Issues in Packet Radio Network Design

BARRY M. LEINER, SENIOR MEMBER, IEEE, DONALD L. NIELSON, MEMBER, IEEE, AND FOUAD A. TOBAGI, FELLOW, IEEE

Invited Paper

There are many design choices that must be made in the development of a packet radio network. There is usually no single correct choice, and the decisions are dependent on the environment that the network must work in, the requirements for performance and other functionalities, and the cost and other limitations. In addition, as new hardware and software technologies become available, the parameters governing the decisions change and often result in different selections.

This paper outlines a number of design issues and choices available. The intent is to provide an overview of the design decisions that must be made so as to provide a context for the decisions made in a number of existing and developing packet radio networks. It is hoped that this will allow future designers to take advantage of the wealth of experience available as well as new technologies. Three areas of design decisions are identified. The first area deals with the physical aspects of the network and concentrates on the radio connectivity and channel sharing. The second area deals with the automated management of the network and concentrates on issues such as link management and routing. The third area deals with the interface of the network to the users and some practical aspects of operating and maintaining a network.

I. INTRODUCTION

Packet radio networks represent the extension of packet switching technology into the environment of broadcast radio. They are intended to provide data communications to users located over a broad geographic region, where direct radio or wire connection between the source and destination users is not practical.

Packet radio networks have been and are being designed to operate in a number of environments using a number of different technologies [1]. There are packet radio networks making use of ground mobile radio (narrow-band 16 kbit/s [2] and more wide-band 400 kbit/s [3], [4]), amateur radio [5], HF for use in Navy applications [6], and satellites [7]. Yet all of these networks share some common characteristics. They are all based on the notion of packet switching applied to (usually broadcast) radio usually sharing a single channel. They are intended to handle mobile users, although some of the amateur and commercial applications de-emphasize this capability. They are for the most part based on a store-and-forward operation, although the simpler satellite networks involving use of a single satellite do not include store and forward operation.

Fig. 1 shows a typical packet radio network structure [8]. A packet radio unit consists of a radio, antenna, and digital controller. The radio provides connectivity to a number of neighboring radios, but typically is not in direct connectivity with all radios in the network. Thus the controller needs to provide for store-and-forward operation, relaying packets to accomplish connectivity between the originating and destination users.

There are a number of common issues involved in the design of these networks. These include efficient methods for sharing the common radio channel, methods for de
terming connectivity and using that connectivity to route data through the network, methods for achieving reliable communications in a typically noisy radio environment, and methods for managing and controlling the distributed network. Thus there are many design choices that must be made in the development of a packet radio network. There is usually no single correct choice, and the decisions are dependent on the environment that the network must work in, the requirements for performance and other functionalities, and the cost and other limitations. In addition, as new hardware and software technologies become available, the parameters governing the decisions change and often result in different selections.

This paper outlines a number of design issues and the various choices available. The intent is to provide an overview of the design decisions that must be made and to do so in the context of a number of existing and developing packet radio networks. It is hoped that this will allow future designs to take advantage of both the wealth of experience available as well as new technologies. Furthermore, as the rest of this Special Issue will provide details on a number of packet radio systems, this paper should provide a context for comparison of the various approaches taken in those systems.

Three areas of design decisions are identified. The first area deals with the physical aspects of the network and concentrates on the radio connectivity and channel sharing. The second area deals with the automated management of the network and concentrates on issues such as link management and routing. The third area deals with the interface of the network to the users and some practical aspects of operating and maintaining a network. The various issues are highlighted throughout the paper by showing them in raised or capitalized text. A compilation of these issues should help guide the design of a packet radio network.

Just as with any packet communication system, the functions to be performed by a packet radio network may be organized into a linear hierarchical structure, as defined in ISO's OSI Reference Model [9]. This structure consists of a layered architecture comprising a number of independent layers. This allows the discussion of design issues underlying a packet communication network to be done by focusing on one layer at a time. While there will always be coupling between the various layers in terms of design issues, requirements, etc., in the case of packet radio (as will be clear from the discussion below), the various layers are highly interdependent. Therefore, the task is not complete until a cross examination of design tradeoffs at all layers is performed. Nevertheless, to achieve an orderly presentation, the issues are presented one layer at a time.

II. PHYSICAL AND DATA LINK LAYERS

In this section, we focus on the first two layers of the ISO model. Issues to be addressed include physical connectivity, bandwidth-time-space management, channel access, and data link control.

