that it is delivering to the intended recipient. (If the noise level is 20 dB over the target receive power, then the threshold of significance of one fourth, or -6 dB, of the level of the noise is 14 dB over the target receive power. Choosing 13 dB here is one decibel more conservative.) Assuming $\frac{1}{r^2}$ propagation, this threshold will be exceeded only when the receiving station is more than five times as far away as the interfere-ee (from the transmitter). For example, a station would need to be located inside the circle for a transmission from station A to station B to significantly affect the noise level:

If the most distant stations we are communicating with are at a distance of $2\rho^{-\frac{1}{2}}$, then the expected number of stations inside a circle with a radius of one-fifth this distance is only $\pi\left(\frac{2}{5}\right)^2 \approx 0.5$. This number is well under the interference threshold of 4 nearby transmitters selected above. Therefore this form of interference will not often be a problem.

When high power must be used, an additional constraint can be placed on the scheduling (to avoid interfering with a neighbor's reception). Those packet transmissions that will require high power must not be scheduled at a time that overlaps with a receive window at a neighbor who is too close. Using the notation from Chapter 1, station k is too close to station i for station i to send to station j during a receive window of station k if $h_{i,j}^2 < 20h_{i,k}^2$. The factor of 20 means that if the signal at a nearby neighbor will, at a neighbor, be more than 13 dB above the usual reception power level, then it cannot be sent during that neighbor's receive windows. While in simulations this constraint seldom came into play, we will see in the next chapter, it succeeds in effectively eliminating packet receptions with low signal-to-noise ratios.

3.3 Conclusion

This chapter has presented a pseudo-random scheduling method that can be used to ensure that packets are not received at times when the signal-to-noise ratio would be unacceptably lowered due to nearby sources of interference. The method requires neither global synchronization nor global coordination. Each station can arrange independently with its immediate neighbors to ensure that its transmissions do not mask the reception of packets either by itself or by any particularly close neighbors. The maximum total throughput (or fraction of time spent sending) out of a station is around 0.21 for a single neighbor, and can be increased as more neighbors are added. A station can even spend more than half of the time sending if it has a sufficient number of neighbors (see Figure 3.3). The expected delay per hop due to the scheduling method is a few scheduling slots which, with four subslots per slot, is around two dozen packet times (or subslot times).

The methods of this chapter and the design strategy of Chapter 2 together yield a design for an effective packet radio network that can scale to seemingly arbitrary density and requires no centralized coordination of channel use. In Chapter 2 we showed that even at high densities the network can maintain the ability to communicate between nearby neighbors by using spread-spectrum radio techniques. In this chapter we showed how to ensure that packets are sent only at times at which they will not be dropped due to collisions, but we did not require any global synchronization or coordination.

In the next two chapters we will present some simulation results that explore the performance of the network we have designed. Some further discussion of performance is presented at the end of Chapter 5 and in the last chapter.

Chapter 4

Simulation of Channel Noise Levels

This chapter will present simulations of the techniques presented in earlier chapters. Simulations will not only demonstrate the management and control techniques presented in earlier chapters, but will also allow observation of some aspects of network performance that would otherwise be difficult to discern.

The most important observable is the actual signal-to-noise ratio achieved at the receiver for each packet. The analysis of the noise levels from Chapter 2 was only suggestive of what the probable range of noise levels might be in an actual network. Since stations will communicate over distances other than exactly $\rho^{-\frac{1}{2}}$, the signal-to-noise ratios will vary from the levels predicted in Chapter 2.

The simulator, *prisim*, explicitly simulates the level of interference contributed by each transmission at each receiver. The inputs to the simulation are the geometry of stations, a model of traffic, and the transmission scheduling and control algorithms with their associated parameters. The geometry is used to determine the path gains between stations (proportional to $\frac{1}{r^2}$ where r is the distance between the stations). The path gains are used to determine routes. For now, the routes are minimum-energy routes. The results of the simulation include explicit traces of packet transmissions, traces of noise levels at receivers, and the statistics of the lowest level of the signal-to-noise ratio during the reception of each packet.

The performance of the network can be examined at two different levels: locally and globally. Locally, the throughput from one station to a neighbor (by direct communication) can be observed. This local throughput between two neighbors is a function of the bandwidth of the channel, the signal-to-noise ratio at the receiver, and the fraction of time the transmitter spends using the channel to send to the receiver. At the local level, it is important to ensure the robustness of the communication: packets that are transmitted must be received with high probability. By simulation, the range of signal-to-noise ratios can be observed and compared with the rate of growth of the noise level predicted in Chapter 2. Then an appropriate level of processing gain can be selected for the spread-spectrum radio system.

Globally, the end-to-end throughput and delay will depend not only on the local throughput, but also on congestion and queueing delays, which depend on the offered load on the network. These global issues are not unlike the issues in any other packet network. However if the proposed wireless networks were to be used independently of

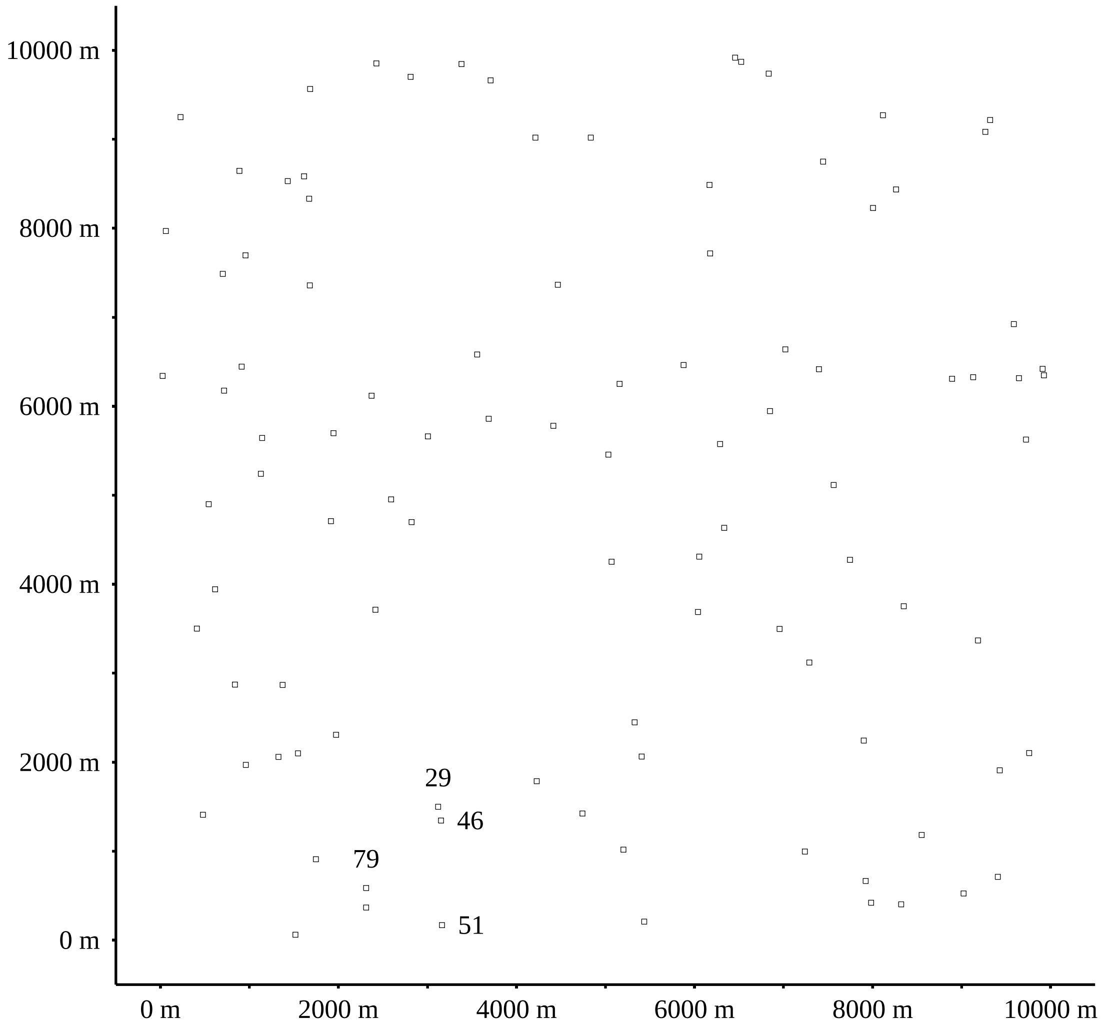

Figure 4.1: 100 stations randomly placed in a 10km by 10km square. (Four of the stations are labeled by station number as they will be referenced in the text.)

any installed infrastructure, there may be no opportunity to upgrade congested areas of the net. The next chapter will examine these issues and use the routings computed by the simulator.

4.1 Simulation to observe SNR

This section will present simulations to observe the signal-to-noise ratio at which packets are received.

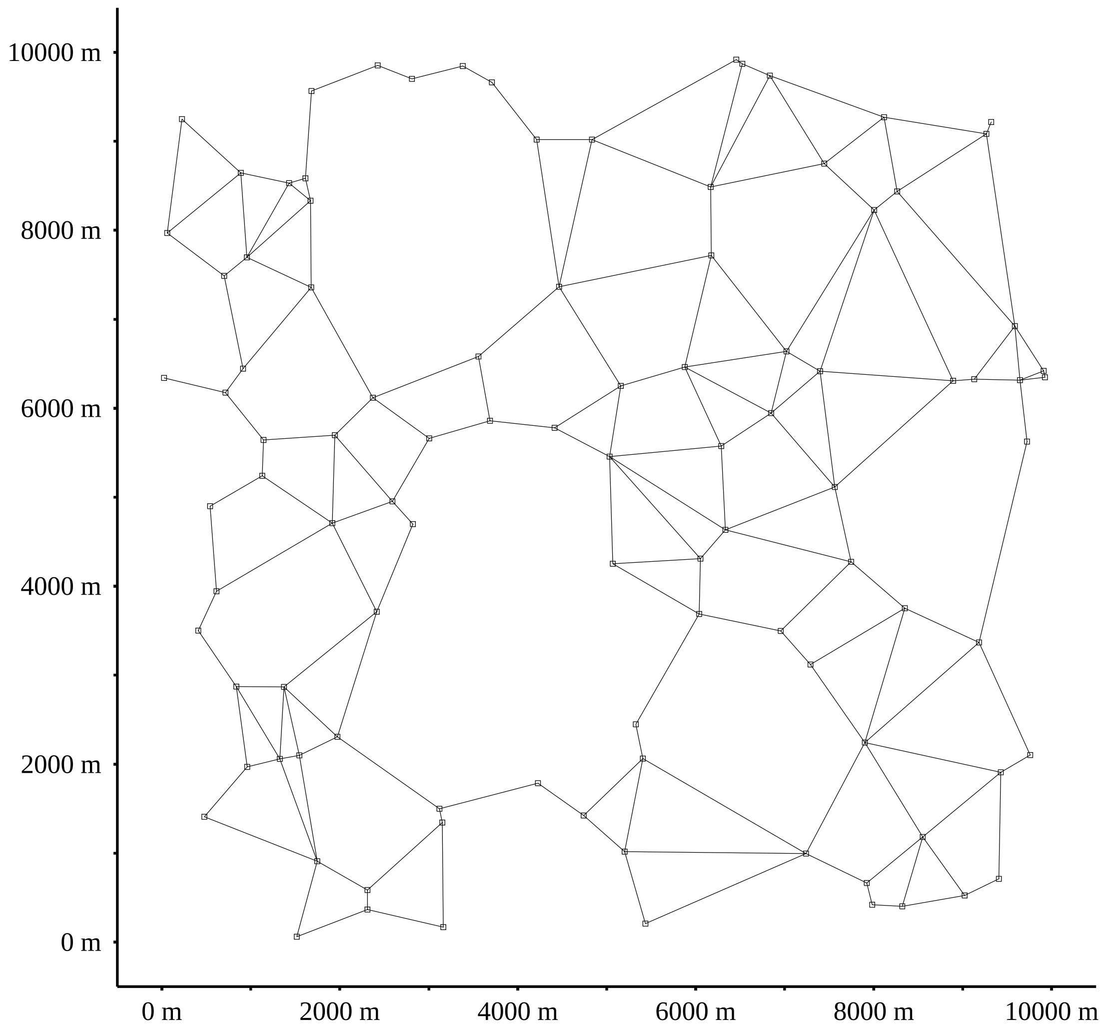

Figure 4.2: Direct hops used by minimum-energy routes.

4.1.1 Geometry and routing

The simulator places the stations in a 10km by 10km square. The locations are selected randomly by selecting two random numbers for the x and y coordinates for each station. Neither the size of the area nor the density of stations is important to our simulations. As we saw in Chapter 2 only the total number of stations and the average amount of power radiated by each station affect the signal-to-noise ratios (when there are many interfering stations and the thermal noise floor is therefore insignificant). The same number and arrangement of stations would produce the same simulation results if it were scaled to any size. Figure 4.1 shows the locations of 100 stations.

Minimum-energy routes are then computed[1] assuming path gains proportional to

[1]While the routing computation in the simulator with only 100 or 1,000 stations is straightforward, in a network of 1,000,000 stations (simulated or otherwise) the computation and storage of

$\frac{1}{r^2}$ (where r is the distance between stations). These minimum energy-routes have the property that a route from A to C that goes through B will use the same route from B to C as any other route that goes through B to get to C.[2] Hence the result of this routing computation can be stored as a routing table at each station that indicates the next hop station for each possible destination. Figure 4.2 shows all of the direct hops that are used by any route for the network of 100 stations.

4.1.2 Operation of simulator

The simulator's core operation is to explicitly record the contribution to the noise level at each receiver for each transmission. The simulator is run by an event queue. Each station calls the core simulator to begin and end each transmission. The core simulator calls the receiving station to report the reception of successful packets. Packet reception might be unsuccessful if, at some point during the reception, the signal-to-noise ratio deteriorated to an unacceptable level. For now, all packets are considered to be successful, regardless of signal-to-noise ratio, and the lowest signal-to-noise ratio during the reception of each packet is recorded for later observation.

Each station's clock is set to a random time and each station schedules its transmissions to fit within its scheduling constraints (using the methods from Chapter 3) and the constraints of the receiving station. The pseudo random slot schedules have a receive-window duty cycle of 0.3. If the respect-neighbor constraint (which we developed at the end of Chapter 3) is enabled, then the transmissions are further constrained to not contribute more than -133 dBW to the noise at any receiver, (a factor of 20 down from the target power level).

Upon receiving a packet, each station selects the next hop station from the routing table and then books the packet into an unreserved subslot at some time in the future. The subslot is then marked reserved to prevent other bookings into the same subslot. The core of the simulator is then called to schedule the event to actually transmit the packet.

4.1.3 Other parameters for simulation

The calibration parameters for the simulator were chosen to match the 26 MHz-wide license-free band at 915 MHz.[3] The selection of frequency determines a constant factor in the determination of all of the path gains (the path gains have a quadratic dependency on the wavelength) but since all propagation is affected equally by this parameter, the choice of band does not affect the simulation results. The thermal noise floor has a linear dependence on the bandwidth, but it is not important either since the level of interference from other stations will make thermal noise insignificant. The regulatory limit on power in this band is 1 Watt and was used as the maximum

the routing tables would be difficult. Solving this problem remains beyond the scope of this work.

[2]If this were not true, then the route from B to C would be a lower energy route than the B to C suffix of the route from A to C. But then the route from A to C could be improved by replacing its suffix with the route from B to C.

