New Technological Vistas for Systems and Control: The Example of Wireless Networks*

P. R. Kumar†
Department of Electrical and Computer Engineering, and
Coordinated Science Laboratory
University of Illinois, Urbana-Champaign

May 8, 2000

Abstract

Over the past four decades the field of control has woven a rich tapestry of a larger systems theory, with sustained investigations into fundamental issues such as distributed control, estimation, adaptation, stability, optimality, etc. These issues are the fundamental ingredients in many new proposed technologies, which are now within our collective purview. They provide a profusion of practical examples of problems which may have sometimes abstractly engaged our attentions in the past, and offer a wealth of opportunities for imaginative solutions. The opportunities are ours to seize. That is the central thesis of this article, which illustrates it using the example of wireless networking, an area of great interest in the emerging field of information technology.

*This material is based upon work partially supported by the Office of Naval Research under Contract No. N00014-99-1-0696, the Air Force Office of Scientific Research under Contract No. AF-DC-5-36128, by EPRI and the U.S. Army Research Office under Subcontract to Cornell University Contract Nos. WO8333-04 and Army OSP 35352-6086. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the above agencies.

†University of Illinois, CSL, 1308 West Main Street, Urbana, IL 61801. Email: prkumar@decision.csl.uiuc.edu. Web: http://black.csl.uiuc.edu/~prkumar

1 Introduction

Over the past four decades, research in the field of control theory has woven a rich tapestry of systems theory. Notions such as stabilizability, reachability, optimality, identification, adaptation, robustness, estimation, information structures, games, control with partial noisy observations, distributed control and estimation, linearity, nonlinearity, infinite dimensional systems, discrete-event systems, hybrid systems, nonholonomic systems, etc., have been systematically explored in depth. In its search for deeper answers, the community as a whole has been willing to invest in learning various mathematical fields to exploit any and all tools. Complex analysis, Martingale theory, heat equations, differential geometry, automata theory, functional analysis, calculus of variations, algebra, category theory, algebraic geometry, Brownian motion, etc., have all been avidly studied, with the result that their usage has become routine. The result has been the creation of a rich metaphor of concepts which allows us to quickly locate the central issues in a given problem formulation.

One can broadly discern two thrusts in the above research. One has been the focus on the core problem of control loop design for linear and nonlinear differential equations. This has given us LQG control, self-tuning control, identification methodologies, H_∞-control, ℓ_1-control, nonlinear control, backstepping, Kharitonov extensions, etc. Over the past decade, a profusion of software tools has been developed which has made these results widely accessible. The other thrust has been expansionary, and has served to expand the frontiers of system theory. Topics such as learning, distributed estimation, distributed asynchronous algorithms, stochastic networks, team theory, stochastic algorithms, optimization theory, stability, etc., have explored. The result has been that researchers with roots in control theory have made distinguished contributions to other fields such as power systems, robotics, learning theory, economics, signal processing, optimization, stochastic processes, communication networks, wireless networks, operations research, financial economics, manufacturing systems, etc. Our workshops and flagship conferences are routinely attended by researchers from all these disciplines. It is interesting, if for nothing else than as a purely sociological phenomenon, that a field beginning with problems such as Watts' governor has gone on to make such a broad impact.

What then does the future hold? What are the vistas and opportunities in the coming one or two decades? What are the current trends affecting technology in general? What lessons should we learn from the past? How can we build on our past successes and assure ourselves of an even brighter future?

Perhaps two principles could profitably be kept in mind. The first is that the control field may be well served by paying careful attention to technology and its trends. Many interesting problems abound in automotive technology, semiconductor manufacturing, power systems, filter design, networking, smart materials, etc. Moreover, future society may be radically altered by trends in nanotechnology, the deregulation of the radio spectrum, advances in materials science, biotechnology, robotics, etc. It may be beneficial for researchers to each become domain experts in at least some one technology. By learning the fundamental details of a technology, one is in position to discover problems which may lie hidden. By starting at the bottom with the basic problem, one can play a role in the very modeling of the issue, which opens up many possibilities for design and analysis. Also by knowing the technology, one is in a position to know where the current bottlenecks are, since advances on bottlenecks have a greater impact than advances made on non-bottleneck issues. Moreover, by starting with technology, the motivation of our research will be transparent to others. In the current utilitarian problem driven environment, it is important for the broader scientific and engineering community to understand the context of our work.

The second issue is that we may be well served by preserving our culture of rigorous and deep thinking, and the tradition of being unafraid to invest in learning new approaches and theoretical tools. A broad and deep knowledge is invaluable in capitalizing on research opportunities. It allows one to quickly discard unworkable approaches, and identify promising ones. Over the long haul, a case can be made that investment in learning does pay off.

I will now take up the field of wireless networks as an example of one fertile future vista which can serve to illustrate some of the above comments. We will see that the issues of stability, adaptation, information structures, uncertainty, distributed control, all concepts which our field has researched, sometimes in abstraction, are present in practical profusion. We will see that modeling the phenomena, and identifying the research issues, allows much scope for original work. The field is of intense current interest, and regarded as possibly revolutionary in its impact on society. The bottleneck is indeed the design of a workable system. Finally, wireless networking is just one topic in the broader field of information technology, which is currently the subject of much attention.

Indeed the central thesis of this article is just this: Issues which have collectively engaged our community over the past four decades, such as optimality, estimation, stability, adaptation, robustness, uncertainty, partial observations, distributed systems, etc., are the fundamental ingredients in many future technologies, and by bringing to bear our tradition of rigorous analysis and design, our field has a bright future, as well as a significant role to play in the realization of several technologies. The present age is characterized by numerous attempts to build such complex systems. All these future systems can well be be added to our list of core control systems.

2 Wireless Networks

The type of wireless networks that we will describe have, at various times in the past, been called packet radio networks or multi-hop mobile radio networks; the current term in vogue is *ad hoc-networks*.

Currently there is much excitement about the future possibilities realizable by wireless networks. Wires have simply become a barrier to prolification. When nodes are untethered, new possibilities arise. For example, in recent decades there has been a great growth in embedded devices. These are present in alarm clocks, automobiles, cell phones, toasters, oil wells, and in fact all around us. However they have not hitherto operated in concert since they have been disconnected from each other. In the future, toasters may coordinate with alarm clocks, travelers may acquire maps or videos from local information warehouses, automobiles may warn each when braking, and wireless trading may become a reality. Of course, it cannot be predicted which applications will be successful and which will not. All one can say is that such technological capabilities are on the horizon.

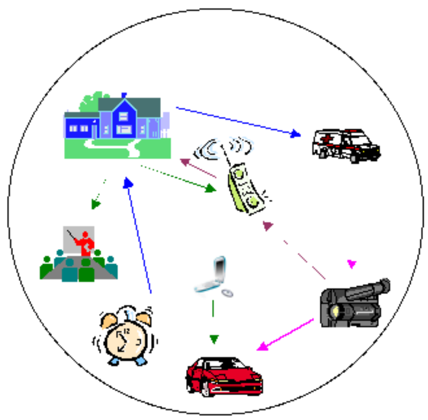

Figure 1: An ad hoc wireless network

Consider a domain within which are located a number of nodes, as shown in Figure 1. Each node is equipped with a wireless transmitter capable of transmitting packets to its neighbors. Packets can be relayed from node to node until they reach their final destination. Nodes may be mobile, changing their location over time. Also, nodes may switch themselves off from time to time.

The objective of operating such a network is to transfer packets from their sources to their destinations reliably and efficiently. Nodes must choose the power levels at which they broadcast, since that influences the range, and time their transmissions. Nodes must also cooperate in relaying each others' packets.