A. Physical Connectivity

The physical layer in packet radio networks establishes digital link connectivity among nodes in the network so that information (data) paths may be established from traffic sources to their destinations. Link connectivity from a node A to another B refers to B's ability to correctly receive information transmitted by A at a specified minimum rate. Link connectivity clearly depends on radio propagation parameters, such as the radio frequency, the distance between nodes, the type of terrain, and the transmit power. Connectivity depends in addition on the data rate requirement, the channel RF bandwidth, and the data encoding and modulation schemes. Thus the design problem at this level can be formulated as follows. GIVEN THE TYPE OF TERRAIN IN WHICH THE NETWORK IS TO OPERATE, THE USERS' LOCATIONS, AND THEIR REQUIREMENTS IN TERMS OF TRAFFIC, MOBILITY, ANTIJAMMING CAPABILITIES, ETC., SELECT

A. THE RADIO FREQUENCY AND RF BANDWIDTH,
B. THE SIGNALING, ENCODING, AND MODULATION SCHEMES,
C. THE NETWORK TOPOLOGY.

Network topology refers to the radio nodes constituting the network, their roles (user interface or repeater), their density, their locations (fixed, mobile), the antenna design associated with each (its height, directionality, etc.), and their transmit powers.

The connectivity resulting from the design decisions may be represented as a graph in which vertices represent nodes, and edges represent link connectivity. Due to changes in the environment, nodal mobility, and other effects, variations in digital link connectivity may result, rendering the graph representing the network topology a probabilistic one.

Design decisions at the physical layer interact with those at higher layers. For example, propagation is typically better at lower frequencies (longer distances can be achieved with less sensitivity to terrain). On the other hand, data rates are lower at lower frequencies, and therefore, the network's ability to cope with mobility may be reduced. Thus if the user data requirements can be satisfied with relatively low data rates, a low frequency can be used, thereby possibly achieving full link connectivity and simplifying the network design. On the other hand, higher data rates would require a higher frequency/bandwidth and probably line-of-sight (LOS) propagation. This would result, though, in more capability to support the overhead of network algorithms to handle changing connectivity resulting from mobility (coupled with LOS propagation.)

Nevertheless, by initially ignoring effects due to higher level functions, an approximation to the design can be achieved, which then can be refined as the higher level design issues are resolved. For example, it is possible to estimate a range for the required link data rates given the users' traffic and delay requirements, independently of higher level protocols, and proceed with the design based on these estimates. A more precise determination of the required rates would then have to be done after an understanding of higher level issues has been acquired.

1) Choice of a Frequency Band: The first issue to be addressed is WHAT FREQUENCY BAND TO OPERATE IN? The tradeoffs between higher frequency/bandwidth LOS operation and lower frequency/bandwidth operation over extended distances, while present in the design of any radio system, have special implications in a packet radio network.
Not only does this tradeoff determine the degree of connectivity between the user nodes, but the choice of frequency may require additional functions to be supported.

For example, if a frequency with good propagation characteristics is chosen, sufficient connectivity may be achieved just using the radios at the user nodes. Higher frequencies, because of the limited propagation distances at LOS frequencies, may require "repeater" nodes be installed to achieve full network-wide connectivity. (A repeater node is a packet radio unit with no user directly connected.)

Another consideration in the choice of frequency is the available bandwidth. This choice is partially determined by the user data requirement as well as other factors such as the need for spread spectrum. However, the dynamics of the network topology, and therefore the required amount of control traffic, also help determine the required data bandwidth and therefore the minimum frequency band.

Networks where the connectivity is changing slowly (e.g., HF networks in a Navy environment [6] or fixed-site networks such as the amateur packet radio networks [5]) can afford to have limited control traffic. Ground mobile networks [3, 4], due to their rapidly changing connectivity, require considerably greater control traffic and therefore increased data bandwidth.

2) Propagation and Interference Considerations: In addition to frequency band, link connectivity also depends on other factors, including terrain, distance between nodes, transmitter power, antenna height and directionality, etc. Given the terrain, antenna parameters, and node locations, link connectivity is achieved by appropriate selection of a data-encoding scheme and transmitter power. When the radio channel is being shared by many nodes, multiuser interference and the near-far problem are introduced. Because these issues pertain more specifically to the problems of bandwidth-time-space allocation and channel access, discussion is deferred until the next section.