[3]The authorization is in 47 CFR 15.247.

transmitter power, but with the parameters for these simulations it was never a factor in limiting station power when power control was enabled. The antennas are assumed to be ideal isotropic (sometimes called “zero gain” or 0 dBi) antennas. The target power for power control was chosen to be −120 dBW. This target was chosen to be as low as possible while still maintaining a comfortable 30 dB margin above the thermal noise floor (which is at around −150 dBW for a 26 MHz-wide channel).

The simulation results and conclusions are largely independent of all of these parameter choices. Alternative values would just shift all the signal levels or thermal noise levels by a few dB. The amount of self-interference (which dominates the performance of this system) would not change since the transmitted and received signal levels will change proportionately together.

The power level delivered to a receiver from the station's own transmitter was chosen to be 40 dB down from the transmitted power. Forty decibels of isolation (at one frequency) would be remarkable performance from an antenna system incorporating a duplexer or circulator and a more realistic choice might be in the range of 10 dB to 20 dB. However, `double`'s in the C programming language used to construct the simulator provide mantissas that can support only around 150 dB of dynamic range.[4] With the thermal noise floor at around −150 dBW (at 26 MHz bandwidth) and transmitter output powers as high as 0 dBW, any less duplexer isolation would result in problems due to roundoff error in the simulation. Nevertheless, with 40 dB of transmitter-receiver isolation, the power received from a station's own transmitter will still be high enough to ensure obliteration of any other received signal. In any case, the scheduling algorithms assure that no reception will be attempted during a transmission, so the simulation results (the signal-to-noise ratios at which packets are received) are unaffected by this choice.

A slot rate of 64 per second and a subslot (or packet) rate of 1024 per second (or 16 subslots per slot) was chosen for simulation. Assuming that 20 dB of spread-spectrum processing gain is needed, and within the constraints of the three license-free spread-spectrum bands in the U.S., the system with this set of parameters could provide for roughly 30 to 120 bytes per packet (depending on which band was used). In the future with anticipated advances in electronics, a system operating in a bandwidth of one or two GHz (probably operating at a few tens of GHz) would be able to support 2000-byte packets at a packet rate of 1024 per second.

The hash function used for scheduling was:

$$\text{hash}(x) = x^{17} \bmod 2^{31} - 1$$

Where x is the top 38 bits of a 64-bit clock kept at each station. The relative phases of the clocks are randomly initialized. The result of the hash function, an essentially uniformly-distributed 31-bit number, was then compared with $2^{31}p$ where p is the receive duty cycle. If the hash value is less, then it is a receive slot. The hash function was empirically tested for low self-correlation and it appeared to be good.[5]

[4] IEEE 64-bit floating point numbers have 52-bit mantissas. A 52-bit mantissa provides a dynamic range of 2^{52}; $\log_{10} 2^{52} = 15.6$, hence 156 dB.

[5] The form of this hash function was suggested to me by Derek Atkins.

4.1.4 Traffic to load each station's transmitter

The worst-case signal-to-noise ratios will determine the level of processing gain required to make the system operational. Since transmissions will depend on the traffic, observing worst case signal-to-noise ratios requires selecting a traffic load designed to maximize the amount of time spent transmitting. This maximum would occur when there is boundless traffic demand on each direct hop used by routes. We can simulate this by generating and enqueueing (or booking) one packet at each station for each direct-hop neighbor. When the packet transmission is actually begun, a new packet for the same neighbor is generated, enqueued, and scheduled. This scheme for packet generation ensures that every transmitter is operating at all times at which it might transmit, and provides that at times when there are more than one possible receiver, the transmissions will rotate evenly among the possible receivers.

4.1.5 Observed signal-to-noise ratio

Figure 4.3 illustrates the operation of the simulator. With the simulator operating with the 100 station geometry and routing of Figures 4.1 and 4.2, this figure shows the received power at station 29 as a function of time. Power control was enabled, so the power of all transmissions are controlled to deliver -120 dBW (or one picowatt) to the intended receiver. The high peaks near -65 dBW in Figure 4.3 correspond to transmissions to the neighbors numbered 14 and 66. Since in this simulator the received power level of a station's own transmitter is 40 dB down from the transmitted power level, the output power levels of these transmissions was around -22 dBW (or about 6 milliwatts). The peaks at -83 dBW correspond to transmissions to a nearby neighbor, station number 46. These transmissions occur at a lower power level (in this case, approximately 50 microwatts) since the path gain is higher to a nearby station.

The peaks at -103 dBW are not from the station's own transmitter, but are from a nearby neighbor transmitting at high power. In these cases station 46 is transmitting to either station 79 or station 51.

The packet receptions are illustrated on the received power plot by the small segments with arrowheads drawn at the received power level. In this case (with power control) all receptions occur at a received power level of -120 dBW. Note that some receptions overlap in time.

Without power control, the power at which packets are received varies. This variation is illustrated in Figure 4.4. Here, all transmissions occur at .02 Watt (-17 dBW) output power (which in the local receiver is observed as -57 dBW). Notice that without power control, the packets arrive at station 29 from the nearby neighbor (station 46) at an unnecessarily high power level.

From these plots we can quickly estimate what the signal-to-noise ratio is. The distance from the reception line to the received power level is the ratio (in dB) of the signal level to the signal plus noise level. When the signal level is significantly below the noise level (by a factor of ten or more), this ratio is essentially equal to the signal-to-noise ratio.

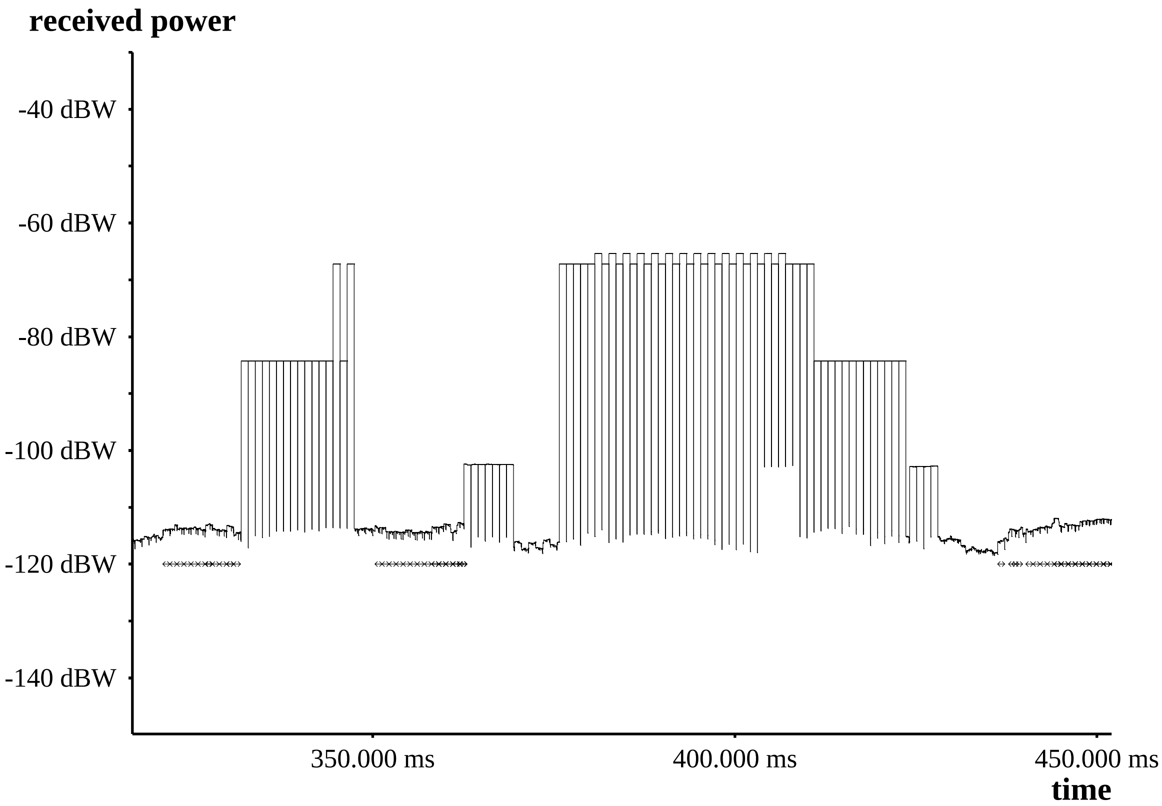

Figure 4.3: Example of received power as a function of time at station number 29 in the 100-station network. The lines with tiny arrowheads below the plotted receive power level are the receptions plotted at the time and power level at which they occur. Since power control is being used, all receptions occur at the same power level, -120 dBW. The higher levels of received power are caused by operation of the station's own transmitter. The variation in the transmitter's power level can be seen. (The receiver sees the transmitted power level down 40 dB.) A few of the high received power levels (those near -100 dBW) are caused by a transmitter at a nearby station (number 46 on the map in Figure 4.1).

For each packet reception the simulator records the highest received power level that occurred during the reception. This peak power level is used to determine the lowest signal-to-noise ratio experienced during the reception of each packet. Figure 4.5 shows distributions of these signal-to-noise ratios from two simulations. The first simulation, shown on top, is without power control. The second simulation, shown on the bottom, is with power control. Both of these distributions are centered around approximately -7 dB. This value roughly matches the signal-to-noise ratio predicted for a network of 100 stations by the model of Chapter 2 and Figure 2.1, if we assume a transmit duty cycle in the range, of one half to one fifth. Most actual transmit duty cycle values were indeed in this range as shown in the top of Figure 4.6. The transmit duty cycle is not affected by power control so the one distribution at the top of Figure 4.6 applies to both cases in Figure 4.5.

Successful reception will require that the signal-to-noise ratio during reception

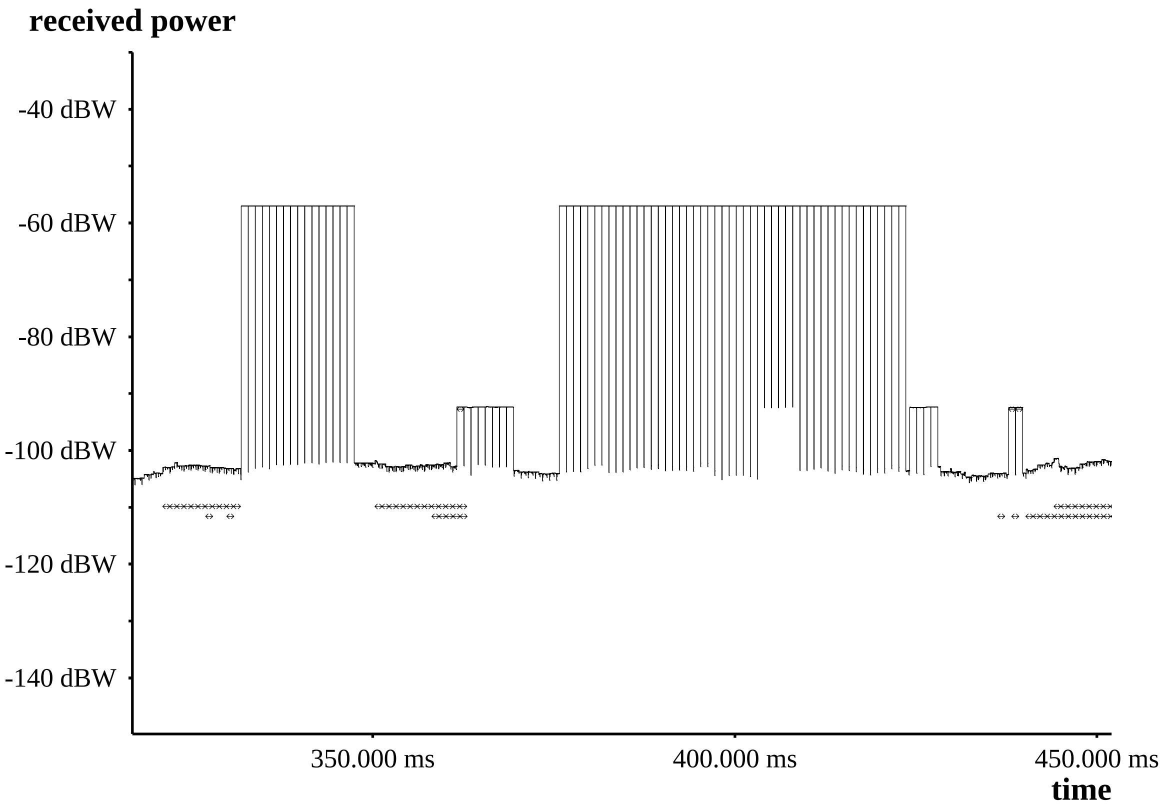

Figure 4.4: Simulation of received power at station number 29, with power control disabled. All transmissions are at -17 dBW, or 0.02 Watts, output power. As before, the lines with tiny arrowheads plotted below the received power level represent the receptions at this station and are plotted at the time and power level at which they are received. The variation in received power is a direct result of disabling the power control as it now depends on the distance (or path gain) between the stations. (Both the background noise level and the level of individual receptions are around 10 dB higher here than with power control, but that is the result of the particular fixed power level chosen for the transmissions. This 10 dB shift is not important. What is important is the signal-to-noise ratios, some of which are improved by use of power control, but the effect is too subtle to be seen by comparing this figure with the previous figure. We will be able to see the difference in Figure 4.5.)

remain above some threshold. A reasonable expectation for the performance of the the spread-spectrum radio detectors is that reception will be successful when the received signal-to-noise ratio (in dB) plus the processing gain (in dB) remains above 5 dB.[6] With a processing gain in the range of 20 dB to 25 dB, the threshold would be somewhere between -15 dB and -20 dB. The vast majority of packets will be received successfully in either of these cases.

Power control is clearly an improvement. It reduced the number of packets received both at excessively high power and at undetectable (low) signal-to-noise ratios, without changing the center of the distribution. This change is an improvement over

[6] [SOSL94], Chapter 5, page 1109

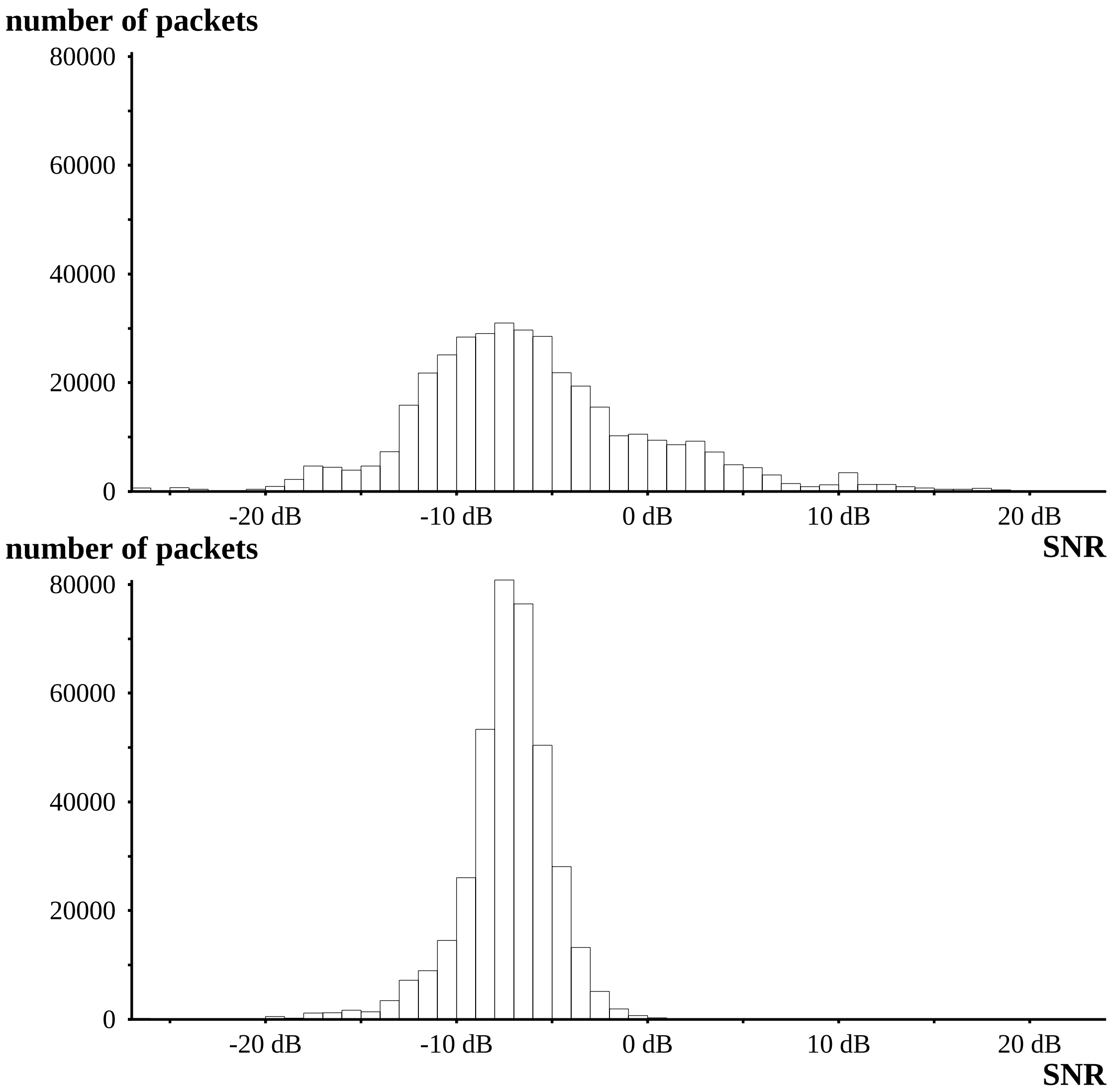

Figure 4.5: Distribution of worst signal-to-noise ratio during reception of each packet. Top plot is without power control and bottom plot is with power control.

not doing power control because it moves packet receptions out of the left tail of the distribution (where packets are below the threshold and will be dropped) by reducing the amount of interference from those transmissions that previously formed the right half of the tail of the distribution.