The wireless medium is an unreliable medium. Transmission are subject to obstacles, reflections, multipath effects, and fading, all of which effect the quality of the received signal.

Transmissions can moreover interfere with each other. All these conspire to make packet reception much more unreliable than in wired media.

Given the characteristics of the wireless medium, how does one design a system which can reliably and efficiently transport packets from sources to destinations?

3 The Goal: Quality of Service

There are two very important performance measures that affect an user's experience of a wireless network, or any other communication network for that matter. One is *throughput* which is measured in bits per second. This is the rate at which bits are transferred from the source to their final destination. The second is *delay*, which is the difference in time from when a bit enters the network at its source node to when it is eventually received correctly at its destination node.

For real-time or interactive services such as voice or video, timelessness of receipt of packets is crucial. Hence delay is an important performance measure. However, real-time voice can afford to lose some packets and still sound intelligible to the listener. So it may be a reasonable strategy to drop some packets if they have been excessively delayed en route, or to not retransmit them again if they have been corrupted en route. On the other hand, for data transfer, e.g., file transfer, delay may not be important, but accurate and complete reception may be paramount.

There are also other performance measures which are derived from the above performance measures. For example, the mean delay of packets, or the standard deviation of the delay, or even the 99-th percentile of delay, may all be important indicators of performance. Another could be *fairness*. When many users are active in the network it may be important that each receives a nearly equal throughput and/or delay. Or, if some users are more important than others, for example if they have contracted for a greater quality of service, their throughput or delay may be weighted more heavily than others.

While the internet currently does not provide much in the way of guaranteeing quality of service to users, it is felt that future users will require specified levels of quality of service [1]:

> The Holy Grail of computer networking is to design a network that has the flexibility and low cost of the Internet, yet offers the end-to-end quality-of-service guarantees of the telephone network.

How this is to be done in the present internet, let alone future wireless networks, is a topic of much current research.

4 The Power Control Problem

Our fundamental issue faced by a node on every transmission is how to choose the the power level at which it should transmit. Clearly the power level should be high enough that the intended receiver receives a signal of adequate power. However, it should not transmit at too high a power level since that may interfere with other receivers receiving separate transmissions of interest to them. One more factor to be kept in mind is that during periods of deep fading it may be better to lose a packet and retransmit it later than to broadcast it at a very high power level. This is related to the “waterfilling” solution in Information Theory [2].

These issues, and others to be described below, give rise to the *power control problem.* It can be regarded as a feedback based set-point regulation problem. The receiver can provide a feedback signal to the transmitter allowing it to regulate its transmitted power level so that the net received power level at the receiver, or the net received signal-to-interference plus-noise ratio at the receiver, is at a desired level. (We should mention here that every problem in wireless networks is considerably exacerbated if the links are not bidirectional).

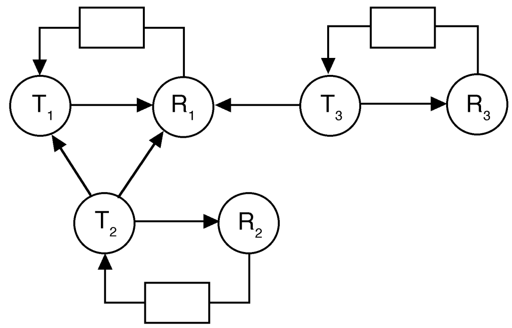

Figure 2: Coupling between power control loops of transmitter-receiver pairs

There is another issue to be considered though. Several transmitter-receiver pairs may be simultaneously operating, with each having its own feedback loop; see Figure 2. Therefore, when one transmitter increases its power level based on its own feedback, this is experienced as increased interference at another receiver receiving a different transmission, thus causing it to send a feedback signal to its own transmitter to increase its power level. Clearly, this causes coupling between loops, with each transmitter raising its power level in response to another transmitter doing so. Thus, one needs to study and design schemes for power control which assure the convergence of the power levels of all transmitters. This is a problem close to the heart of all control researchers. It can be solved, for example, by discretizing the power levels, and each transmitter raising its power level just enough to satisfy its own receiver's desired signal-to-interference ratio; see [3] Solutions are also available for this "physical layer" problem of maintaining a link of adequate quality when the power levels are continuously varying; see [4],[5],[6].

Unfortunately there are other complicating issues in ad-hoc networks. High power transmissions cause interference, which results in a low throughput for other users. Thus the power control problem impinges on the "congestion control" problem, which is managed in the current internet by protocols such as TCP [7],[8],[9],[10]. A user's throughput is increased until a packet is lost, at which time the throughput is drastically curtailed. This is done by giving each user a "window," which is the number of unacknowledged packets it can have in the network, and increasing or decreasing the window. The essential point is that when the network is congested, packets are dropped at full buffers, and so packet loss is taken as the appropriate "feedback" signal to regulate congestion.

However, in wireless networks, one's congestion can be caused by another's usage of an excessive power level. How does one regulate the behavior of others? How can one use feedback across transmitter-receiver pairs? These questions require further investigation.

Another issue is for all nodes to cooperate in maintaining systemwide network connectivity. It is shown in [11] that if n nodes are randomly (uniformly iid) located in a disk of area A sq. m., and each user chooses the same range $r(n)$, then the network is asymptotically connected with probability one as $n \to \infty$ if and only if

$$r(n) = \sqrt{\frac{\log n + \gamma(n)}{\pi n}},$$

with $\gamma(n) \to +\infty$.

Finally, it is shown in [12] (see also Section 6), that under certain models, systemwide efficiency of the wireless network is optimized when every "hop" covers a very short distance. Nodes should thus relay packets over very short distances to nearby nodes, thus allowing them to transmit at low power. (However, as noted above, care must be taken to prevent disconnectivity of the network). This will entail a packet to be relayed repeatedly, traversing several hops before reaching its destination. This strategy is preferable to broadcasting at a very high power level and reaching the remote destination in just one hop. Thus, power control affects the capacity of the system, and also impinges on the routing problem considered in Section 5.

As can be seen from the above discussion, power control cuts across many layers in the design hierarchy of wireless networks and hence requires a proper conceptualization.

4.1 The Medium Access Control Problem

Several problems in wireless networks arise from the fundamental shared nature of the wireless medium. One is the power control problem described in the previous section. Another is the Medium Access Control Problem described in this section.

For a receiver to intelligibly receive a packet, other nearby transmitters should refrain from broadcasting. Otherwise, one has the phenomenon of "collisions" where packets destructively interfere with each other.

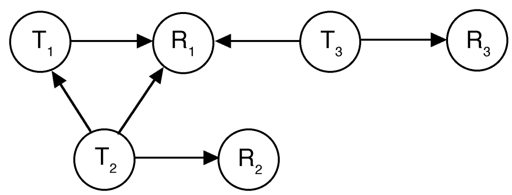

Figure 3: Interference and hidden nodes

Consider the scenario shown in Figure 3. Assume that T_1, T_2, and T_3 are all within range of the reception by R_1. Then, for R_1 to successfully receive a packet from T_1, both T_2 as well as T_3 have to remain silent.

Note that "carrier sensing" does not guarantee that conflicts can be avoided. For example, if T_1 can hear T_2, then it can sense the carrier of T_2 and refrain from transmitting at a time that T_2 is transmitting. This avoids collisions with T_2. However such a strategy does not work in avoiding collisions with all potential interferers. Consider T_3. Suppose T_3 can be heard by R_1 but cannot be heard by T_1. Such a terminal T_3 is said to be a "hidden" terminal [13]. Even though T_3 can cause collisions at R_1, T_1 cannot detect when T_3 is transmitting and thus when collisions will occur.