B. Bandwidth-Time-Space Management

Once a frequency band and RF bandwidth have been selected, the question remains as to HOW TO ALLOCATE THE BANDWIDTH IN TIME AND SPACE TO THE NODES IN THE NETWORK. Four techniques are available, all of which may coexist: frequency division, time division, code division, and spatial reutilization of the bandwidth resources. Frequency division refers to the partitioning of the bandwidth into separate radio channels, orthogonal in the frequency domain. Time division refers to the allocation of a given radio channel to different users at different times. Code division refers to the provision of orthogonal spread spectrum codes to different (radio channel) users, so that these may use the same channel simultaneously without interference. Finally, spatial reutilization of the bandwidth-time resource refers to the simultaneous use of a given portion of the channel bandwidth in different localities without causing interference. A few examples are presented here to illustrate some design considerations and tradeoffs.

The use of frequency division in order to provide channels for use by individual nodes (or pairs of nodes) may be adequate if the period of use for channels is long, and the channel utilization high; that is, if the users' demand is non-bursty and predictable. Otherwise, overall bandwidth efficiency will be low. Furthermore, such a mode requires management procedures to dynamically allocate these channels. From another, perhaps more important, point of view, frequency division may be useful in the provision of several channels, each of which is used for a different functionality, but is shared in the time or code domains by many users [8]. This constitutes a natural way of providing a hierarchical network. (In the HF band, for example, radio distance varies significantly with the channel frequency. A channel centered around a frequency at the low end of the band could be used for overall network connectivity and control purposes, while other channels operating at higher frequencies are used for communication among neighbors.) But in general, it eases network deployment and resource allocation and management to have all users tuned to the same channel frequency, and employ time and code divisions [2, 3]. The specific means by which a radio channel is shared in the time and code domains are discussed below under channel access and capture modes. Suffice it to say at this point that such schemes are devised so as to achieve efficient bandwidth utilization. Furthermore, if nodes employ omni-directional antennas, then local broadcast communication results, and mobile users are more easily supported without the need for complex network management procedures.

Spatial reutilization of the bandwidth-time resource achieves a higher overall utilization by allowing multiple transmissions to take place at the same time (and with the same code) in different geographic areas of the network. Such reutilization may either be the result of RF propagation characteristics, or intentionally designed for. Using the UHF band for ground mobile operation, for example, the radio range is inherently short, rendering spatial reuse a natural outcome. In this and other cases, reutilization may also be achieved by using directional antennas, or low transmission power.

It is not clear when to intentionally enforce spatial reuse. Consider, for example, the problem of using transmission power to control the range of nodes in a wide geographic area with a relatively dense population of nodes. If source-destination pairs are distant, then high transmit power leads to a smaller number of transmission hops, but a higher degree of interference with other nodes results, which in turn limit the network's throughput; a low transmit power leads to less interference, and thus a higher degree of spatial reuse of the bandwidth; but it then requires a larger number of transmissions to reach the destination, thus increasing channel load for the same user traffic, and thus perhaps again limiting the network's throughput [10].

C. Channel Access

Two mechanisms to share a single radio channel are time division and code division (in the case of spread-spectrum transmissions.) WHAT CHANNEL ACCESS PROTOCOL AND CODE ASSIGNMENT ALGORITHM SHOULD BE USED?

In sharing a channel, there is a need to deal with conflicts which result from contention. This is achieved either by a priori fixed assignment of the channel resources (time and codes) to different users so as to prevent contention, or by providing some dynamic channel access algorithm which defines when users are permitted to transmit based on
channel conditions and traffic demands. Although such algorithms may not prevent conflicts from occurring, the objective is to maximize the overall network throughput. As channel access protocols and their performance are closely related to channel signaling methods and the capture effects that result, we will discuss them for narrow-band systems and spread-spectrum systems separately.

1) Narrow-Band Systems: Aside from the limited effects of power and FM capture, the overlap of two or more packets at some receiver in narrow-band systems results in the destruction of all. We say in this case that the system operates under a zero-capture mode.

At first glance, the solution appears to be one which guarantees orthogonality in the time domain. Fixed TDMA, whereby time slots are permanently assigned to nodes (or pairs of nodes), suffers from the same limitation in efficiency as does FDMA when nodes' traffic is bursty. Dynamic allocation of time slots to match traffic requirement is more efficient, but then requires scheduling algorithms which tend to become complex in distributed environments [6].