Even with power control, some packets remain below the threshold for successful reception. These low signal-to-noise ratios are caused by interference from high-power transmissions occurring near (in space) to receptions that overlap in time. These few remaining cases of unsuccessful reception can be eliminated by turning on the respect-neighbor constraint. Figures 4.7 and 4.8 show that turning on the respect-neighbor constraint eliminates the tail.

The respect-neighbor constraint is actually more effective at tightly cutting off the tail of the distribution than has been demonstrated here. The threshold for the respect-neighbor constraint (13 dB) was chosen with the assumption that the

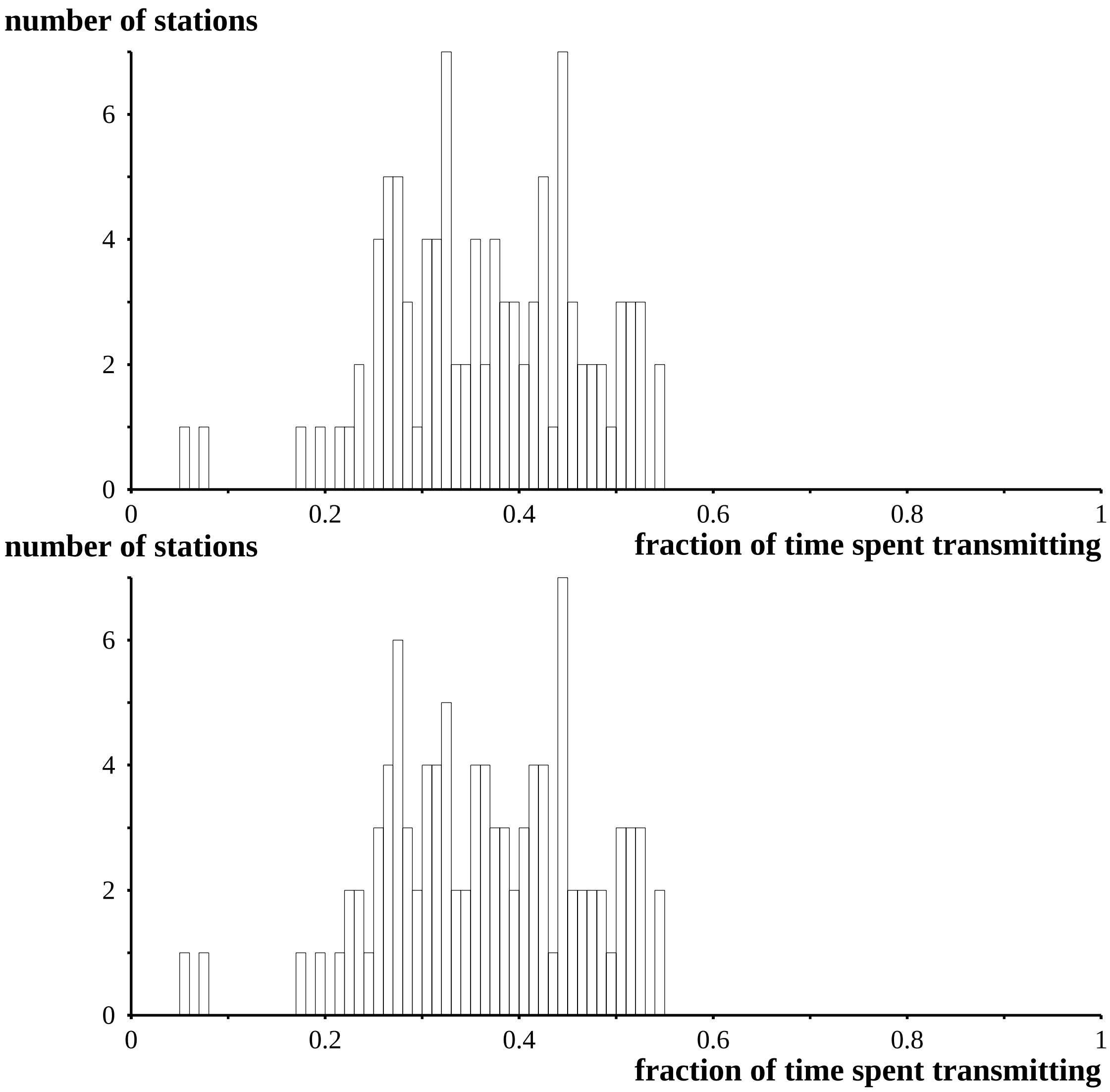

Figure 4.6: Distributions of the transmit duty cycle. The top is without the respect-neighbor constraint, and the bottom is with the respect-neighbor constraint. There are no significant differences between the transmit duty cycles with and without the respect-neighbor constraint.

background level of interference was already 20 dB above the received signal levels. But the overall level of interference here is on average only 10 dB above the target reception power level (because this is a small network). Hence a nearby neighbor, within the 13 dB constraint imposed by the respect-neighbor constraint, may in fact decrease the signal-to-noise ratio down to as much as -15 dB. As the network scales, the overall noise level increases, reducing the impact of those high-power transmissions allowed under the respect-neighbor constraint. Thus, the cutoff will become tighter as the number of stations increases.

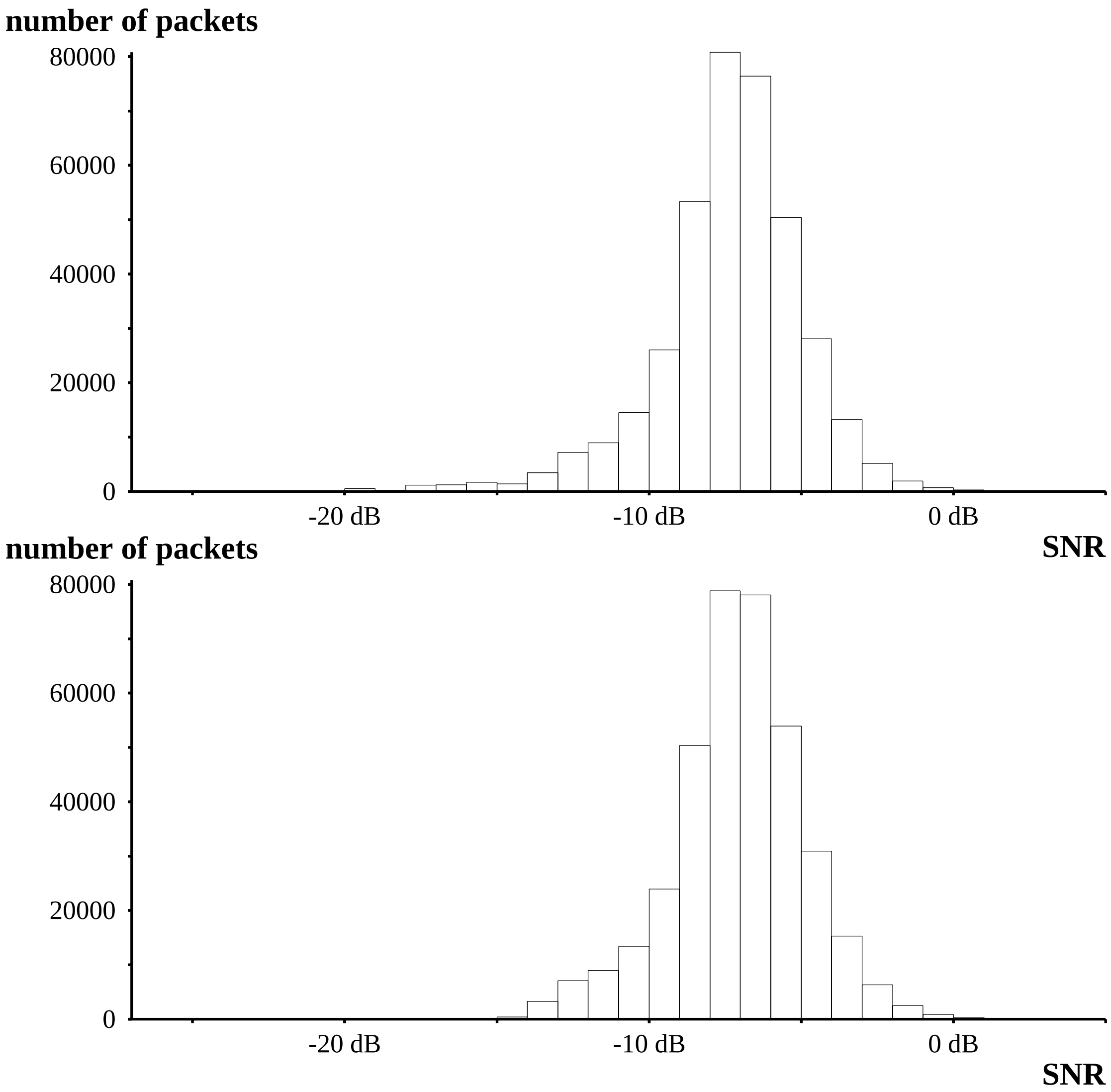

Figure 4.7: Impact of the respect-neighbor constraint on signal-to-noise ratio. Top distribution is the same as the bottom of Figure 4.5 (though the scale has changed) which is from a simulation with power control but without the respect-neighbor constraint. The bottom is with power control and with the respect-neighbor constraint. The respect-neighbor constraint has reduced to zero the number of packets received with low signal-to-noise ratio (this reduction can be seen more clearly in Figure 4.8 where the vertical scale is magnified), but has otherwise had little impact.

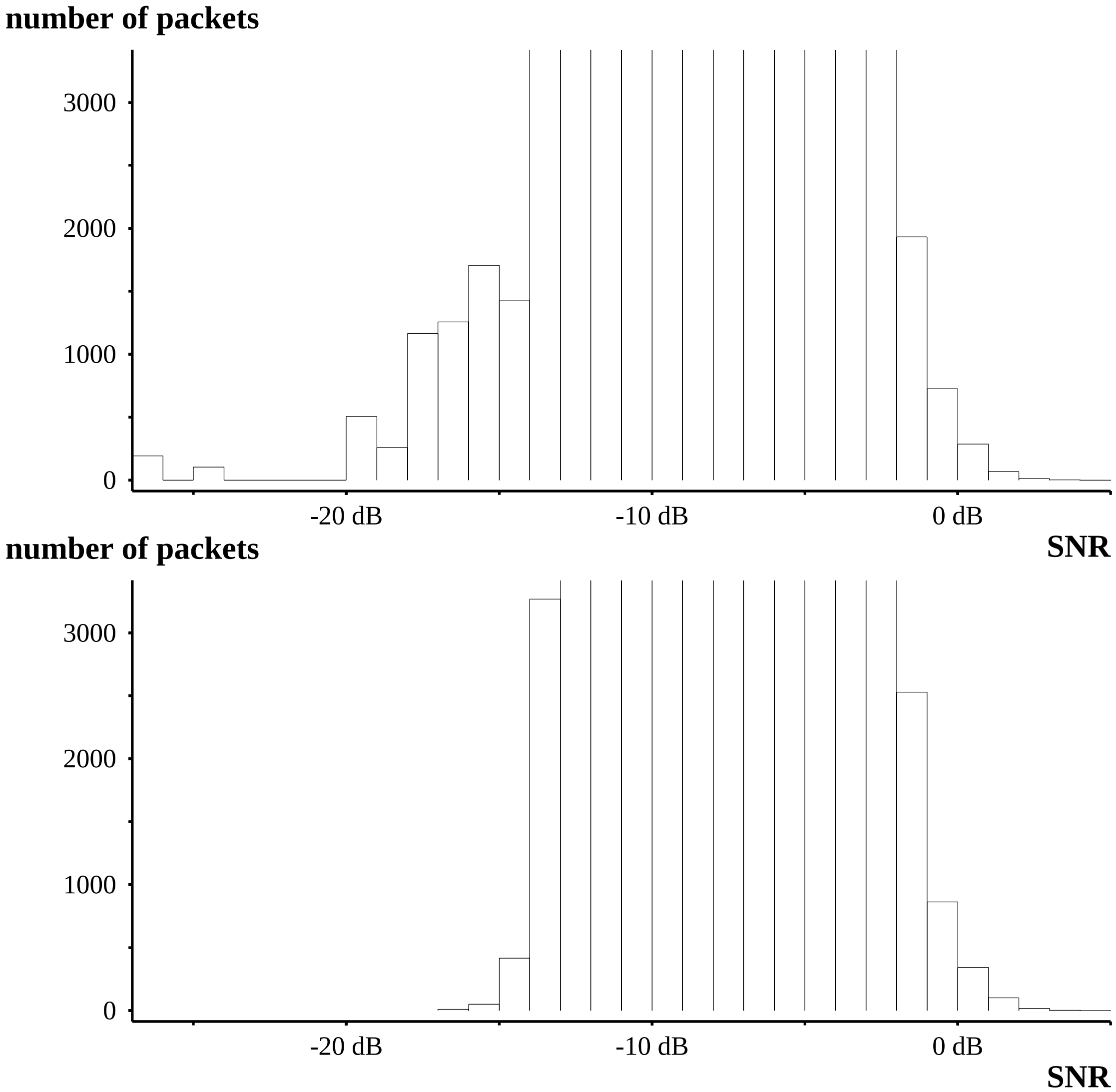

Figure 4.8: Same as Figure 4.7 but at a magnified vertical scale so that the tails can be seen more clearly. In the bottom distribution (for the case with the respect-neighbor constraint) no packets were received with a signal-to-noise ratio of less than -17 dB.