This gives rise to the *Medium Access Control Problem*. How should nodes schedule their transmissions in order to have their packets received intelligibly by their intended receivers?

This problem has its genesis in ALOHA, a packet radio communication system set up in the 1970s to link the Hawaiian islands [14]. (Apparently, the goal of communication at that time was only to remote login; email was not envisaged!) The ALOHA protocol consisted simply of transmitting whenever there was a packet to send. If there was a collision, it was assumed that all nodes could detect it, which is the case if all transmissions can be heard by all nodes or if all transmissions can be heard by a common satellite. All nodes involved in the collision would then "back-off" for a random time and retry their transmission. Note that it is necessary for the retry interval to be random; otherwise if each node involved in the collision waits exactly the same time before retrying, then all nodes are destined to collide again. Calculations show that the maximum throughput of such a protocol over a common channel is only $\frac{1}{2e}$. This can be doubled by "slotting" time into intervals and synchronizing all transmissions to fit into these time slots. One adverse property of the ALOHA protocol is that if a very large number of users collide, then there are likely to be continued collisions in the future as nodes retry. This causes the throughput of successful transmissions to drop, leading to even more backlogged transmissions as more packets arrive to the system. Put simply, ALOHA is unstable when there is an unlimited number of users. Variants of ALOHA do exist today, the most-notable being the Ethernet protocol, where of course the number of users is bounded.

Ad-hoc networks however differ from the one common channel or satellite channel scenario, since not all nodes can hear all transmissions. Indeed, as we have seen there can exist "hidden" transmitters of whose existence one is not even aware. Our goal is to spatially reuse the radio frequency spectrum. That is, by restricting the range of a transmission, one can limit the interference of one's transmission, thus allowing a distant transmitter-receiver pair to carry on a conservation on the same frequency at the same time. One therefore needs to spatio-temporally schedule transmissions. One scheme which has been proposed is embodied in the IEEE 802.11 standard [15] (see also [16] and [17]). The protocol employs reservation packets to reserve the channel locally in space for data packets.

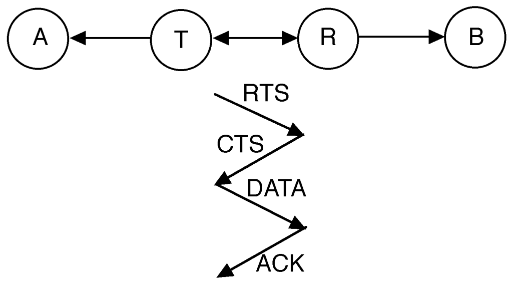

Figure 4: The nodes involved in an RTS-CTS-Data-ACK handshake in IEEE 802.11.

When a node T has a data packet to send to a node R, it first sends out a "request to send" packet, called RTS; see Figure 4. (Here we are assuming that the node T is not currently under an order to remain silent as described in the sequel). This transmission can be heard by all the neighbors of T, including R. Suppose, as in Figure 4, that T has one additional neighbor, A, and R has one additional neighbor, B. If A is not currently in range of any other transmissions, then it can hear T's RTS. It should then refrain from transmitting for a while. Similarly, if R is not currently in range of any other transmissions, then it too can hear T's RTS. Assuming that it is not currently under an order to remain silent, it sends a "clear-to-send" CTS packet to T. This CTS packet will be heard by B, assuming that it is not in range of any other transmission, and it should then refrain from transmitting for a while. Assume this to be the case. The CTS from R is heard by T since A was previously silenced, and does not therefore cause interference by transmitting. Then T sends its data packet to R. This is received successfully by R since its neighbor B has been silenced. Then R sends back an "acknowledgment" packet ACK to T. At that point, B is released from silence. Node A also is released from silence when it hears the end of T's data packet, after an obligatory pause to allow node T to receive the subsequent ACK from node R. In the event of a failure to do so for any reason (e.g., not receiving a CTS) then node T simply restarts the whole process after a random time.

This handshake needs to be done for each and every data packet on each of its hops.

Several issues have to be considered in evaluating this protocol: What is the overhead experienced by the RTS-CTS reservation process? What is the efficiency of this protocol? It should be noted that defining "efficiency" is not as easy as it may seem since it depends on the geographical location of all nodes. Indeed, due to spatial re-use of frequency, efficiency at one location cannot be defined independently of other locations. Moreover efficiency cannot be separated from "fairness." How equitably does this protocol allocate throughput when there are several contending sources? How does this fairness depend on geometry, i.e., the location of contenders and other nodes? How can one provide more throughput to users requesting more? What is the delay per hop? How does the protocol function in a multi-hop scenario where a receiver has to transmit (i.e., relay) every packet that it receives? How is the protocol influenced by channel errors? Note that if a node does not successfully receive a reservation packet (e.g., RTS or CTS) then it is liable to transmit at an inopportune time when it is supposed to have been silent, thus causing a collision. The answers to many of these questions are frequently based only on simulation.

It is worth noting that the central problem here is one of achieving *distributed coordination*. In order to communicate, the nodes need to coordinate their transmissions. However this coordination can itself take place only over the communication medium. So, in order to coordinate, the nodes need to communicate. Thus we have a vicious circle: Communication needs coordination which itself needs communication. Can we communicate somewhat more with perhaps less coordination packets? The following new protocol is a suggestion in this regard which requires further evaluation.

4.1.1 The SEEDEX Protocol

The key idea behind the proposed new SEEDEX protocol is for each node to publish a schedule of its intentions. Once all concerned nodes know of each other's schedules, they can find opportunities for transmission of their own packets.

There are several questions. The first is what sort of a schedule to publish. Our suggestion is to publish a pseudo-random schedule, i.e., a stochastic process. The second is how to publish an entire stochastic process. The key idea we will employ is to publish the initial *seed* for a pseudo-random number generator, since that will determine the entire future of the random schedule. The third question is how to exploit this schedule to find transmission opportunities for individual transmitter receiver pairs. Here we will simply choose the transmission times as the times when a node is certain that its intended receiver is certain to be listening, and then to transmit with a probability which reduces the likelihood of a collision. Finally, we design a mechanism to acknowledge packets so that transmitters become aware of packets which have undergone collisions and can therefore retransmit them.

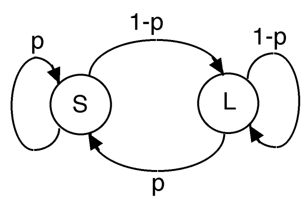

Figure 5: The state transition diagram of a node

We suppose that time has been divided into slots of equal length. In every slot, a node is in either a "listen" mode denoted by the state L, or a "possibly send" mode, denoted by the state S. Each node i chooses to be in state S with probability $p(i)$, or in state L with probability $(1-p(i))$, independently in each slot. Thus the dynamics of the state of each node can be modeled as a discrete-time, two-state Markov chain, with transition probabilities as shown in Figure 5. Each node implements this Markov chain by employing a pseudo-random number generator, which is a simple linear congruence relationship whose output produces numbers which appear to be random. If the initial state of the pseudo-random generator, known as the *seed*, is specified, then the entire sequence of pseudo-random numbers is predetermined. Thus once other nodes are made award of node i's seed, they know what state node i is in in each subsequent slot. Thus, in order to publish its entire future schedule, node i only needs to let others know its initial seed. Note that a node's packet can collide only with the packets of nodes in its 2-hop neighborhood (recall the notion of a "hidden" node in Section 4.1), and hence only such nodes need to communicate their seeds with each other.