The time sharing of a channel by users can also be achieved via random-access schemes [10]–[12]. The decision as to whether to transmit or not is entirely left to the nodes, and thus collisions may occur. If a transmission collides with others, then it is repeated at some later time. Random-access protocols are of basically two types, the ALOHA type and the carrier sense type. In the former, no knowledge of current activity in the network is required, and nodes act completely independently. The ALOHA scheme, which allows a node to transmit any time it wishes, is one such scheme [13]. In the carrier sense type, some knowledge of transmission activity in the network is acquired and used in the decision. For practically realizable protocols, the rules embodied in such protocols are constrained to be in terms of information that can be acquired locally at the node; typically, this consists of the transmission activity of neighboring nodes, which is acquired mostly via carrier sensing. Carrier Sense Multiple Access (CSMA), in which a node may transmit only if it does not sense any carrier (which otherwise would be due to neighbors transmitting), is one example [14].

While it may appear that CSMA would greatly outperform ALOHA, this is not always true and depends to a large extent on the network topology and traffic. In a fully connected network, in which the propagation delay is a small fraction of the transmission time of a packet, CSMA is significantly superior to ALOHA. But in a general multipath topology, the existence of hidden nodes (i.e., nodes within range of the intended destination but not of the transmitter) can drastically degrade the performance of CSMA [12]. Finally, we note that the implementation of CSMA also requires special hardware, and the ability to switch rapidly from the receive mode to the transmit mode in order to keep the efficiency of the scheme high.

The problem of collisions caused by hidden nodes can be alleviated by the use of a busy tone which is transmitted by a node on a separate channel to indicate that it is currently receiving a packet [13]. This activity-signaling channel requires additional bandwidth and hardware resources which increase the cost of the radios. Furthermore, deciding which node should transmit the busy tone and under what condition, and the possibility of blocking transmissions which would otherwise have succeeded without interfering with ongoing ones, complicate matters further.

The above mentioned schemes, and variants thereof which may prove appropriate for particular situations, offer tradeoffs and performance results that are not simple to assess. Performance analysis via mathematical modeling and simulation has been carried out for some schemes to some extent, leading to results which can be helpful in understanding their behavior [10], [12].

2) Spread-Spectrum Systems: SHOULD SPREAD-SPECTRUM CODING BE USED? The selection of a signaling method may be based on considerations other than digital connectivity or jamming. For example, spread spectrum may be used to combat multipath, or FH spread spectrum may be used to overcome the near-far problem. On the other hand, as will be clear from the discussion below, there may be advantages in using spread spectrum specifically for sharing a channel, because of the resulting reduction in multiuser interference. But in general, given that spread-spectrum coding requires wider bandwidth, it is not clear whether overall bandwidth efficiency is improved or not [16].

The main features which distinguish spread-spectrum systems from narrow-band systems are code-division and time-capture. Code-division refers to the fact that transmissions with orthogonal spreading codes may overlap in time with little or no effect on each other. Time-capture refers to the ability of a receiver to successfully receive a packet with a given code despite the presence of other time-overlapping transmissions with the same code. To simplify the discussion, no distinction is made between direct-sequence pseudo-noise (PN) modulation and frequency hopping (FH) as the means for achieving spread spectrum, and multipath is ignored. In either case, it is assumed that the transmission of packets is asynchronous, and hence there is a need to precede the transmission of the packet by that of a preamble, which receivers use to acquire bit and packet synchronization. Furthermore, depending on the need, some form of data encoding for error correction may be performed to recover from erased or incorrectly received symbols.

Once the form of spread-spectrum modulation is selected, the method for ASSIGNMENT OF CODES USED IN THE PREAMBLE AND DATA PORTIONS OF THE PACKET must be selected. (The data portion includes control headers for network layers and above, as well as user data.) Two basic alternatives exist for the spreading codes. The preamble may consist of a known code with strong autocorrelation properties, which is used throughout the network and which idle receivers constantly search for. In this case, it is reasonable to assume that the overlap at some receiver of preambles belonging to different packets would cause errors in the processing of all preambles, and no packet is then acquired. Otherwise, the preamble is correctly processed (assuming that the background noise level is not too high), and the packet is locked onto. This case is referred to as the space-homogeneous preamble code assignment.

An alternative is to use preamble codes which are specific to intended receivers, referred to as the receiver-directed preamble code assignment. This alternative results in reduced preamble interference, as fewer packets would share the same code. However, information regarding the assignment of codes to receivers must be disseminated...
throughout the network. Furthermore, as we will see later, broadcast reception (the ability of all neighbors of the transmitting unit to hear the transmissions) can be quite helpful in dissemination of information for routing and network management.