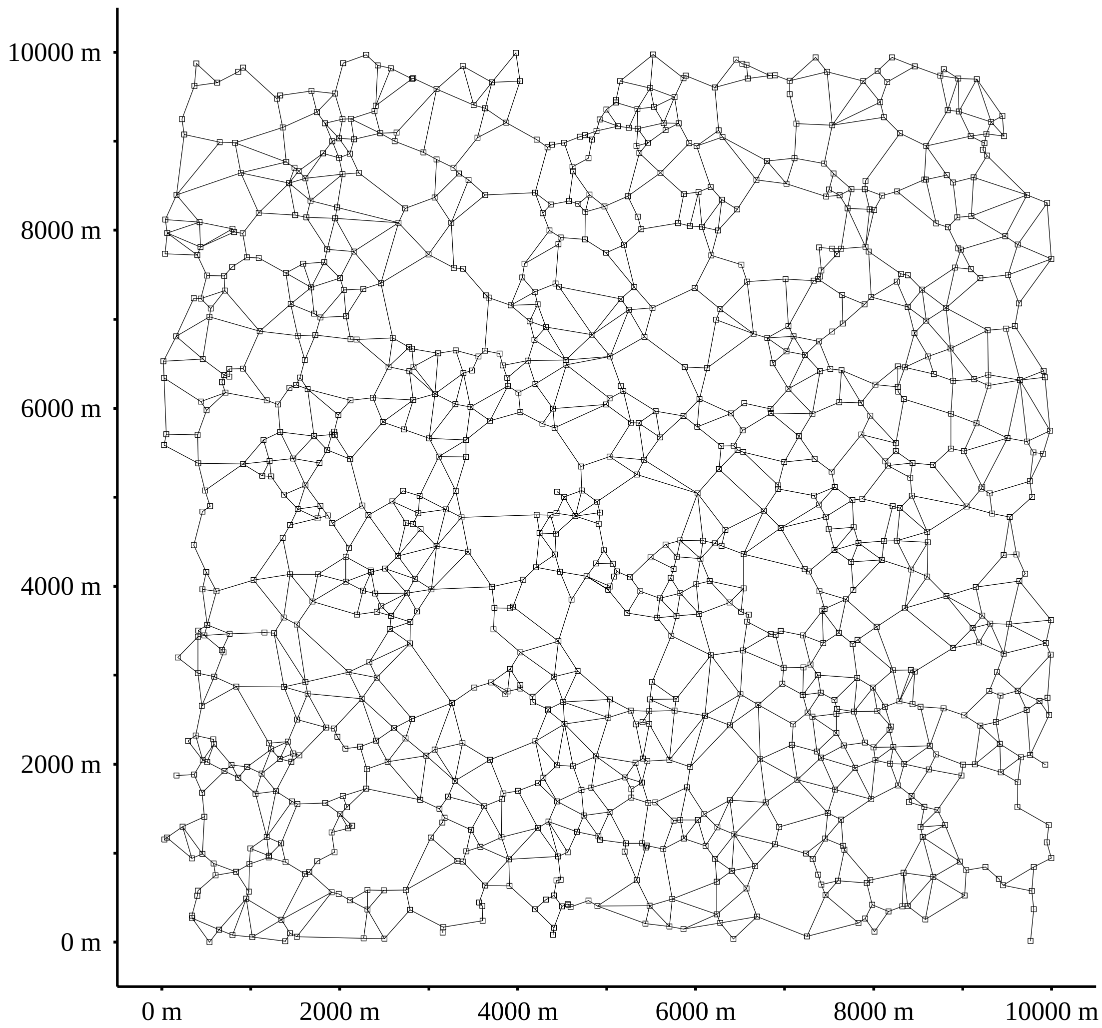

Figure 4.9: A network of 1,000 stations. Links shown are the direct hops used by minimum-energy routes.

4.2 Reduction in signal-to-noise ratio as network scales

As the network scales in number of stations (or in density within a region) the signal-to-noise ratios are expected to grow as predicted in the model of Chapter 2 and Figure 2.1. While direct simulation of the interference in a network with millions of nodes is not possible, the simulator can be used to investigate the growth of noise with a modest amount of scaling. This section presents the results of scaling the network to 1,000 nodes.

As earlier in the chapter, the stations (this time one thousand in number) were placed randomly in a 10km by 10km square. Minimum-energy routes were computed between all pairs of stations in the network. The direct hops used by these routes are

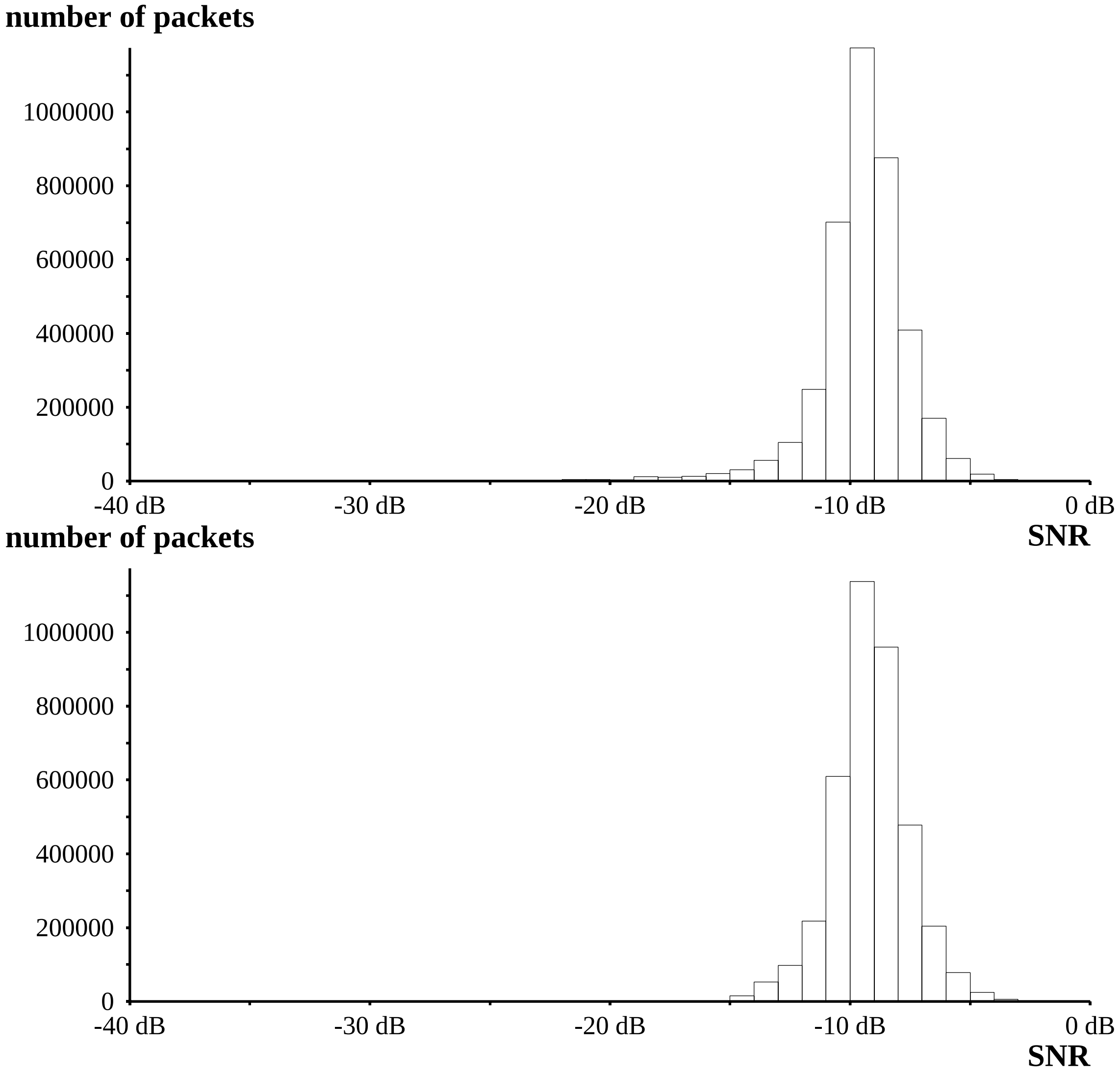

Figure 4.10: Signal-to-noise ratios for a network of 1,000 stations. The top plot is without, and bottom plot is with, the respect-neighbor constraint.

shown in Figure 4.2. Again, the network was loaded heavily by generating traffic to cross each direct hop that is used by any route.

The distributions of signal-to-noise ratios are presented in Figures 4.10 and 4.11. With 1,000 stations, the distributions are centered slightly lower, near -9 dB. This value corresponds closely to the shift expected from the model in Chapter 2, and can be read directly off of the eta=0.5 curve line of Figure 2.1.

Without the respect-neighbor constraint, the tail is much longer this time, as can be seen in Figure 4.11. This lengthening of the tail is caused by a few stations that happened to have been placed extremely close together (in some cases only 5 meters apart). The respect-neighbor constraint again eliminates the interference that leads to the low signal-to-noise ratio tail of the distribution.

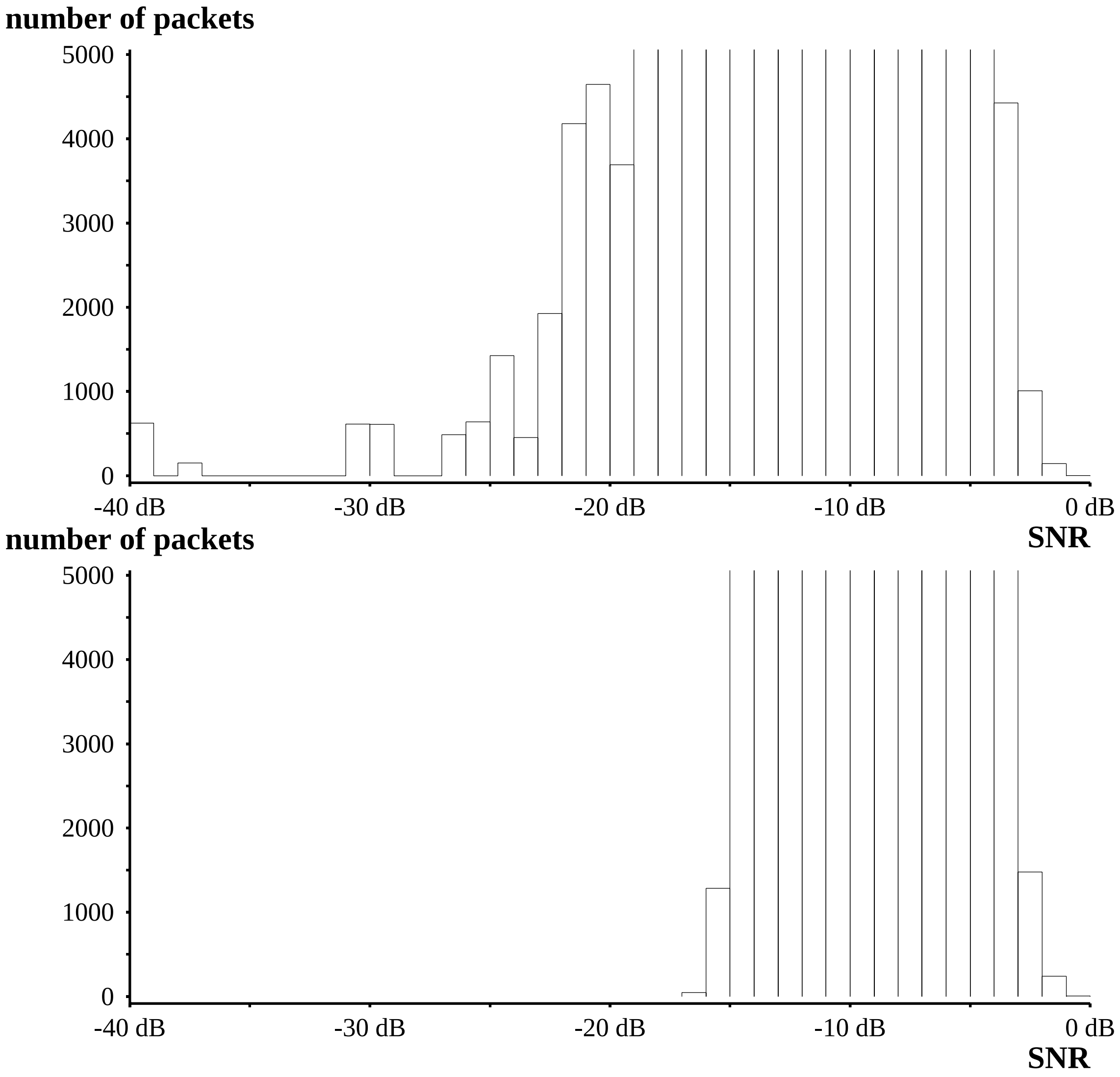

Figure 4.11: Same as Figure 4.5, but at a magnified vertical scale so that the tails can be seen more clearly.

4.3 Summary

The simulations presented in this chapter have verified the model for noise-level growth from Chapter 2 and have exercised the scheduling methods from Chapter 3. In networks of 100 and 1,000 stations randomly placed in a square area, both minimum-energy routing and the respect-neighbor constraint have been demonstrated to allow all packets to be received at a reasonable (and useful) signal-to-noise ratio. If stations with a moderate amount of spread-spectrum processing gain are employed, then we can be reasonably certain that transmitted packets will be received successfully. The packet radio network is essentially collision-free.

Chapter 5

Multihop Routing and System Performance

Earlier chapters have shown how a system of packet radio stations can maintain the ability to communicate between neighbors while the system scales in number and density. While this is a meaningful result, an important limitation (with minimum energy routes) is that stations can directly communicate with only their neighbors. Packets traveling further will be routed over multiple hops. The sharing of resources implied by this limitation may have a significant impact on overall system performance. Accurately measuring overall system performance in light of this requires knowledge that, in general, we do not have: What are the end-to-end demands for service?

Even without knowledge of the traffic, some observations can be made about overall system performance. The routing tables produced by the minimum-energy routing can be examined to observe the amount of sharing. For example, if the traffic is uniformly distributed to and from each station, then the number of end-to-end paths that pass through a station is a measure of how much traffic is competing for resources at that station.

5.1 Congestion

Multihop traffic will be limited by contention for use of the resources at shared transit stations. Only single-transmitter stations were considered in this work, so at most one packet may be transmitted at a time by any one station.[1] The receivers will have multiple despreading channels and detectors and may receive a number of packets simultaneously. Therefore even with a receive duty cycle of only 0.3, stations may be able to receive packets at a higher rate than they can be transmitted.

For the same 1,000-station network with minimum-energy routing that was pre-

[1]Multiple transmitters per station might make sense if directional antennas were used, but without directional antennas, the additional power from multiple simultaneous transmissions throughout the network would degrade the noise level such that any gain in throughput from multiple transmitters would be lost due to the increase in noise.