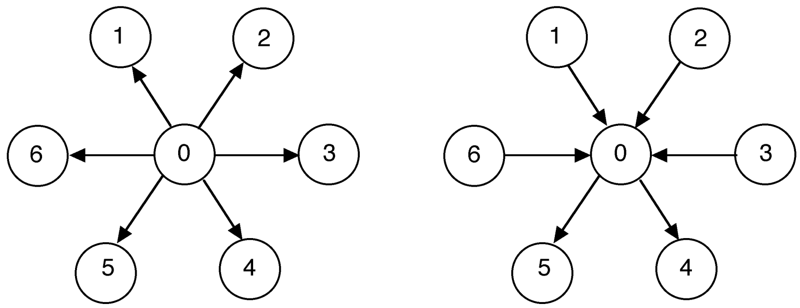

Figure 6: The two phases of the seed exchange procedure

This is done by a two-phase seed exchange procedure illustrated in Figure 6, from which SEEDEX derives its name. In the first phase, every node transmits its seed to each of its neighbors. This can be done by simple broadcast. In the second phase, each node broadcasts the seeds of all its neighbors to all of its neighbors. Thus, at the end of the two phases, every node finds out the seeds of all its neighbors, as well as the seeds of the neighbors of all its neighbors. That is, each node is made aware of the seeds of all nodes in a 2-hop neighborhood of itself. In the first phase shown in the left in Figure 6, node 0 broadcasts its seed to its six neighbors 1, 2, 3, 4, 5, and 6. Similarly, in the first phase, other nodes also broadcast their own seeds to their neighbors. Thus at the end of the first phase, all nodes know the seeds of their 1-hop neighbors. In the second phase, shown in the right in Figure 6, nodes 1, 2, 3, 4, 5, and 6 broadcast the list of seeds of their own neighbors to node 0. Thus at the end of this phase, node 0 knows the seeds of all its 2-hop neighbors.

We note that each phase will take several time slots. The reason is that all nodes cannot broadcast simultaneously. They may do so randomly, a la ALOHA. Also it is not necessary to broadcast the initial seed itself; they can broadcast the current seed, i.e., the current state of the pseudo-random number generator.

The next issue is to choose transmission opportunities. Suppose node T has a packet to send to node R. It then waits for a slot in which it is in state S and node R is in state L, guaranteeing that node R is in a listen mode. However there is another complication. There may be other neighbors of node R which are also in state S, and thus also liable to transmit, which would result in a collision. However, node T knows precisely how many such nodes there are due to the the seed exchange procedure; let this number be n. It then transmits on that slot with probability $\frac{1}{1+n}$.[1] The idea here is to reduce one's probability of transmission if there are many possible interferers, and to increase it if there are few or none. This is illustrated in Figure 7 where node T transmits with probability $\frac{1}{3}$. Note that if each of the other n neighbors of R actually had a packet to send to R, then they too would each transmit with probability $\frac{1}{1+n}$, which would lead to an expected value of only one transmission, which is what one aims for to avoid a collision. In that case, the throughput μ_T in packets per slot that node T obtains for its transmissions to R is

$$\mu_T \;=\; p(1-p_R)\sum_{i=0}^{F-1}\binom{F-1}{i}p^i(1-p)^{F-1-i}\left(\frac{1}{1+i}\right)\left(\frac{i}{1+i}\right)^i \tag{1}$$

assuming that all nodes other than the receiver have $p_i \equiv p$, and that F is the number of transmitting neighbors of R (i.e., "flows" through R). For $F = 3$, and $p_R = Fp$, this is maximized at $p^* = 0.118$ giving a throughput of about $\mu_T = 0.049$ packets/slot.

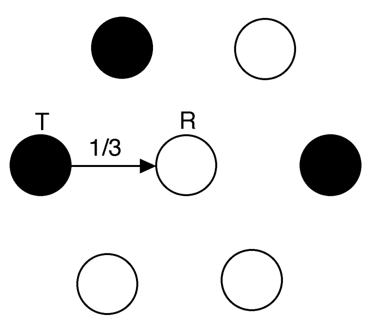

Figure 7: The transmission probability of the node T is $\frac{1}{3}$ since its intended receiver R has two other nodes in state S. All nodes in state S are shown shaded.

[1]This can be replaced by $\mathrm{Min}[\frac{\alpha}{1+n}, 1]$ where $\alpha > 0$ is some parameter to be optimized.

In general, however, the other neighbors of node R which are in state S may not have any packets to send, or if they do, they may want to send them not to R but to another of their neighbors. In that case, such a node will transmit with a different probability $\frac{1}{1+n'}$ where n' is the number of neighbors of its recipient R' which are in state S, assuming of course that R' is in state L. Hence the above calculation is only an approximation. This is illustrated in Figure 8. To assess the "efficiency" of this scheme is therefore more complicated.

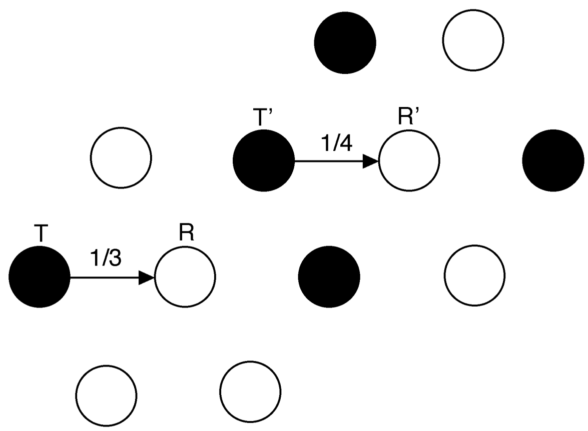

Figure 8: Node T transmits with probability $\frac{1}{3}$, while node T' transmits with probability $\frac{1}{4}$.

To roughly evaluate how good the performance is, consider the scenario shown in Figure 9, where there is one flow of packets from node 0 to node 1 to node 2 to node 3 to node 4. Suppose the flow achieves μ packets/slot. Note that when node 1 is receiving a packet (from node 0), node 2 cannot transmit. Similarly, when node 1 is transmitting to node 2, node 2 cannot transmit. Thus, out of 3 slots, node 2 can transmit in only one slot. Hence μ cannot exceed 1/3. If we normalize by this upper bound of $\frac{1}{3}$ we obtain a rough measure of "efficiency," $\frac{\mu}{1/3} = 3\mu$ (it is actually a lower bound on efficiency). Note that there is one more constraint that we have ignored; node 2 cannot receive a packet while node 3 is transmitting to node 4.

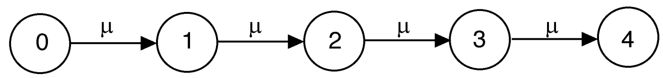

Figure 9: A flow of packets

To extend this to a situation where there is more than one flow through a node is again more complicated since there is some simultaneity in transmission opportunities which could be exploited. Consider the situation shown in Figure 10 where there are three flows through node 0. Then nodes 3 and 1 can receive at the same time while node 0 is silent. Nevertheless we will take 3×(Sum of throughputs through a node) as a measure of the efficiency of the scheme. By this measure, from (1), with $p^* = 0.118$, we see that the efficiency is about 44%.