Codes must also be assigned to the data portion of the packet. One desirable characteristic for the code assignment is that, once a packet is locked onto by a receiver, other overlapping packets do not interfere with its correct reception. Here too, there are several alternatives. In the space-and-bit-homogeneous code assignment, all data bits are encoded with the same code. In this case, an overlapping packet would not interfere with a packet locked onto as long as its autocorrelation peaks do not coincide with those of the earlier packet; i.e., unless the bit periods of the overlapping packets are within a few chip times of each other, causing the correlation peaks to overlap. As with preamble code assignment, this kind of interference can be reduced if receiver-directed bit-homogeneous code assignment is used.

It is possible to almost totally eliminate interference from overlapping packets by using a bit-by-bit code changing method, and equipping the receiver with a programmable matched filter which follows the pattern as it varies from bit to bit. If the pattern is long enough so that it does not repeat itself during the transmission of the packet, and the packets arrive with at least a few chip times of relative delay, then no interference will ever take place. This truly approaches perfect capture. A similar level of capture can also be achieved by assigning orthogonal codes to nodes which the latter use to encode their packets when transmitting. We refer to this as the transmitter-directed code assignment. The preamble must contain information regarding the spreading waveform used, thus allowing the receiver to program its matched filter accordingly.

Given the form of modulation and code assignment, the **SELECTION OF A CHANNEL ACCESS PROTOCOL** must be made. Similar to that of narrow-band systems, we distinguish two types: the ALOHA type and the activity-sensing type. The ALOHA schemes are identical to those in narrow-band system (transmit as long as not already transmitting, nor locked onto a packet worth receiving). A protocol of the activity-sensing type, on the other hand, may or may not be feasible depending on the ability of a node to dynamically acquire knowledge regarding the state of other nodes. Consider CSMA for example. A node must be able to sense activity due to its neighbors. This is possible with space-and-bit homogeneous codes, and is done by observing the output of the matched filters corresponding to the desired waveforms. In transmitter-directed or receiver-directed bit-homogeneous code assignments, a node will have to possess a bank of filters matched to all possible codes used by the neighbors. While this is clearly possible, it is rather impractical. With bit-by-bit code changing, activity sensing is difficult to achieve. From another point of view, namely that of overall network performance, it is not definitely clear that CSMA actually provides any improvement, since spread spectrum already exhibits strong capture properties, and inhibiting transmissions may actually decrease network throughput.

In particular situations one may be able to devise simple but useful schemes. Consider, for example, the case of receiver-directed code assignment, and let the node wishing to transmit monitor the channel for transmissions using the code assigned to the intended destination. If activity is sensed, then it is likely that the intended destination is busy, locked onto a packet destined to it. The existence of hidden nodes, and the possibility that the intended receiver is free (not locked) despite the presence of activity, introduces complexity similarly to that of a narrow-band system.

In spread-spectrum systems, performance is often not the only issue. Because the use of spread spectrum is often driven by operational requirements, considerations such as security, feasibility of implementation, cost, etc., must also be taken into account.

### D. Data Link Control

Data link control pertains to the functions at the data link Layer which achieve reliable communications between adjacent nodes (i.e., nodes which are connected directly by a digital radio link). (The relation of link reliability to end-to-end reliability is discussed below when network management and routing are discussed.) As in any packet communication system where some degree of reliability across links is needed, acknowledgment mechanisms (ARQ) are typically used to notify a device of its success in the transmission of a packet. In packet radio networks, however, where the performance of a digital link is highly variable (due to radio propagation characteristics and user contention), and at times poor, acknowledgment procedures alone may not be sufficient and have to be augmented by forward error correction (FEC) coding. Indeed, if the likelihood of errors in a packet is high, then ARQ schemes would result in very low throughput, as most of the packets would be rejected. FEC would greatly improve the chances of correct reception. This is particularly important when spread spectrum is being used with a pseudo-random generation of the code on a bit-by-bit basis. The probabilities of generating two codes that have high correlation sometime during the packet (on two different simultaneous transmissions) is often significant, and therefore the probability of incurring at least a few errors during the packet can be very high. Forward error correction can be used to correct these few bit errors so that the effective packet probability of error can be driven down to an acceptable level for an ARQ scheme to be effective. The rate of coding for error correction, on the other hand, should not be too low, since the information throughput across the link would then be low.

Thus the primary issue in data link control is **HOW TO COMBINE FEC AND ARQ SO AS TO ACHIEVE AN ADEQUATE LEVEL OF LINK PERFORMANCE** in the highly variable conditions that are typical of packet radio network environments. We note here that the balance between the two techniques must vary from link to link, and dynamically in time to match the current conditions.