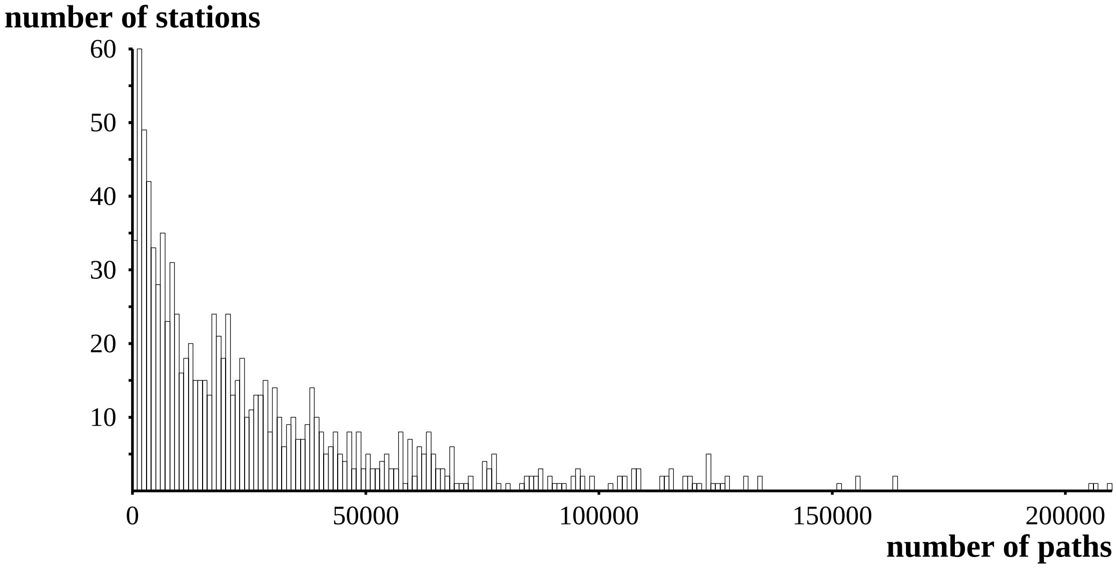

Figure 5.1: Distribution of number of paths routed through each station in the network of 1,000 stations using minimum energy routes (as shown in Figure 4.9). For each source-destination pair there is one path, for a total of 1,000,000 paths.

sented in the previous chapter, Figure 5.1 shows the distribution of the number of paths routed through each station. There are 1,000,000 total paths, one for each pair of stations. Each of these paths is the route over which the lowest amount of energy would be used to transfer the packet end-to-end. (In two dimensions this is equivalent to finding the path with the shortest sum of the squares of the distance across each hop.) The most notable part of the distribution is the group of three blips at the extreme right side. Each of these three stations have over 200,000 (or 20%) of the paths routed through them. If the distribution of traffic is at all uniformly distributed among the 1,000,000 different source-destination pairs, then at each of these few stations more than one fifth of all traffic in the network will be vying for the throughput capacity of the station.

Furthermore, if the traffic is distributed uniformly among the source-destination pairs, then there are a number of stations that would carry more than 10% of the traffic (have more than 100,000 paths through them, from the right half of the distribution of Figure 5.1). These stations are shown in Figure 5.2. The stations identified by this criterion are either part of a busy path (that is used to get from one large region to another), or are located at the crossroads of two busy paths (that alone were not busy enough to cross the threshold, but at the crossroads station the combined number of transit paths was sufficient).

Communication across the single hops chosen by minimum energy routing was shown (in Chapter 4) to remain functional when all such single hops are fully loaded. The system could try to route around the hot spots while continuing to use only the same set of single hops that were used in the minimum energy routing (thus continuing to ensure that packets would not be lost in the noise). By distributing the traffic differently, the congestion might be alleviated. But this scheme would fundamentally change the routing computation. The route between two stations would then depend

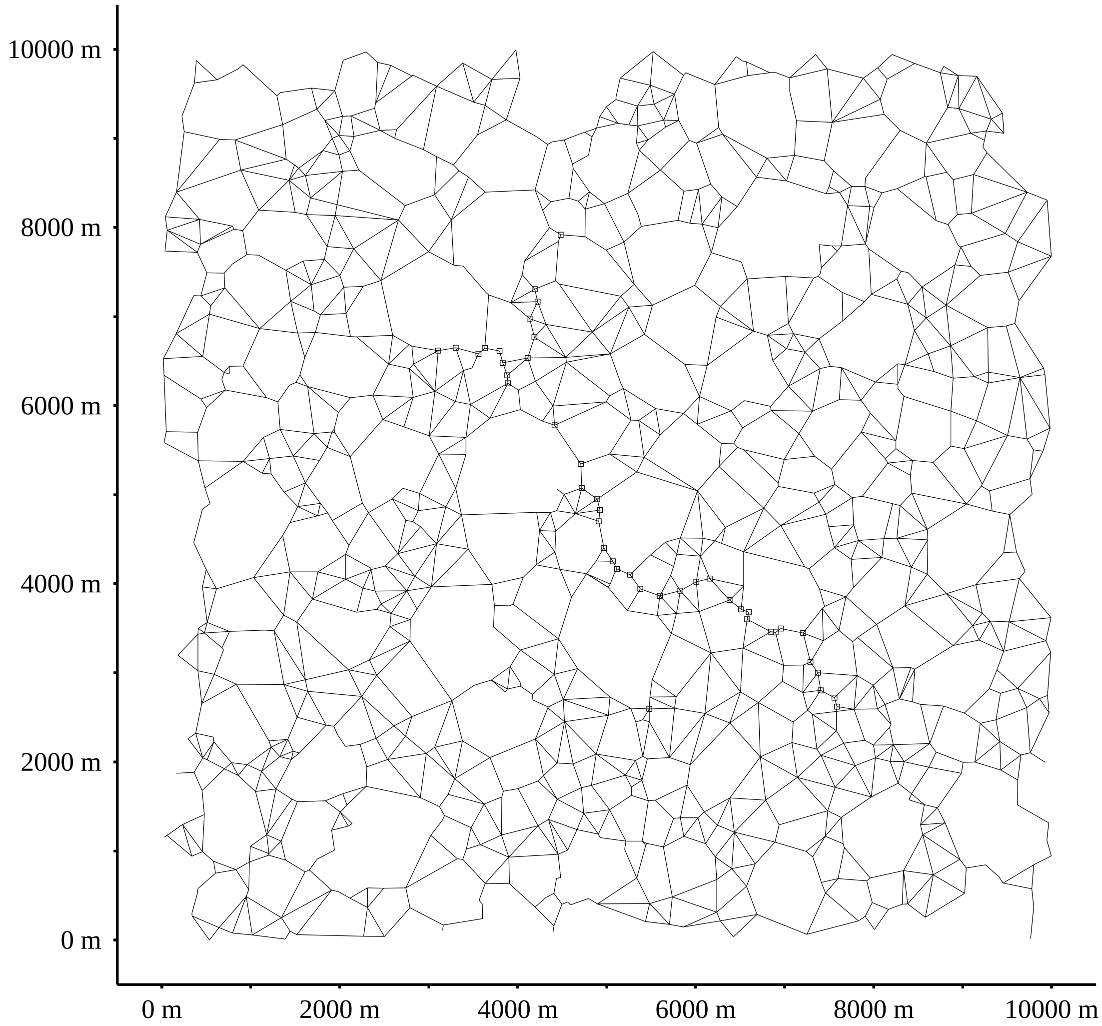

Figure 5.2: Stations with more than 10% of the end-to-end paths routed through them. Shown here is the network of Figure 4.9 with a square drawn only at the stations carrying more than 100,000 paths.

not only on the geometry of candidate transit stations but on the routes of (and hence the geometry of) all other stations in the entire network. Even so, there is no guarantee that a sufficient number of alternative uncongested routes will exist. Figure 5.3 identifies all stations with more than 1% of the end-to-end paths routed through them. This figure shows almost all alternative paths are already blocked by a station with more than 1% of paths routed through it, showing that a factor of 10 is an upper bound on the reduction in sharing we could expect if we were to attempt to redistribute routes away from the stations carrying more than 10% of the paths.

Another problem with using minimum energy routes for traffic that transits over a large distance is the accumulated delay introduced by the many hops. The top distribution in Figure 5.5 shows that many end-to-end paths in the 1,000-station network are over 30 hops long and some are as many as 70.

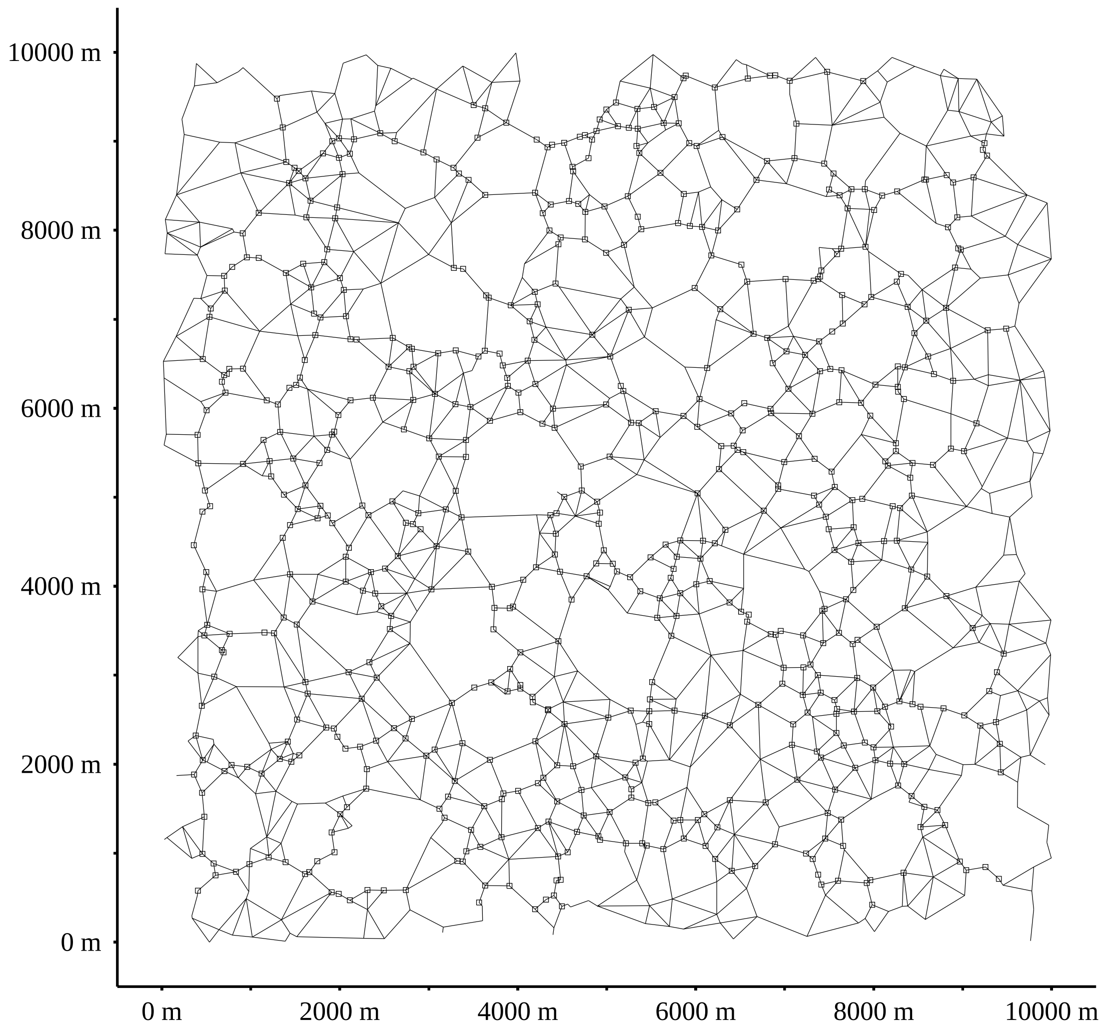

Figure 5.3: Stations with more than 1% of the end-to-end paths routed through them.

It seems that routes other than minimum energy routes might be able to improve performance if they can reduce the number of hops and amount of sharing. A simple modification of minimum energy routing that attempts to improve both of these situations was explored. The modification consisted of adding a per-hop constant term to the cost function so that the cost of a route is now a linear combination of the energy of the route and the number of hops on the route. By preferring paths with fewer hops, this modification will, one hopes, reduce the amount of sharing, but do so in a way that does not require that the computation of routes take explicit note of the amount of sharing.

The relative weightings in the linear combination should be chosen to have a moderate effect on the routes. Weighting the number of hops too light would result in no significant differences from the minimum energy routing. Weighting the number of hops too heavy might result in too many transmissions at high power (to cover the longer single-hops), which would increase the amount of interference. Increased

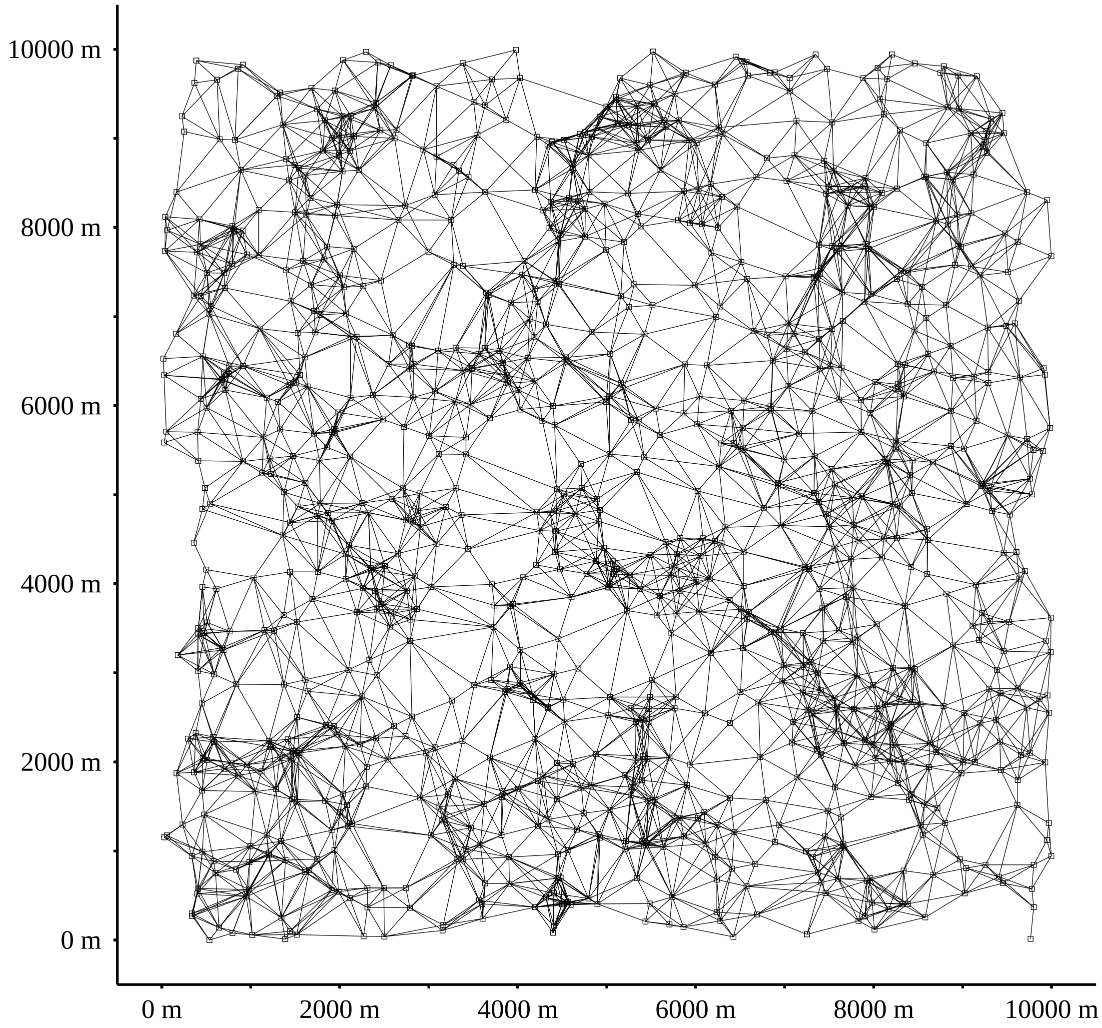

Figure 5.4: Map of the 1,000-station network showing single hops used by routes produced by a modification to minimum-energy routing. Here the cost of a route is a linear combination of the energy and the number of hops. The weightings were ad hoc and chosen to have a moderate effect on the routes.

interference would reduce the achievable raw throughput of the spread-spectrum radios because by decreasing the signal-to-noise ratio it would increase the amount of processing gain needed. Increasing the processing gain while holding the bandwidth constant reduces the achievable data rate.