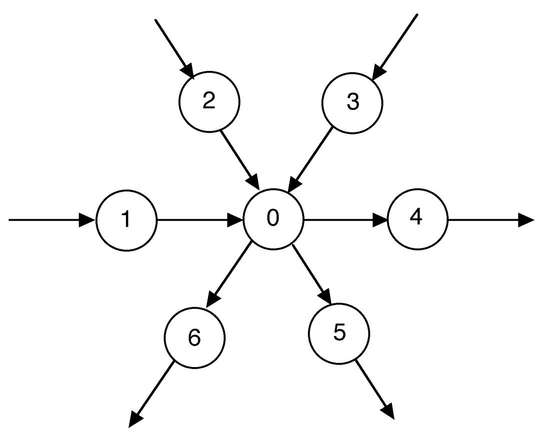

Figure 10: Many flows through a neighborhood

Let us now turn to the issue of Quality of Service. Consider a scenario where a node 0 has F transmitters $1, 2, \ldots, F$ who have packets to send to it. Let p_i be the probability that node i is in state S in a slot. Denote by μ_i the service rate in packets/slot that node i obtains. Then, similar to (1) we have

$$\mu_1 = p_1(1-p_0) \sum_{\substack{0 \leq k_2 \leq 1 \\ 0 \leq k_3 \leq 1 \\ \cdots \\ 0 \leq k_F \leq 1}} \left(\Pi_{i=2}^F p_i^{k_i}(1-p_i)^{1-k_i}\right) \left(\frac{1}{1+\sum_{i=2}^F k_i}\right) \left(1 - \frac{1}{1+\sum_{i=2}^F k_i}\right)^{\sum_{i=2}^F k_i}$$

$$
\begin{aligned}
&= p_1(1-p_0)E\left[\left(\frac{1}{1+\sum_{i=2}^F k_i}\right)\left(\frac{\sum_{i=2}^F k_i}{1+\sum_{i=2}^F k_i}\right)^{\sum_{i=2}^F k_i}\right] \text{(where the } k_i\text{'s are independent} \\
&\qquad \text{Bernoulli random variables with probability } p_i \text{ of being 1 and } (1-p_i) \text{ of being 0)} \\
&= p_1(1-p_0)E\left[\frac{1}{1+x}\left(\frac{x}{1+x}\right)^x\right] \text{ where } x := \sum_{i=2}^F k_i \\
&\geq p_1(1-p_0)\frac{1}{1+E(x)}\left(\frac{E(x)}{1+E(x)}\right)^{E(x)} \text{ (by Jensen's inequality since } \tfrac{1}{1+x}\left(\tfrac{x}{1+x}\right)^x \text{ is convex)} \\
&= p_1(1-p_0)\frac{1}{1+\sum_{i=2}^F p_i}\left(\frac{\sum_{i=2}^F p_i}{1+\sum_{i=2}^F p_i}\right)^{\sum_{i=2}^F p_i} \text{ (since } E(x) = \sum_{i=2}^F p_i) \\
&\geq \frac{p_1(1-p_0)}{1.4+e\sum_{i=2}^F p_i} \text{(since } \tfrac{1}{1+y}\left(\tfrac{y}{1+y}\right)^y \geq \tfrac{1}{1.4+ey}).
\end{aligned}
$$

Similarly, for $i = 2, \ldots, F$, also, we can obtain

$$
\mu_i \geq \frac{p_i(1-p_0)}{1.4+e\sum_{\substack{j=1 \\ j\neq i}}^F p_j}.
$$

Note that this provides a relationship between the parameter setting p_i and the service rate μ_i that is obtained for a flow. Also,

$$
3\sum_{i=1}^F \mu_i \geq \frac{3(1-p_0)\sum_{i=1}^F p_i}{1.4+e\sum_{j=1}^F p_j}.
$$

As argued earlier, we take the left hand side as a measure of efficiency.

Assume now that

$$
0 \leq p_i \leq \bar{p} < 1 \text{ for all } i \tag{2}
$$

and

$$
\sum_{i=1}^F p_i \geq \underline{P}. \tag{3}
$$

Then

$$
3\sum_{i=1}^F \mu_i \geq \frac{3(1-\bar{p})\underline{P}}{1.4+e\underline{P}}
$$

Table 1: Guaranteed Service Rate and Efficiency for $\bar{p} = 0.3$, $\underline{P} = 0.6$.

Number of Flows F	Lower Bound on $\frac{\mu_i}{p_i}$	Lower bound on $3\sum_{i=1}^{F} \mu_i$
2	0.32	0.42
3	0.23	0.42
4	0.18	0.42
5	0.15	0.42
6	0.13	0.42

as well as

$$\frac{\mu_i}{p_i} \geq \frac{1-\bar{p}}{1.4 + e(F-1)\bar{p}}. \tag{4}$$

This latter lower bound can be taken as a QoS guarantee on the guaranteed throughput, when the guidelines (2,3) are used to allocate the p_i's at each node. The numbers shown in the Table 1 for $F = 2, 3, 4, 5, 6$ are lower bounds only and the performance in general will be better.

The next issue that naturally arises is how much μ_i to "provide" to a requesting flow, and how to allocate its p_i. One scheme that suggests itself is that when setting up a flow, a leader packet traverses the route gathering information on how much service rate is available at each node. On the forward path it finds out where the bottleneck is and on the reverse path it requests the minimum of the service rates that can be provided at the nodes it has transited, and obtains an appropriate p_i at each node. There are several phenomena that need to be studied. Altering any p_i also alters the service rate that other already existing flows obtain. This causes a reverberating distributed adjustment mechanism whose equilibration is of interest.

It is useful to note that regulating the service rate obtained by a flow at each node it traverses opens up the possibility of congestion control on a hop-by-hop basis. If λ_i is the throughput of the flow, and μ_i is the service rate, then approximation by a $M|M|1$ queue would show that the number of backlogged packets has the probability distribution

$\pi(n) = (1 - \rho_i)\, \rho_i{}^n$ with a mean of $\frac{\rho_i}{1-\rho_i}$ and a standard deviation $\frac{\sqrt{\rho_i}}{1-\rho_i}$, where $\rho_i := \frac{\lambda_i}{\mu_i}$. This suggests strategies for cutting the flow when the buffer length is unduly large, and increasing it otherwise, which can be incorporated in a window based mechanism such as TCP.

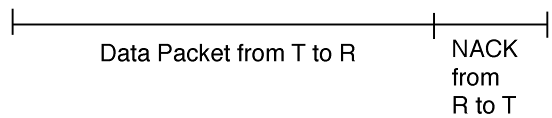

Figure 11: Structure of a slot

Let us also address the issue of how acknowledgments can be obtained. Consider the structure of a slot shown in Figure 11. The bulk of the slot is reserved for the data from T to R (as well as the header bits). However the tail end of the slot is reserved for a "negative acknowledgment" or NACK. If packets numbered $1, 2, 3, \ldots, k-1,\ k+1, \ldots$ have been received, but not k, then the receiver requests packet k. Note that this is done persistently until it obtains packet k, at which time it may switch to requesting the next missing packet. No protection against collisions is provided to this reverse channel. However, since it is persistent, it will eventually be heard. One also needs to calculate the throughput on the reverse channel to show that missing packets do not become the bottleneck for the forward link.

There are several other properties of the proposed protocol that may be noted. One is that a node with just a packet or two still gets opportunities to transmit even without reserving a flow. Such an opportunity occurs whenever all neighbors of the intended receiver are in state L as well as the receiver itself, guaranteeing an unhindered path for its packet. One other point is that the protocol is relatively unaffected by channel errors since these are simply treated as collisions which are retransmitted following NACKs.

5 The Routing Problem

In an ad-hoc network the address of a node is not indicative of its location or how to reach it. The address is simply a name. This gives rise to the routing problem [18] where the goal is determine the route to be followed by packets from their sources to given destinations.