The second issue to be addressed at the data link control level is **HOW TO IMPLEMENT HOP-BY-HOP ACKNOWLEDGMENTS**. One alternative is to have the receiving node transmit explicit short acknowledgments consisting typically of only the header, since the header uniquely identifies the packet. Another alternative, due to the local broadcast property of a radio channel, allows the relaying of a packet by the next node to be the acknowledgment to the current node. This scheme is referred to as the echo of

1The reader may wish to contemplate the view of spread spectrum as a form of FEC.
passive acknowledgment scheme. Clearly, in this scheme, the acknowledgment at the last hop has to be explicit.

While echo acknowledgments might appear to save channel time, as compared to sending explicit acknowledgments, specific problems are associated with them. Consider, for example, what happens when nodes are implementing a single first-in-first-out (FIFO) transmit queue with only one outstanding packet awaiting acknowledgment from a neighboring node. Let A and B be two consecutive nodes on the path for some source-destination pair. Node A transmits the packet at the head of its queue to node B. When received by node B, this packet is put at the bottom of its transmit queue. Then, the echo acknowledgment awaited for by node A will not be received until node B services that packet, which is at the bottom of its queue. This problem is particularly severe if packets are traveling along a string of repeaters.

Another problem associated with echo acknowledgments is that, since they are as long as the original packet, they are more likely to be interfered with than would be the case with shorter explicit acknowledgment packets. This may cause severe degradation in performance, especially in highly congested regions, since the original node would be required to undertake additional transmissions beyond the first successful one, due to simply having missed the echo acknowledgment. Thirdly, echo acknowledgments cannot be used in spread-spectrum systems with receiver-dependent codes, since the next node would not be using the PN waveform corresponding to the original node. Explicit acknowledgments, on the other hand, seem to present benefits which would outweigh the presumably additional channel time that they would require. In addition to being less prone to interference, they can be given priority over regular packets, thus speeding up the freeing of the buffers.

III. NETWORK MANAGEMENT

Once the methods for getting data from one node to the next have been determined, the next set of issues pertain to the techniques for moving data through the network. Given the particular environment in which a network is to operate, the characteristics of the radio links, the capabilities of the digital processing, and the desired functional capabilities of the network, design choices must be made as to the various network management algorithms and techniques to be used. In this section, we discuss some of the choices that are available and why certain choices are more applicable for different applications.

For convenience, the discussion is broken into three broad areas: link determination and control, routing and packet forwarding, and other network management concerns such as monitoring the status of the various network nodes. Again, as is true for most areas of packet radio network design, it must be kept in mind that these areas are by no means independent and the actual design choices must be made as a coherent whole.

A. Link Determination and Control

In Section II, we discussed how data can be moved at a suitable level of reliability from one node to its adjacent node (the next node along the path to the destination.) However, there are many parameters associated with transmission and it is the responsibility of network management techniques to determine the logical link connectivity and to control the radio parameters to assure that this link connectivity is maintained and managed appropriately.

HOW SHOULD TWO PACKET RADIO UNITS DETERMINE THE EXISTENCE OF A LINK BETWEEN THEM AND PASS THAT INFORMATION TO THE NETWORK MANAGEMENT ALGORITHMS? The network management techniques (to be discussed below) rely on the ability to pass packets/data from one node to the next and to have knowledge about that packet-passing capability. Because of the dynamics in a mobile packet radio network, it is desirable for this determination to be done as quickly as possible. However, there is also the desire to minimize the number of logical connectivity changes due to transient conditions, such as noise or temporary connectivity outages (such as that which might occur when a mobile radio goes under a bridge.)

Radio connectivity must be determined by the two ends of the radio link (i.e., the two packet radio units which are connected). The information from each node can be collected at a central location where connectivity is then determined, or it can be determined by the nodes themselves through a cooperative mechanism, such as exchange of the number of transmitted and received packets. In either case, a decision must be made as to the nature of the information that will be used to determine the existence of a link.

One set of methods that can be used to determine the existence of a link is to directly use measurements made on the radio channel. These can include such measurements as signal strength, signal-to-noise ratio, and bit error rate. These measurements, made on a packet-by-packet basis and associated with the individual radio-to-radio link, can then be integrated over several packets to declare whether a particular link is useable or not.

It is relatively straightforward to include such measurements in a radio [17], [18]; however, they have the disadvantage of requiring suitable hardware to perform the measurements and therefore may drive up the cost of the radio units. Furthermore, some systems are based on the use of existing radios [2] and adding special hardware would be difficult and expensive.