Figure 5.4 shows the single hops that are used in the resulting routings for one particular relative weighting of the number of hops. The weighting was chosen so that the number of hops on each path was on average reduced by roughly one half. The distributions of path lengths in both the original minimum-energy routing and the modified routing are shown in Figure 5.5. Figure 5.6 shows the resulting distribution of the number of paths routed through each station. These plots show that the path lengths have indeed been reduced by a factor of two, and that the amount of sharing

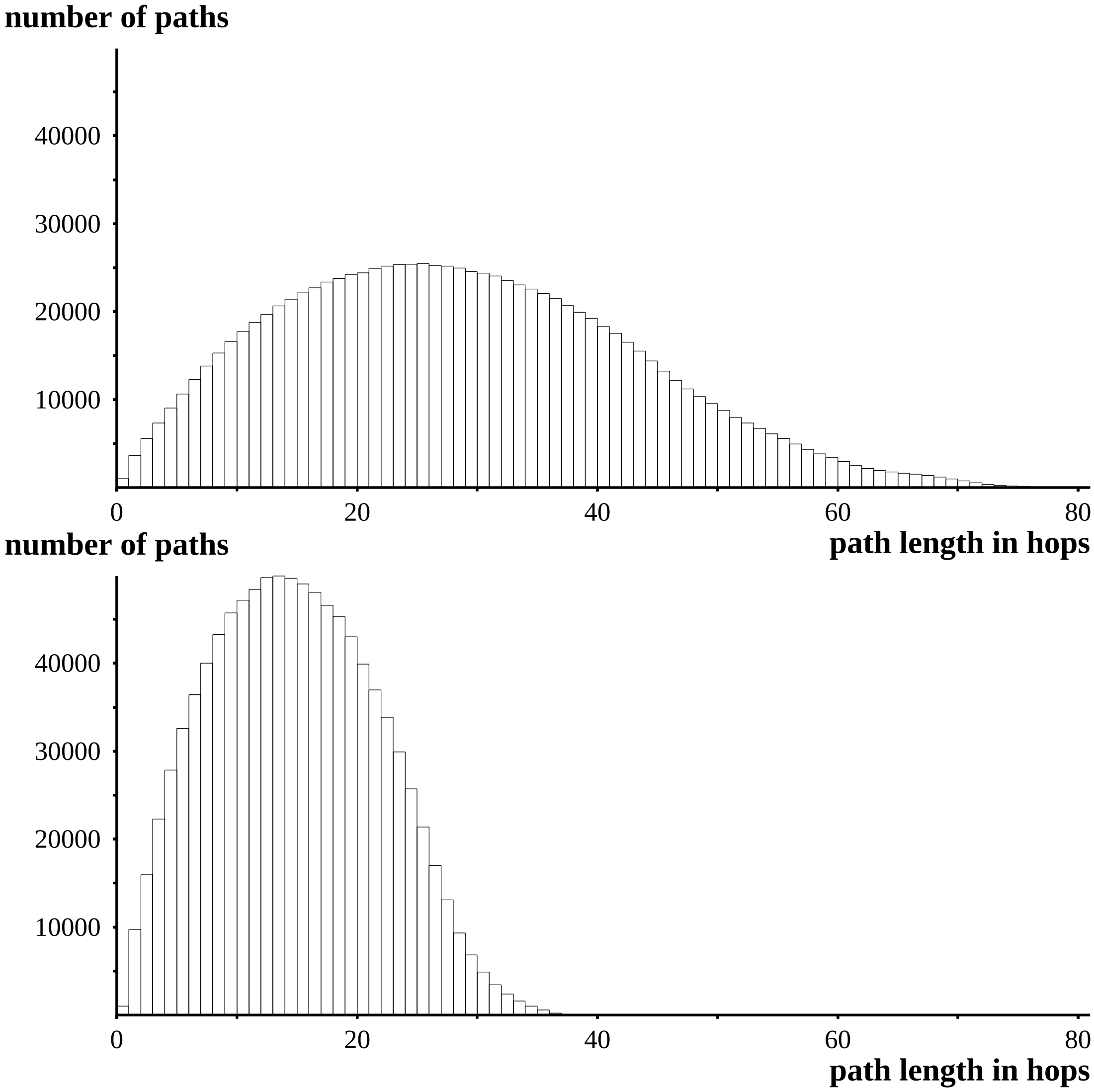

Figure 5.5: Distribution of path lengths of the 1,000,000 different paths in the network of 1,000 stations. The top plot is for minimum energy routes. The bottom plot is for the paths produced by the modified routing computation.

has dropped by roughly a factor of two as well (including the worst-case blips on the right). These results may appear to be an improvement in performance. The number of packets (of a fixed length in time) that can be sent between stations has been doubled, but whether or not we have increased throughput depends on the impact of this departure from minimum energy routes on the signal-to-noise ratios.

Figure 5.7 shows the change in signal-to-noise ratio due to the modified routing. The modified routing has reduced the signal-to-noise ratios by a factor of 4 (or 6 dB). A decrease of approximately this amount is understandable: The packets are on average being sent twice as far in each single hop, so the power control algorithm is on average increasing the power of each transmission by a factor of four (to cover twice the distance with $\frac{1}{r^2}$ propagation). The received power levels of each packet at the intended receiver will be the same, but the interference has quadrupled in power.

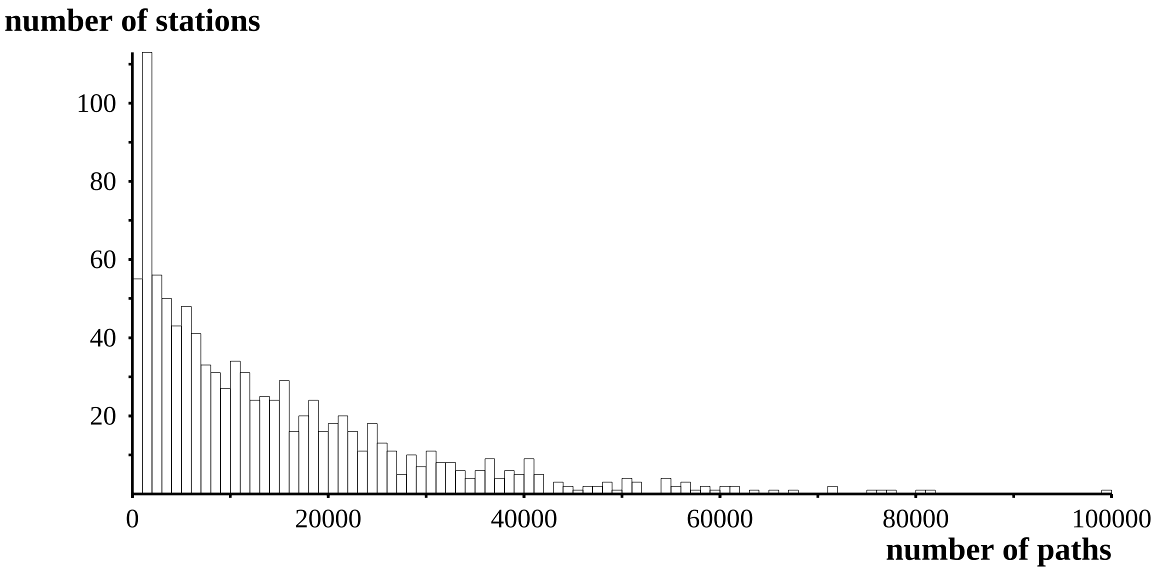

Figure 5.6: Distribution of number of paths routed through each station in the network of 1,000 stations using routes produced by the modified routing computation.

Hence the signal-to-noise ratio has been reduced by a factor of four, and would be reduced by a factor of four in a system of any scale. In the already interference-limited channel the modified routing will result in a reduction, by roughly a factor of four, in raw data rate achievable.

The factor of two gain in throughput due to reduced sharing from the modified routing algorithm has been more than lost due to a resulting factor of four reduction in the raw per-hop data rate. The result is a net loss of a factor of two in system performance. Unfortunately, it appears that any system-wide deviation from minimum energy routing that is able to make a significant reduction in the number of hops taken, or a significant reduction in the sharing, will result in a net loss in system performance.

Minimum energy routing is important to system performance. While it remains unknown if improvements might be possible on minimum energy routing, any improvements are unlikely to change the routes much. In the experimental attempt to reduce the delay and the amount of sharing in the routes, the number of hops was reduced. Any routing that reduces the end-to-end delay (which is dominated by the per-hop scheduling delays) must reduce the number of hops taken. Reducing the number of hops requires traversing longer distances with each hop, resulting in a net penalty in system throughput due to increased noise, more than offsetting any gains in reduced sharing from taking longer hops.

More importantly, the results of this section illustrate that if the traffic demand on the network has no locality, then packet radio networks cannot scale to large numbers of stations while maintaining constant per-station throughput performance. Even a moderate size network of only 1,000 stations was shown to have serious congestion problems. As any packet radio network scales in number of stations, the size of at least some cross section will grow no more than as the square root of the number number of stations. If there is no locality to the traffic, the demand for service that

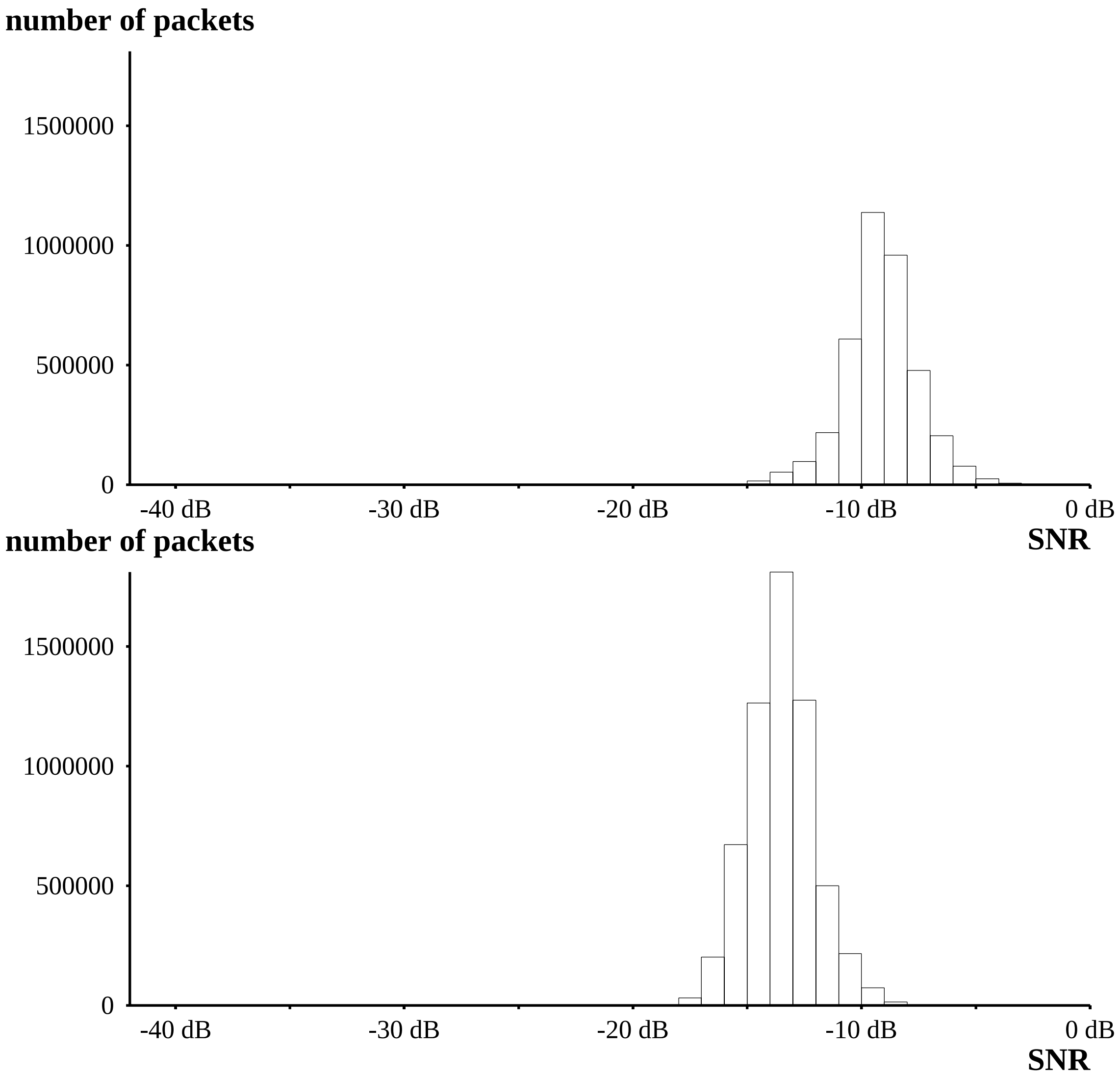

Figure 5.7: Signal-to-noise ratios in a network of 1,000 stations. The top plot is with minimum energy routes, the bottom plot is for the modified routing. The modified routing has reduced the signal-to-noise ratios by 6 dB, or a factor of four.

traverses this cross section will grow linearly in the number of stations. Thus as the system grows, the demand to carry traffic through this cross section will grow faster than the capacity across the cross section grows.

5.2 Multihop delay measurements

End-to-end delay is an important part of performance in a communication system. This section presents a few measurements of delay from simulations of lightly-loaded networks that use minimum-energy routes to measure the delay in the absence of congestion.