Suppose that one wishes to find the path with the minimum number of hops from a given node 0 to a particular node i. This can be viewed as a dynamic programming problem. Let d_{ji} be the shortest path from node j to node i. For each node j let N_j denote the set of 1-hop neighbors of j. Assume, for simplicity, that all links are bi-directional, and that the network is connected. The d_{ji} uniquely satisfy the dynamic programming equation

$$\begin{aligned} d_{ii} &= 0, \\ d_{ji} &= 1 + \operatorname*{Min}_{k \in N_j} d_{ki} \text{ for } j \neq i. \end{aligned}$$

To solve this one can use the Bellman-Ford algorithm [19]. Let $d_{1k}^{(h)}$ be the shortest path from 1 to k with the constraint that the total number of hops is less than h. $d_{1k}^{(h)}$ is defined as $+\infty$ if there is no path from 1 to k with less than h hops. Define

$$d_{11}^{(h)} := 0 \text{ for all } h.$$

Initially, let

$$d_{1k}^{(0)} = +\infty \text{ for all } k \neq 1.$$

Then for $h = 1, 2, \ldots,$ let

$$d_{1k}^{(h+1)} = \operatorname*{Min}_{j \in N_k} \left[1 + d_{1j}^{(h)}\right]$$

for all $k \neq 1$.

Another algorithm is Dijkstra's algorithm. The procedure employed is to find the shortest paths to nodes in the order of increasing path length. Let X be the set of nodes to which we have obtained the shortest paths frim node 1. Set

$$X = \{1\}, \qquad d_{11} = 0, d_{1j} = 1 \text{ for all } j \in N_1, j \neq 1, \text{ and } d_{1j} = +\infty \text{ for all } j \not\in N_1 \cup \{1\}.$$

Then, at each step we perform the following operations:

(i) Find $i^* \in X^c$ such that

$$d_{1i^*} = \operatorname*{Min}_{j \in X} d_{1j}$$

(ii) Set $X = X \cup \{i^*\}$.

(iii) Set $d_{1j} = \text{Min}[d_{1j}, d_{1i^*} + 1]$ for all $j \in N_{i^*} \cap X^c$.

There are also distributed versions of these algorithms. For example, in the distributed asynchronous Bellman-Ford Algorithm [20], a node can broadcast its current estimate of its distance to its neighbors. Each neighbor then adds one to this distance (the additional one is for the added hop), and compares this quantity with its own current estimate of its shortest distance. If the new distance is lesser, then it adopts it, and also notes down the node through which packets should be sent.

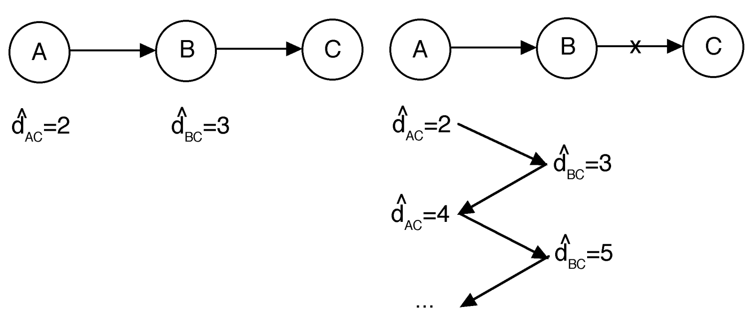

Figure 12: The counting to infinity problem

In a dynamic context, when nodes are mobile and links are volatile, problems can arise. One such problem is the "counting to infinity problem." Consider the scenario shown in Figure 12. In the beginning, there is a path from node A to node C via node B. The distances are $\hat{d}_{BC} = 1$ and $\hat{d}_{AC} = 2$, and both nodes A and B are aware of this. We use a "hat" to signify that these are merely estimates of the distances as perceived by the nodes. Now suppose the link from B to C is broken. Then node B notes that A had a path to C of two hops, and since A is its neighbor, it updates its estimated distance to C as $\hat{d}_{BC} = 3$. When this is passed on to A, it revises it distance to C as $\hat{d}_{AC} = 4$, and so on.

This illustrates the problem of maintaining routes which are loop free. Other issues of importance are to reduce the overhead in the amount of information transported just to maintain routes, and the time that routing algorithms take to reconverge to a solution when nodes move or links fail.

Routing protocols are broadly classified as either based on "link state" or "distance vector." In link state protocols each node broadcasts its complete view of the network to all nodes. Hence each node can form its own picture of the complete network. In distance vector protocols, each node only broadcasts its view of its distances to other nodes (as for example in the distributed Bellman-Ford algorithm). An early example of a distance vector protocol is [21], while a more recent one is [22].

Several protocols have been proposed for usage in ad-hoc networks. In the Dynamic Source Routing protocol [23], the sender explicitly lists the route of the packet in the header of each packet. When a route is needed, and it is not available in cache, then a route discovery procedure is used. A route request packet is repeatedly rebroadcast from node to node until it reaches its final destination, where the reversed route can be used to intimate to the source the sequence of nodes followed by the route discovery packet. Several modifications are made to maintain loop free routes, cache routes, etc. In the Ad-hoc On-demand Distance Vector Routing protocol [24], an attempt is made to preserve the advantages of distance vector routing while eliminating the overhead of global periodic routing advertisements. Several modifications are made to repair broken paths, and to maintain loop free routes. In the Temporally Ordered Routing Algorithm [25], shortest path routing is eschewed. The goal is to construct and maintain a set of directed acyclic graphs rooted at the destinations. Such graphs can provide a multiplicity of routes to a destination.

An idea that has been suggested in the System and Traffic Adaptive Routing Algorithm [26] is to use an adaptation algorithm to adapt routes to the network delays. Let $D^d_{sk}(t)$ denote an estimate made at time t of the mean delay from source s to destination d, if the packet is sent on its first hop to node k. In this algorithm, packets at source s destined for node d are sent on their first hop to node k with probability p^d_{sk}. These probabilities are adapted according to the following adaptation law:

$$p^d_{sk}(t) \quad = \quad p^d_{sk}(t-1) + \alpha(t)\left[D^d_s(t-1) - D^d_{sk}(t-1)\right]$$

where $D^d_s(t-1)$ is an estimate of the average delay,

$$D^d_s(t-1) \quad = \quad \sum_{k \in N_s} p^d_{sk}(t-1) D^d_{sk}(t-1).$$

The parameter $\alpha(t)$ is an adaptation step-size parameter. The adaptation update reduces the probability of forwarding a packet via a node k if the estimated delay of the route via k is larger than the average estimated delay, and increases it otherwise.

Thus, in equilibrium, the probabilities will be such that the mean delay via every utilized relay node k is the same as the average delay, and all the routes via the unutilized relay nodes k have a greater potential delay. Hence:

> Every path between nodes s and d which has a positive fraction of traffic traversing it has the same mean delay. All unutilized paths have larger mean delay.

This type of an equilibrium is apparently well known in transportation; it is called a Wardrop equilibrium [27].

As is clear from all the above, the central problem is to design an efficient adaptive distributed routing algorithm. It necessarily must be adaptive due to the volatile and mobile nature of the network, and it needs to be distributed since the nodes are neither spatially colocated nor informationally centralized. One needs to understand the time constants of adaptation and how they compare with the time constants of mobility, and also the trade-off between the value of maintaining routes and the overhead of doing so.

6 How Much Traffic Can Wireless Networks Carry?

Given the inherent shared nature of the wireless medium, what is the ultimate traffic carrying capacity of wireless networks? This will provide a design goal to strive for.