To avoid the use of special hardware as well as make measurements that are directly related to the operation of the link in the network, direct observations of the link logical level can be made. This is commonly done by simply counting the percentage of packets that are received correctly over some period of time.

Early versions of the DARPA packet radio network [19] utilized this method of determining packet radio link quality. The disadvantage of this method is that it requires a measurement interval sufficiently long to obtain a reasonable estimate of correctly received packets. Thus recognition of a change in connectivity will be delayed until such a time interval has passed. This can result in limitations on the speed of tracking for mobile units (as they move out of connectivity with one neighboring unit and into the range of another.)

In addition to assessing link connectivity, network management algorithms must be concerned with connectivity control. As described in Section II, link connectivity at the physical level (and therefore the logical level) depends on several parameters including data rate, coding rate, and transmitted signal power. To the extent that these are variable, link connectivity assessment must account for the
variations and permit network management algorithms to exploit these choices. For example, even the earliest DARPA packet radio units had the capability to control transmitted power and data rate [3]. Data could be transmitted at either 100 or 400 kbits/s at a constant spread-spectrum spreading bandwidth. The result is that the lower data rate could be used when connectivity was poor (either due to increased propagation attenuation of multipath). Algorithms were developed to dynamically select between the two data rates on a per-transmission basis, and this markedly increased the performance of the network over using either data rate exclusively. On the other hand, because of the complex interactions between transmitted power and congestion in the network, satisfactory algorithms to dynamically control transmitted power were not available, partly because the necessary signal strength monitoring tools were not in the earlier radios.

WHAT IS THE CORRECT BALANCE OF LINK PARAMETERS, FORWARD ERROR CORRECTION, HOP-BY-HOP ERROR DETECTION AND RETRANSMISSION (ARQ), AND END-TO-END ARQ? At the link level, decisions have to be made as to the mechanism to obtain some level of reliability. This is due to the fact that a typical packet radio channel has a nonnegligible packet error and loss rate. Relying on only end-to-end mechanisms can result in an overly large number of retransmissions. Therefore, by paying some cost at the link level, overall network channel utilization and delay is improved. In the section above on data link control, the tradeoff between the various link parameters was discussed. In addition, there must be an interaction between network level routing algorithms (discussed below) and the control of the link parameters [20]. If link connectivity is lost, the network must determine whether it should try harder on that link (by, for example, increasing power or coding gain) or it should attempt to find a different route, thereby possibly suffering some delay and lost packets while the new route is determined.

WHEN DOES THE SET OF AVAILABLE LINKS CONSTITUTE AN ACCEPTABLE NETWORK? Once the available links are determined, the network management algorithms need to determine whether or not the links are sufficient for the network algorithms to proceed to constitute a network. For example, in order that the network be robust in coping with failing nodes and links, it may be desirable to consider only networks where there is a minimum of two (or more) neighbors for every node. Another example is that of partitioned networks. It is possible that the set of links does not form a connected graph (i.e., the radios are clustered with no connectivity between the clusters). This could be managed as two separate networks or as a single partitioned network. Finally, there is the issue of minimal supported traffic levels. If there is a minimal user requirement stated for the network, it might be desirable to accept only those combinations of links (and associated capacities) that will support that traffic environment. For example, if two clusters of radios have a high degree of traffic between them (by specification) and a low degree of connectivity, one must decide whether or not to permit degraded operation.

The issue here then is to determine the minimal acceptable network. At one extreme, one can take the "what is given is what you have" approach, and require that the network management algorithms be prepared to cope with the available digital link connectivity, regardless of what it is. The other extreme would involve a substantial amount of pre-deployment engineering and require that a certain minimal amount of that connectivity be supported. Most of the current approaches to packet radio networking have favored the former approach, recognizing that locations of radios are usually determined by other factors than radio connectivity (such as the user location and mission).

B. Routing and Packet Forwarding

The basic job of the network management algorithms is to allow data packets to be routed through the network in an efficient and reliable manner. This entails two basic tasks. The first is the establishment of routes through the network, and the second is the forwarding of packets along those routes.

At this point, we should note that many of the network management algorithms discussed here are used for other networks in addition to packet radio. However, the unique environment of packet radio, having to do with the unpredictable and changing topology coupled with the local broadcast capability of the radio channel, gives rise to a set of concerns in designing the network management strategies that is significantly different than in other networks (such as long-haul wire networks or local area networks).