The top plot in Figure 5.8 shows the distribution of delays in the network of 100

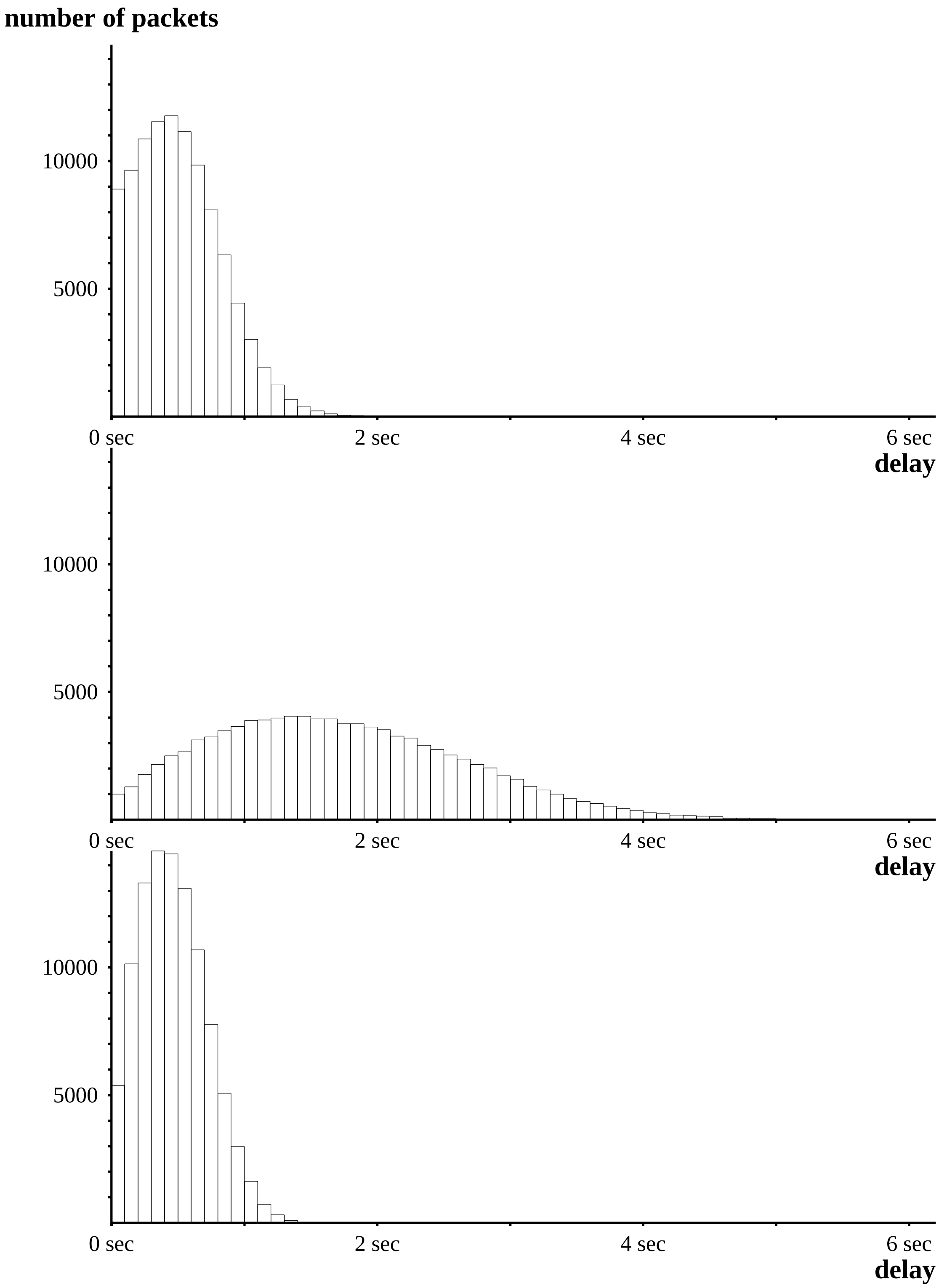

Figure 5.8: Distributions of measured end-to-end delays under extremely light load in simulated network. The top is for the network of 100 stations, the middle and bottom are for 1,000 stations. Both of these simulations used a slot size of $\frac{1}{64}$ seconds with packet sizes of $\frac{1}{1024}$ seconds. In the bottom plot the slot size was reduced to $\frac{1}{256}$ seconds, reducing the average wait for a suitable scheduling slot by a factor of four.

stations using minimum energy routing, with a very light traffic load (one packet generated network-wide an average of once every ten seconds), a slot rate of 64 per second, and 16 subslots per slot. Each generated packet is assigned a random source and destination and then sent through the network from the source to the destination, and the end-to-end delay is recorded. The middle plot in Figure 5.8 shows the distribution of delays for the same experiment, but in the 1,000-station network with minimum energy routing.

Ignoring congestion, there are two components that contribute to the delay of multi-hop traffic. One is the number of forwardings (the path length). The other is a delay at each station due to waiting for a suitable slot to arrive. As the size of the network was increased from 100 to 1,000 stations (by a factor of 10), the distance across the network increased by approximately $\sqrt{10}$, which explains the factor of 3 increase in delay evident in the second distribution of Figure 5.8.

In the bottom distribution in Figure 5.8, the slot rate was increased to 256 slots per second, and the number of subslots per slot reduced to 4 (maintaining 1024 subslots per second). Delay is reduced by a factor of four because the expected time to a suitable slot has been reduced by a factor of four. While delay has been improved, possibly at the expense of reduced throughput (since now one fourth of subslots now overlap two slots at a neighbor—see Chapter 3), not much more can be obtained here.

The delay performance of packets wishing to transit a large network will be poor. Scaling from 1,000 nodes up to 1,000,000 nodes would increase the delays across the network by a factor of $\sqrt{1000}$ or approximately 32. End-to-end delays would then be in the range of one minute, without any congestion in the network. These delays would be unsuitable for any of the interactive applications that are used in the Internet today, but might be suitable for non-interactive applications such as large file transfer, e-mail, and distribution of Usenet news (but not interactive browsing or reading).

5.2.1 Delay per hop

To demonstrate the results in Section 3.1.6 for the delay across a single hop, the experiment that produced the middle plot of Figure 5.8 (1,000 stations, 64 scheduling slots per second) was run again, but the end-to-end delay observed for each packet was divided by the number of hops the packet traversed. Figure 5.9 shows the distribution of average per-hop delays seen by each packet. The average delay over all hops taken by all packets was 0.061 seconds per hop. At 64 slots per second, the delay is 3.9 slot times, which is less than (but within one slot time of) the value of 4.76 slot times predicted by the simpler Bernoulli model in Section 3.1.6.

5.3 Summary of system performance

In this section, a summary of parameters and performance will be presented for a design of the packet radio system.

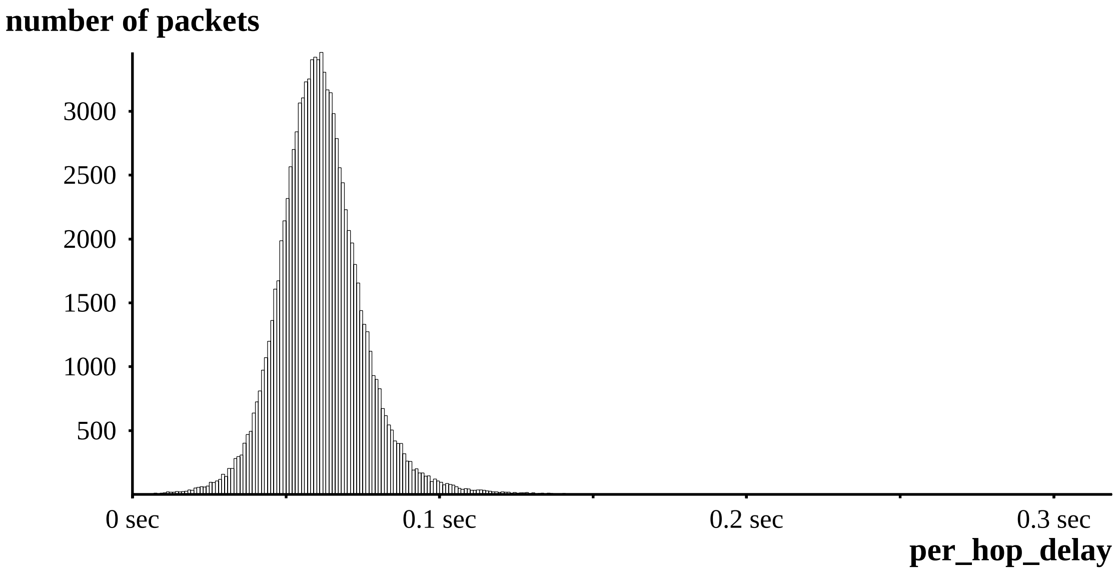

Figure 5.9: Distribution of measured end-to-end delays divided by number of hops, or the average per-hop delay, experienced by each packet in a lightly-loaded 1,000-station network with minimum energy routes, 64 slots per second, and 16 subslots per slot.

5.3.1 Single hop performance

Here we determine the performance across a single hop, from one station to a single neighbor. The throughput determined here will be the maximum throughput to a single neighbor, but does not include additional throughput capacity that could be used to send to additional neighbors. The delay estimated here will be the delay across a single hop when no other traffic is competing for capacity (so there are no queueing delays). The single-hop performance is governed by the following parameters:

- W, The bandwidth in Hertz. The amount of radio spectrum allocated to the system. This will also be roughly the highest clock frequency that must be present in the digital circuitry used to generate and detect the spread-spectrum signals.
- s, The number of scheduling slots per second.
- PG, the processing gain used in the spread-spectrum system, at least 100, maybe as much as 300.
- m, The number of subslots per slot, suggested value of 4.
- p, the receive duty cycle, the probability that a given pseudo-random scheduling slot is a receive slot. Recommended value of 0.3 as determined in Chapter 3 and from Figure 3.3.
- OH, the overhead bits per packet. Probably a dozen bits for spreading code synchronization plus whatever overhead is needed by the network, for a total of somewhere between 50 and 500 bits (depending on the network protocols used).

Values are suggested or estimated for all parameters except for the first two. The suggested or estimated values are largely independent of the bandwidth available to the system and are believed to be appropriate for a wide variety of situations. The bandwidth available, W, will be limited by government regulation or by technological limits on signal processing. Once the bandwidth is determined, the system designer must select a value for s, the number of scheduling slots per second. Here we will develop expressions for the throughput, delay, and packet size as a function of the parameters and then examine the tradeoffs involved in selecting a value for s.

The spread-spectrum chip rate will be approximately equal to the bandwidth W. Direct-sequence spread spectrum can send one data bit for every PG chips. Hence the raw data rate during transmission is $\frac{W}{PG}$. As was seen in Section 3.1.2 the fraction of time we expect to be able to spend sending to a single nearby neighbor is $\frac{m-1}{m}p(1-p)$, which for the values estimated above will be approximately 0.16. Combining these factors, the maximum raw throughput α (including overhead) from one neighbor to another is approximately

$$\alpha = \frac{W}{PG}\,\frac{m-1}{m}\,p(1-p). \tag{5.1}$$

For the suggested values above, α/W will be around 1 bit per second per kilohertz. But α is the total throughput including the overhead bits. The number of overhead bits per second, β, is

$$\beta = OH\,\frac{(m-1)}{m}\,p(1-p)\,sm. \tag{5.2}$$

So the data bits per second excluding overhead is $\alpha - \beta$. The efficiency while transmitting is then

$$\frac{\alpha-\beta}{\alpha} = \frac{W/PG - OH\,sm}{W/PG} \tag{5.3}$$

$$= 1 - \frac{OH\,sm}{W/PG}. \tag{5.4}$$

The number of data bits carried in a packet is

$$\frac{W}{PG\,sm} - OH. \tag{5.5}$$

Either the efficiency, $(\alpha-\beta)/\alpha$, or the number of data bits per packet, or both will limit how high s, the number of scheduling slots per second, may be set. It should not be made so large that the efficiency drops significantly below one. In the case of the packet size, there may be some minimum packet size that the network must be able to carry while avoiding the complexities of fragmentation and reassembly.

Finally, the expected amount of delay per hop due to scheduling, D_{hop}, in an unloaded network is

$$D_{\text{hop}} = \frac{1}{s}\,\frac{1}{p(1-p)}. \tag{5.6}$$

This makes higher values of s desirable, as scheduling delay is then reduced.

5.3.2 Raw system performance with purely local traffic

In some cases where there is some amount of locality in traffic load, the techniques described in this work will perform better than more traditional multiplexing techniques (such as system-wide time division). As an example, imagine that the traffic demands were ideally matched to the network's capability. This occurs when all traffic crosses only one hop (using the minimum energy routes). It is the same as the traffic that was generated in Chapter 4 to load up all of the single hops. In this case, for networks that are at least moderately large (of 1,000 stations or more), spread spectrum's simultaneous sharing of the channel will allow greater throughput than an ideally scheduled system that lets each transmission have exclusive use of a clear channel.

For a single channel of bandwidth W (in Hertz), the maximum throughput when the channel is otherwise clear is in practice limited to a few bits per second per Hertz of bandwidth. Even with extremely good signal-to-noise ratios, the Shannon bound grows as only the log of the signal-to-noise ratio, so at most a small factor k times W bits per second would be possible. If some ideal scheduler were able to schedule the transmissions among all stations so that no overlap occurred but so that the channel was always in use, then the total throughput of the system would be at most kW bits per second. The total system throughput would remain constant, independent of the number of stations. As the number of stations increases, the fraction of time available to each station will be reduced.

For a network with purely local traffic using the scheme presented in this work, multiple transmissions will overlap and interfere with each other. The signal-to-noise ratio in a large system will be roughly 0.01, or -20 dB, as was shown in Chapter 2 and demonstrated in Chapter 4. At this low signal-to-noise ratio, each transmission will only be able to communicate at roughly 0.01 bits per second per Hz of bandwidth, or $0.01W$ bits per second. But unlike the case in the previous paragraph, at any given time a constant fraction of stations, independent of network size, may make simultaneous use of the channel. This constant fraction may be somewhere around one third to one half of the stations (which is suggested by Figure 4.6). For a network of only 100 stations, this scheme will not perform as well as a scheme that has only one transmission at a time (assuming at most one bit per second per Hertz is possible in the clear channel), as 50 (one half of 100) transmissions in parallel cannot win back the factor of one hundred penalty in raw throughput. But at 1,000 stations, where 300 to 500 stations may be transmitting simultaneously, then a factor of 300 in parallelism is more than sufficient to make up for the lost raw performance of the channel. This advantage (for this contrived traffic load) improves as the system scales to even more stations because in this case the fraction of time available to each station will remain constant as the system scales.

Hence the simultaneous sharing of the channel is advantageous for performance even with the reduction in raw channel performance, if there is sufficient locality in the traffic. For purely local traffic this advantage increases as the system scales. In other words, a system that allows only one transmission at a time cannot scale while maintaining any level of per-station performance. A system with distributed, simultaneous use of the channel can scale while maintaining the level of performance if the traffic is local.

The system can scale reasonably even if the the traffic is not strictly local. In the 1,000-station network, the advantage over one-at-a-time would be maintained if traffic were on average taking fewer than 3 hops. The per-station performance of the system can be maintained as the system scales if the average number of hops taken remains small.

5.4 Applicability

Reasonable performance of a multihop packet radio system using the methods presented in this work depends on the locality of the traffic. Traffic loads with insufficient locality will swamp large systems with congestion. Hence large-scale multihop packet radio networks will be limited to applications that have either locality in the traffic or low demand on system performance. In small scale networks (of a few number of stations, not necessarily limited in geographic extent) all traffic is local when compared to large scale networks, hence a wider variety of traffic patterns could be supported if the number of stations was not large.

If the network is large and the traffic demand is not sufficiently local, then this sort of packet radio network alone will not be able to meet the demands. Multihop packet radio networks are not, however, completely useless in this situation. An overlaid wireline infrastructure could be added to an existing packet radio network to alleviate congestion by carrying the long-haul traffic. While the resulting system might not provide any performance advantage over a more traditional asymmetric cellular system (in which user stations communicated directly only with base stations), some important operational advantages of multihop remain. With multihop, the initial deployment can be without any infrastructure at all. The user stations in a multihop system are not entirely dependent on the existence or correct operation of any infrastructure to be useful, unlike cellular telephone equipment.

One area where multihop systems might be useful even if the traffic were not strictly local is for netnews-like flooding. In this application each station originates only occasional traffic, but all traffic is received and collected by every station. This application might seem like a good match for a system where each station makes one transmission of the traffic it originates which is received by all other stations all at once. The maximum throughput in this case would be at most a small constant times the channel bandwidth (for the same reasons as in the previous section). But for large networks, multihop systems will be advantageous. In a multihop system, flooding a message using single unicast hops will require one transmission to each station. With these transmissions performed in parallel by the multihop system, simultaneous flooding of messages at a rate greater than the clear-channel rate is possible. This is because the system can take advantage of the isolation provided by the free-space loss in radio signals to enable spatial re-use. Furthermore, the multihop system does not require that every station be able to hear every other station, and does not require any system-wide management of the channel.

Chapter 6

Conclusions

By using spread-spectrum techniques and by using the scheduling algorithm of Chapter 3, stations in a large packet radio network can communicate, without collisions, with nearby neighbors. The techniques developed require no system-wide coordination to manage access to the channel. The only coordination required is pairwise between neighbors who are in direct communication, and the results of coordinating with one neighbor have no consequence in the coordination with other neighbors, so there is no transitive dependency in the coordination. With this completely decentralized control, and with a practically-constant lower bound on the signal-to-noise ratio of signals from nearby neighbors as the system scales, packet radio networks can scale to almost arbitrary numbers of stations.