To answer this question one needs to model the nature of interference to receivers, or its converse; When is a packet successfully received by a receiver? Let us consider the following model. There are n nodes in a disk of area A square meters with each node capable of transmitting at W bits per second. Nodes can choose the range of each transmission. We will suppose that a transmission of range r creates a wireless footprint in a disk of radius $(1+\Delta)r$ centered around the transmitter, within which no receiver can successfully receive any other transmission, as shown in Figure 13. (For more details as well as other models centered around the receiver, or based on the signal-to-interference ratio, see [12]).

The performance measure we will analyze is the aggregate *bit-meters/second* that the network as a whole can transport. We call this the *transport capacity* of the network. When one bit has been transported one meter, we will say that the network has pumped one bit-meter. Counting all such transactions occurring in the network, how many bit-meters/second can the network pump?

For simplicity of exposition suppose that time is slotted into slots of length τ, and transmission is quantized into packets of length τ seconds which fit into slots. Consider a long

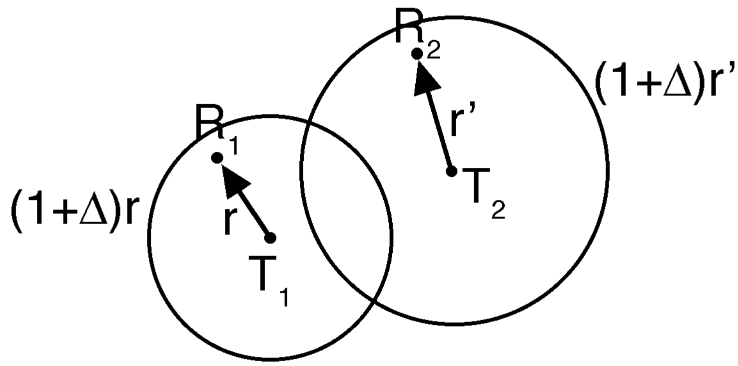

Figure 13: The interference footprint of transmissions

time interval of length T, and suppose that in this time interval a total of B bits have been transported from their sources to their destinations, in a multi-hop manner. Let us suppose that the average distance travelled by a bit is L meters; thus the bit-meters/sec pumped by the entire network in $\frac{BL}{T}$ bit-meters/sec.

We will show that

$$\frac{BL}{T} \leq \sqrt{\frac{8}{\pi}} \cdot \frac{W}{\Delta}\sqrt{An} \text{ bit-meters/sec.} \tag{5}$$

This provides an upper bound on the transport capacity of the network. In the right hand side, the linear scaling in W is obvious. If everyone doubles the capacity of their modems, then the number of bits is also doubled. Similarly, the square root scaling in the area A is also obvious. If the area is doubled, distances are scaled by the square root of two.

What is really interesting is the square root scaling in the number of nodes n. The implication of the scaling is this. If the transport capacity is divided equitably among all the nodes, then each individually obtains only $\frac{c}{\sqrt{n}}$ bit-meters/sec. Thus, as the number of nodes in the network increases, there is a square root drop off in what each node obtains.

Hence, networks should be designed either with few nodes, i.e., small n, or to support mainly nearest neighbor communications, i.e., small distances. Note that the distance to the nearest neighbor is of order $O(\frac{1}{\sqrt{n}})$ meters when there are n nodes in a domain.

We will now demonstrate the square root scaling, which follows from some fundamental constraints. From the picture in Figure 13, it is obvious that

$$\begin{aligned} |R_2 - T_1| &\geq (1+\Delta)|R_1 - T_1|, \\ |R_1 - T_2| &\geq (1+\Delta)|R_2 - T_2|. \end{aligned}$$

Hence, by the triangle inequality,

$$\begin{aligned} |R_2 - R_1| + |R_1 - T_1| &\geq (1+\Delta)|R_1 - T_1|, \\ |R_1 - R_2| + |R_2 - T_2| &\geq (1+\Delta)|R_2 - T_2|. \end{aligned}$$

Thus

$$\begin{aligned} |R_2 - R_1| &\geq \Delta|R_1 - T_1|, \\ |R_1 - R_2| &\geq \Delta|R_2 - T_2|. \end{aligned}$$

Hence

$$|R_2 - R_1| \geq \frac{\Delta}{2}(|R_1 - T_1| + |R_2 - T_2|).$$

This shows that disks of radius $\frac{\Delta}{2}$(Range) centered around receivers are disjoint. Thus, each transmission of range r consumes an area of $\frac{\pi r^2}{4\Delta^2}$. However, if a receiver is close to the boundary of the domain, then some of the consumed area can lie outside the domain, as shown in Figure 14. However at least a quarter, or $\frac{\pi r^2}{16\Delta^2}$, lies within the domain.

Now we simply list some of the constraints on the operation of the network. We need to do some bookkeeping. Let us suppose that the b-th bit makes $h(b)$ hops on its way from its source to its destination, with the h-th hop moving a distance of $r(b,h)$ meters. Note that each hop of a packet occupies one slot and carries τW bits.

Clearly

$$\sum_{b=1}^{B} \sum_{h=1}^{h(b)} r(b,h) \geq BL, \qquad (6)$$

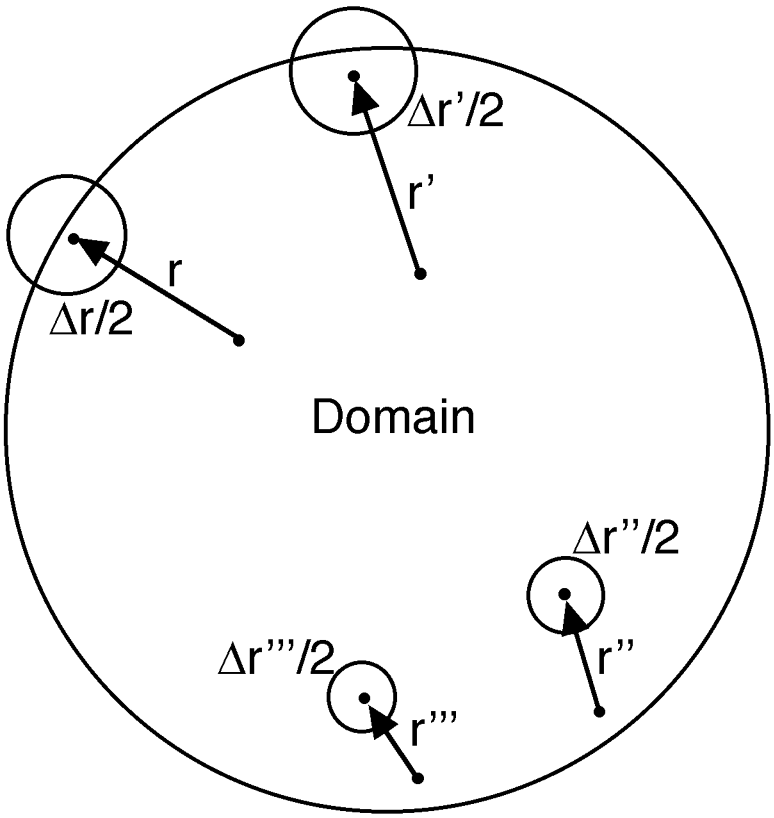

Figure 14: The exclusion zone around each receiver

since the total distance travelled by a bit can be no less than the the shortest path from its source to its destination.

If we count all the bits undergoing hops we obtain the bound

$$H \ := \ \sum_{b=1}^{B} h(b) \ \leq \ \frac{WTn}{2}, \tag{7}$$

since at any given time at most $\frac{n}{2}$ nodes can be transmitting, each at rate W bits/sec, and the total time duration is T seconds.