Multihop routing, using minimum energy routes, can be used to communicate beyond the immediate neighbors. Minimum energy routes also determine which stations are immediate neighbors. In simulation with 1,000 stations, minimum energy routes were demonstrated to result in a fully connected network in which all packets could be received with acceptable signal-to-noise ratios under a worst-case traffic load (which kept the transmitters busy at all times that they could be sending a packet to a neighbor). Depending on the pattern of offered traffic, congestion may be a problem. If traffic is distributed uniformly from and to each area of the net, then congestion will result. This congestion grows worse as the system scales to large numbers of nodes.

An attempt to alleviate the congestion by modifying the routing algorithm to have a slight preference for paths that traverse fewer hops (reducing the number of end-to-end paths sharing transit stations and reducing delay) resulted in a decrease in signal-to-noise ratio that more than offset any gain in throughput from reduced congestion. Minimum energy routes appear to be important for maintaining reasonable signal-to-noise ratios.

The congestion implies that the system cannot scale arbitrarily while maintaining a per-station performance guarantee for arbitrary traffic patterns. But for traffic patterns that are sufficiently local, the techniques developed here are capable of guaranteeing a nondecreasing level of service as the system scales. For a network of 1,000 stations, using the techniques developed here is advantageous over sending all packets directly from source to destination using ideal time division multiplexing on a clear channel, if on average the packets are traversing fewer than three hops.[1] The advantage is due to the spatial re-use enabled by allowing simultaneous use of the channel. The total system throughput in the clear-channel time-division multiplexing case remains constant as the number of stations increases. But with the techniques here, as long as the traffic is bounded to a maximum number of hops, the total system throughput will scale linearly with the number of stations.

6.1 Contributions

The important contributions in this work are:

- This work is the first in scalable multihop packet radio networks at the system level to use a model of propagation and detection that captures the effect of interference from distant stations. The commonly used transmission radius model neglects the effect of combined interference from distant stations which is important to understanding the effect of noise on system performance.
- The analysis in Chapter 2 shows that, as the system scales, the noise level remains bounded with respect to the signal level from a nearby neighbor.
- The novel pseudo-random scheduling method presented in Chapter 3 can achieve collision-free transmission of packets without the need for any system-wide synchronization or system-wide coordination.
- Minimizing the total energy of each packet's transmission through the network is a straightforward and practical method of choosing routes in a multi-hop network. In general, stations might not know their geometric locations, but the path gain between two stations can be measured directly by the stations when they rendezvous.

6.2 Applications

The techniques in this thesis are applicable to more than just large-scale wireless networks. Small-scale packet radio networks can benefit from robustness of high processing gain and decentralized channel management. Such networks would include wireless local area networks for in-building use, or wider area networks of lesser density. Even for networks that are not anticipated to grow to a very large scale, the use of high processing gain adds robustness against interference from other sources, and (perhaps more importantly) guards against failure-by-success should the system become popular and grow to a larger size than anticipated.

Large-scale packet networks will probably need a significant amount of wireline infrastructure to extract the congestion from the network. Even so, the techniques

[1] There are probably more flexible criteria for "sufficiently local" than just bounding the maximum number of hops that the traffic is allowed to take.

described here would work well for "last mile" systems. The multi-hop ability of the networks developed here would allow for initial deployment to be without infrastructure. If the system becomes popular, then the investment could be made in infrastructure incrementally as the number of users grows. For example, a utility company could run wires on the poles in a town to pole-top boxes.

The widest of the existing "Part 15" unlicensed spread-spectrum bands in the U.S. is 125 MHz wide. (It runs from 5725 to 5850 MHz.) The raw throughput (the data rate while transmitting) of a system using 23 dB of processing gain would be around 625 thousand bits per second (kbps). The maximum throughput from one station to a single neighbor, the α of Section 5.3.1, would be about 94 kbps. Two simultaneous streams out of a station could each run at around 86 kbps. This rate is barely enough to match the data rate offered by telephone companies as ISDN. A transit station on a path could barely deliver enough throughput to support one 64 kbps digital telephone call passing through it. With 1024 subslots per second, each subslot could handle a packet of around 76 bytes. With four subslots per slot and an expected delay of around 5 slot times for each hop, the expected per-hop delay would be 20 ms. These delays would likely be intolerable. This performance might not be attractive in the marketplace, particularly when compared with ISDN service which gives each user a 64 kbps.

If a 2 GHz wide swath of spectrum could be obtained (most likely at some frequency over 30 GHz), then 23 dB of processing gain yields a raw data rate of about 10 million bits per second (or 10 Mbps). The maximum throughput to a single neighbor, α, would then be 1.5 Mbps. At this data rate, and again with 1024 packets per second, each packets could hold around 1200 bytes. This packet size and throughput are comparable to the performance of 10 Mbps Ethernet or a 1.5 Mbps leased T1 line. Ethernets, which share their 10 Mbps among all of their users, are commonly used with populations of a few dozen or more on each in the Internet today. While building the radio electronics today to make use of this wide of a bandwidth would be ambitious today, there may still be an unmet demand for this level of networking service in a few years. However the per-hop delay due to the scheduling would still be around 20 ms. (If the delay experienced by short packets is the primary concern, then smaller slots and subslots could be used.)

6.3 Future work

Designing an economically viable system will require additional research. Only a small portion of the large design space for station behavior has been explored in this work. A few ideas for further investigation are presented here.

6.3.1 Directional antennas

One promising method of improving performance is to use directional antennas. Every complete 3 dB improvement in SNR is a factor of two improvement in throughput, as processing gain can then be reduced. Electronically steerable phased arrays that could make a 6 dB improvement in the desired direction over omnidirectional antennas would increase the signal-to-noise ratio by 12 dB (6 dB at transmitter and 6 dB at receiver). Though the need to steer the directional antennas would complicate the initial rendezvous and ongoing control, this increase in gain would amount to a 16-fold increase in throughput that might make the 125 MHz-wide band at 5 GHz viable for a system with reasonable performance. Such a system could probably be built in prototype form with the electronics available today.

A system 2 GHz wide with 12 dB or more in combined gain from directional antennas (6 dB at both the sender and receiver) would be quite attractive. If the system can operate with only 10 dB of processing gain (and 12 dB of gain from directional antennas) then the raw data bit rate while the transmitter is running would be around 200 megabits per second, and α would be 30 megabits per second. There could be 10,000 sub-slots per second for a packet size of around 2500 bytes. With four sub-slots per scheduling slot, there would be 2500 scheduling slots per second, so expected (low-load) per-hop scheduling delay of 5 slots would be 0.002 seconds. This packet size and throughput would be clearly better than Ethernet throughput performance for each hop, and the delays would be comparable to Ethernet. The electronics needed to build this system could be available in a decade.

6.3.2 Propagation modeling

Further refinement of the propagation model is needed. In addition to modeling the delay and multipath propagation, the propagation of a candidate frequency band should be measured empirically from likely deployment locations. Stations with indoor antennas are an attractive possibility, but may be precluded by attenuation and multipath effects. Not much work can be done in this area without making empirical measurements to determine the correct parameters for modeling attenuation and multipath.

6.3.3 Routing

This work has neglected one important implementation problem. If there are millions of stations in an area, how are the stations to find routes to or even learn the address of the stations they wish to talk to? In networks this large it is unreasonable to expect each station to maintain a routing table entry for every possible destination. Some method for automatically aggregating routes to various regions is needed, as well as a method for naming stations and finding the names of stations offering services. A study of the published literature (e.g. [Tsu88]) would be a first step.

Minimum-energy routes need further examination. In this work they were initially tried as a stopgap measure to determine routes for simulating multihop traffic using the scheduling techniques of Chapter 3. The resulting congestion observed in the simulations (discussed in Chapter 5) led to some discussions about how to optimize routing. Though we were unable to formulate the optimization problem in linear form, we came to believe that keeping the route energies low is important for system performance. However there may be more complicated methods of determining routes that can lead to somewhat less congestion in some systems, without impacting the signal-to-noise ratio. Directional antennas may also change the nature of this problem.

One approach is to find a *figure of merit* for the system that collapses many different system parameters and measures of performance into a single scalar. The right figure of merit would aid comparisons between different possibilities for system operation. The goal is to capture the essence of the "goodness" of a system in a single number. What exactly would make one system better than another? For a given traffic load, total throughputs and delays might somehow be combined into a single measure of goodness, say average throughput divided by average delay. But another criterion for comparing two packet radio systems might be their ability to handle a wide variety of differing traffic loads. In [Kar91] Karn proposes $\frac{rD}{BA}$ where r is the total system bit rate, D is the average distance between nodes, B is the channel bandwidth, and A is the area of the system. With this metric, a given technique for controlling a packet radio network yields different values of merit as the system is scaled in linear size, as the capacities between stations would remain identical. But as we have seen, with free space propagation, scaling in linear size does not change the signal-to-noise ratios in a system that is limited by self-interference. Therefore we believe that Karn's figure of merit can be improved upon.

An interesting question is to determine the dimensionality of the network resulting from minimum-energy routes among stations distributed randomly in the plane. For example, if we scaled the network of Figure 4.9 up to 10,000 randomly placed stations and examined the amount of sharing (by plotting a distribution like in Figure 5.1) we do not know what fraction of the 10^8 routes will pass through the most heavily shared stations. If the dimensionality is 2, then we would expect the number of paths through a station to grow $O(M^{\frac{3}{2}})$ (where M is the number of stations). To see this, imagine dividing the region in half. The number of end-to-end paths that cross the divide will grow as the square of the number of stations (for any number of dimensions). In two dimensions, the number of ways to cross the divide would be expected to grow as the square root of the number of stations, so the M^2 paths would be shared among the $\sqrt{M}$ crossings. It is possible that the routes resulting from minimum energy routing lead to a growth in the number of routes through some stations at a rate faster than $O(M^{\frac{3}{2}})$. In that case it would be said that the dimensionality is less than 2.

6.3.4 Infrastructure

A large packet radio network would not exist in isolation. There would likely be many "exit points" that would be connected to service providers or to the global Internet. Even within a single metropolitan area, communication outside of a local area will probably best be provided by exiting the system at some point and traveling over wired networks. The placement of these exit points has not been investigated here. One simple problem that might be investigated (within the spirit of the randomly placed stations) would be to install large-capacity tunnels between randomly selected stations and examine the resulting routes and congestion. It might not take many of these randomly placed tunnels to alleviate much of the congestion seen in Chapter 5. Another question is to determine at what rate the infrastructure must grow as the network grows. These questions again depend on routing and the demand for service; questions that are difficult to determine before real users are involved in the generation of traffic.

Bibliography

[Abr70] Norman Abramson. THE ALOHA SYSTEM–Another alternative for computer communications. In *Proceedings of the Fall Joint Computer Conference*, pages 281–285. AFIPS, 1970.

[Abr93] Norman Abramson, editor. *Multiple Access Communications*. IEEE Press, 1993.

[BG87] Dimitri Bertsekas and Robert Gallager. *Data Networks*. Prentice-Hall, 1987.

[CCI86] CCIR Rep. 719-2. Attenuation by atmospheric gases. In *Recommendations and Reports of the CCIR*, volume V, pages 167–177, Geneva, 1986. CCIR, International Telecommunication Union.

[CEMS83] Charles E. Cook, Fred W. Ellersick, Laurence B. Milstein, and Donald L. Schilling, editors. *Spread Spectrum Communications*. IEEE Press, 1983.

[Dix84] Robert C. Dixon. *Spread Spectrum Systems*. John Wiley & Sons, 1984.

[Fre87] Roger L. Freeman. *Radio System Design for Telecommunications (1-100 GHz)*. John Wiley & Sons, 1987.

[Gal85] Robert G. Gallager. A perspective on multiaccess channels. *IEEE Transactions on Information Theory*, IT-31(2):124–142, March 1985.

[IEE87] IEEE. *Proceedings of the IEEE*, volume 75, number 1, January 1987. Special Issue on Packet Radio Networks.

[Kar91] Phil Karn. Spectral efficiency considerations for packet radio. In *ARRL 10th Computer Networking Conference*, pages 62–66. ARRL, 1991.

[KGBK78] Robert E. Kahn, Steven A. Gronemeyer, Jerry Burchfiel, and Ronald C. Kunzelman. Advances in packet radio technology. *Proceedings of the IEEE*, 66(11):1468–1496, November 1978.

[KS87] Leonard Kleinrock and John Silvester. Spatial reuse in multihop packet radio networks. *Proceedings of the IEEE*, 75(1):156–167, January 1987.

[LB87] Clifford A. Lynch and Edwin B. Brownrigg. *Packet Radio Networks*. Pergamon Press, 1987.

[ML92] Robert K. Morrow and James S. Lehnert. Packet throughput in slotted ALOHA DS/SSMA radio systems with random signature sequences. *IEEE Transactions on Communications*, 40(7):1223–1230, July 1992.

[Nic88] David L. Nicholson. *Spread Spectrum Signal Design.* Computer Science Press, 1988.

[Pur87] Michael B. Pursley. The role of spread spectrum in packet radio networks. *Proceedings of the IEEE*, 75(1):116–134, January 1987.

[RR64] Henry R. Reed and Carl M. Russell. *Ultra High Frequency Propagation.* Boston Technical Publishers, second edition, 1964.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, XXVII(3 and 4):379–423 and 623–656, July (part 1) and October (part 2) 1948. Also republished in [Sle74], [SW93], and [SW63].

[Sle74] David Slepian, editor. *Key Papers in The Development of Information Theory.* IEEE Press, 1974.

[SOSL94] Marvin K. Simon, Jim K. Omura, Robert A. Scholtz, and Barry K. Levitt. *Spread Spectrum Communications.* McGraw-Hill, second edition, 1994.

[SS88] Elvino S. Sousa and John A. Silvester. Spreading code protocols for distributed spread-spectrum packet radio networks. *IEEE Transactions on Communications*, 36(3):272–281, March 1988.

[ST89] James S. Storey and Fouad A. Tobagi. Throughput performance of an unslotted direct-sequence SSMA packet radio network. *IEEE Transactions on Communications*, 37(8):814–823, August 1989.

[Sto87] James Scott Storey. Throughput performance of spread spectrum multiple access packet radio networks. Technical Report CSL-TR-87-319, Stanford University Computer Systems Laboratory, March 1987. Ph.D. Thesis.

[SW63] Claude E. Shannon and Warren Weaver. *The mathematical theory of communication.* University of Illinois Press, 1963.

[SW87] Nachum Shacham and Jil Westcott. Future directions in packet radio architectures and protocols. *Proceedings of the IEEE*, 75(1):83–99, January 1987.

[SW93] N. J. A. Sloane and Aaron D. Wyner, editors. *Claude Elwood Shannon Collected Papers.* IEEE Press, 1993.

[TK84] Hideaki Takagi and Leonard Kleinrock. Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Transactions on Communications*, COM-32(3):246–257, March 1984.

[Tro94] Gregory D. Troxel. Time surveying: Clock synchronization over packet networks. Technical Report MIT/LCS/TR-623, Massachusetts Institute of Technology Laboratory for Computer Science, May 1994.

[Tsu88] Paul F. Tsuchiya. The landmark hierarchy: A new hierarchy for routing in very large networks. In *SIGCOMM Symposium, Stanford, CA*, pages 35–42. ACM, August 1988.

[Ver89] Sergio Verdú. Recent progress in multiuser detection. In William A. Porter and Subhash C. Kak, editors, *Lecture Notes in Control and Information Sciences*. Springer-Verlag, 1989. Also republished in [Abr93].

[ZP85] Rodger E. Ziemer and Roger L Peterson. *Digital Communications and Spread Spectrum Systems*. Macmillan, 1985.