Now we invoke the crucial fact that space is a valuable resource in wireless networks. Since each transmission consumes a portion of the limited domain,

$$\sum_{b=1}^{B}\sum_{h=1}^{h(b)} \frac{\pi\Delta^2 r^2(b,h)}{16} \ \leq \ AWT.$$

The reason is that each packet of $W\tau$ bits with range r occupies $\frac{\pi\Delta^2 r^2}{16}$ sq. m. for the duration of one slot of τ seconds, there are a total of $\frac{T}{\tau}$ slots, and the total area of the domain is A square meters.

We write the above as

$$\frac{1}{H}\sum_{b=1}^{B}\sum_{h=1}^{h(b)} r^2(b,h) \quad \leq \quad \frac{16AWT}{\pi\Delta^2 H}.$$

Since r^2 is a convex function, we have

$$\left(\frac{1}{H}\sum_{b=1}^{B}\sum_{h=1}^{h(b)} r(b,h)\right)^2 \quad \leq \quad \frac{1}{H}\sum_{b=1}^{B}\sum_{h=1}^{h(b)} r^2(b,h).$$

Hence

$$\frac{1}{H}\sum_{b=1}^{B}\sum_{h=1}^{h(b)} r(b,h) \quad \leq \quad \sqrt{\frac{16AWT}{\pi\Delta^2 H}}$$

From (6) we obtain

$$\frac{BL}{T} \quad \leq \quad \sqrt{\frac{16AWH}{\pi\Delta^2 T}}.$$

The result (5) then follows from (7).

This upper bound on the capacity, of order $O(W\sqrt{An})$, can in fact be achieved, by arranging points in a regular grid; see [12].

Hence $\Theta(W\sqrt{An})$ bit-meters/sec is a sharp characterization of the transport capacity. However, to achieve it requires not only merely optimal *operation* of the network (i.e., optimal choice of transmission times, ranges, routes, etc.) but also optimal *location* of nodes, choice of traffic patterns, etc.

The last cannot be realized, of course. To address more common situations, a random scenario can be considered where the nodes are randomly located in a disk and each node randomly chooses a destination. Assuming that all nodes employ the same range, it can be shown that the throughput which can be provided to every node is $\Theta\left(\frac{W}{\sqrt{n\log n}}\right)$ bits/sec [12].

Since the factor $\frac{1}{\sqrt{\log n}}$ is not terribly large, it follows that random networks may be nearly best.

One can also envisage other models for successful reception. Essentially any such model's validity will depend on how receivers are constructed. Suppose that we construct receivers such that a packet is received successfully only when the signal-to-interference ratio exceeds some threshold. Assume also that the path loss of signals follows a $\frac{1}{r^2}$ law for $\alpha > 2$. Then it is shown in [12] that the transport capacity is lower bounded by $c\sqrt{n}$ and upper bounded by $cn^{\frac{\alpha-1}{\alpha}}$. For α close to 2, the case of inverse square-law loss, the bounds are nearly sharp.

One can imagine much more sophisticated receiver design which operates at or near Shannon capacity. Then, one will need to study not only the capacity but also the complexity of decoding and encoding. This remains an area for the future.

7 Concluding Remarks

In the coming decades the field of systems is headed for rapid change and greater technological vistas, a la the phase transition which occurred around 1960. Entire new technologies are within our purview. With the foundations laid over the past four decades, the opportunities are ours to grasp.

References

[1] S. Keshav, *An Engineering Approach to Computer Networking.* Reading, MA: Addison-Wesley, 1997.

[2] T. Cover and J. Thomas, *Elements of Information Theory.* New York: Wiley and Sons, 1991.

[3] Cynara C. Wu and Dimitri P. Bertsekas, "Distributed power control algorithms for wireless network," in *38th IEEE Conference on Decision and Control*, (Phonix, AZ), pp. 3556–3561, 1999.

[4] S. Papavassiliou and L. Tassiulas, "Improving the capacity in wireless networks through integrated channel base station and power assignment," *IEEE Trans. on Veh. Technology*, vol. 47, pp. 417–427, 1998.

[5] J. Rulnick and N. Bambos, "Mobile power management for wireless communication networks," *Wireless Networks*, vol. 3, pp. 3–14, 1997.

[6] R. Yates and C. Huang, "Integrated power control and base station assignment," *IEEE Trans. on Veh. Technology*, vol. 44, pp. 638–644, 1995.

[7] Van Jacobson, "Congestion avoidance and control," *ACM Computer Communication Review*, pp. 314–329, August 1988.

[8] Van Jacobson and M. J. Karels, "Congestion avoidance and control," in *SIGCOMM 1988*, August 1988.

[9] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 397–413, August 1993.

[10] W. R. Stevens, *TCP/IP Illustrated*. Addison-Wesley, 1996.

[11] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming* (W. M. McEneany, G. Yin, and Q. Zhang, eds.), Boston, MA: Birkhauser, 1998. ISBN 0-8176-4078-9.

[12] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. IT-46, pp. 388–404, Mach 2000.

[13] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II—the hidden terminal problem in carrier sense multiple-access models and the busy-tone solution," *IEEE Transactions in Communications*, vol. COM-23, no. 12, pp. 1417–1433, 1975.

[14] N. Abramson, "The aloha system—another alternative for computer communications," *Proc. Fall Joint Computer Conf. AFIPS Con.*, vol. 37, 1970.

[15] IEEE Computer Society LAN MAN Standards Committee, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications." IEEE Standard 802.11-1997. The Institute of Electrical and Electronics Engineers, New York, NY, 1997.

[16] P. Karn, "MACA: a new channel access method for packet radio," in *Proceedings of the 9th Computer Networking Conference*, pp. 134–140, September 1990.

[17] V. Bharghavan, A. Demers, S. Shenkar, and L. Zhang, "MACAW: a media access protocol for wireless LANs," in *Proceedings of the SIGCOMM '94 Conference on Communications Archtitectures, Protocols and Applications*, pp. 212–225, August 1994.

[18] S. Ramanathan and M. Steenstrup, "A survey of routing techniques for mobile communication networks," *Mobile Networks and Applications*, vol. 1, no. 2, pp. 89–104, 1996.

[19] D. Bertsekas and R. Gallager, *Data Networks: Upper Saddle River*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[20] D. P. Bertsekas and J. N. Tsisiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[21] J. Jubin and J. D. Tornow, "The DARPA packet radio network protocols," *Proceedings of the IEEE*, vol. 75, pp. 21–32, January 1987.

[22] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," *Computer Communications Review*, vol. 24, no. 4, pp. 234–244, 1994. (ACM SIGCOMM '94).

[23] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing* (T. Imielinski and H. Korth, eds.), Kluwer Academic Publishers, 1996.

[24] C. Perkins and E. Royer, "Ad-hoc on-demand distance vector routing," in *Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications*, (New Orleans, LA), February 1999.

[25] V. D. Park and M. S. Corson, "A highy adaptive distributed routing algorithm for mobile wireless networks," in *Proceedings of INFOCOM'97*, (Kobe, Japan), April 7-11, 1997.

[26] P. Gupta and P. R. Kumar, "A system and traffic dependent adaptive routing algorithm for ad hoc networks," in *Proceedings of the 36th IEEE Conference on Decision and Control*, (San Diego, CA), pp. 2375–2380, Dec. 1997.

[27] J. G. Wardrop, "Some theoretical aspects of road traffic research," in *Proceedings of the Inst. Civ. Eng., Part 2*, pp. 325–378, 1952.