HOW SHOULD ROUTES BE ESTABLISHED THROUGH A PACKET RADIO NETWORK BASED ON THE BASIC LINK CONNECTIVITY? A route is the set of links traversed by a packet as it proceeds from source packet radio unit to destination unit. To effectively utilize the available links, routes need to be determined. The choices to be made in this area fall into two areas: WHAT TYPE OF ROUTING AND ROUTING ALGORITHM SHOULD BE USED? HOW SHOULD THE ROUTING INFORMATION BE DISSEMINATED?

f) Type of Routing: Because the methods for network management depend heavily on the method for routing of packets, it is important that the type of routing be resolved early in the design of the network. The methods for routing packets fall into two basic categories. The first is flooding techniques, and the second is point-to-point methods.

Flooding methods involve transmitting the packet to every node in the network. No attempt is made to store routes. Rather, nodes keep track of individual packets as they pass through and decide whether or not to retransmit (usually based on whether they have seen the packet previously). The utility of flooding techniques in packet radio networks arises from their utilization of the inherently broadcast nature of the radio channel. The main advantage of flooding techniques is that they usually involve little explicit overhead and require little network management. They are also well suited to distributed control as many such methods do not require any central control at all. On the other hand, flooding methods tend to utilize the network inefficiently, as every node in the network will receive every packet at least once.

Thus flooding methods tend to be well suited to applications where there is a high need for reliable delivery in the presence of uncertain connectivity and when the connectivity is changing so rapidly that it is difficult for routing information to be determined and disseminated throughout the network in a consistent manner. Flooding methods therefore have potential application in two areas of packet
radio network management. The first is for environments where connectivity is changing extremely rapidly so it is inefficient or impossible (given the delays in the network) to track changes in connectivity. The second application is in the area of network control itself. Because flooding techniques do not require a priori knowledge of the network connectivity, they are easily used for disseminating network management and control information which is used to determine that connectivity.

Point-to-point routing methods typically involve the association of a route (a sequence of links) with a source-destination pair. One method of doing point-to-point routing is to explicitly associate information in each node with a source-destination pair (connection). Typically such techniques involve a route establishment phase that occurs when the "connection" is first recognized, and then the information stored at each node is used to perform the actual routing of the packets. Forwarding of packets then simply involves looking up the appropriate forwarding information based on the connection identifier (which is carried in the packet). If topology changes occur, a new route establishment (or re-establishment) phase would occur to ensure that the correct information is stored at all the nodes in the intended route.

Connectionless approaches to routing typically involve the use of routing techniques that take place as a background activity and do not require an explicit route establishment at the time the end-to-end connection (source destination pair) first has traffic. The individual nodes in the network have no knowledge of the existence of an end-to-end connection, and operate based on information contained in the packet network header (such as the destination address and type of service) and information about the network topology that results from the background operation of the network. Thus as topology changes occur, the background activity would cause the nodal information to be updated without regard to any end-to-end connection, and the traffic would keep flowing (except for some possible delays while topology information is out of date.)

The choice of the routing method used depends heavily on the nature of the traffic pattern and the dynamics of the network topology. Connection oriented approaches have the advantage of requiring minimal information in the packet itself (basically just a connection identifier and sequence number) and so lead to better utilization on the channel. Since channel utilization in a radio environment is always an important consideration, such approaches can be attractive.

However, in networks where topology changes rapidly, routing strategies that lead to local adaptive behavior are preferable to connection oriented approaches, which often require re-establishment of the end-to-end connection when any change occurs in the network topology. Connectionless approaches coupled with distributed routing techniques can often deal with topology changes in a way that maintain the end-to-end service.

Thus we see that all three routing methods have a place in packet radio networks. In relatively static networks, it is often most efficient to have the nodes determine their connectivity, and then determine relatively fixed routes (which would then be modified if connectivity changed due to mobility, etc.). For more dynamic networks, where connectivity is constantly changing, higher channel efficiency can be achieved by reducing the connection setups and the associated overhead. Finally, in the most dynamic networks, where network delays preclude tracking of connectivity on any but the most local basis, flooding techniques would appear to be a reasonable approach.

2) Spreading Routing Information: HOW SHOULD THE INFORMATION THAT EACH NODE Requires to ROUTE PACKETS BE DISSEMINATED TO THOSE NODES? For any type of routing method (with the exception of the most simple flooding methods), the local connectivity information must be processed and made available to the nodes so that they may route the packets. Note that this is somewhat independent of the type of routing being used. However, it does depend on the method for determining link connectivity and in particular, where the resulting connectivity information resides.

A popular method for doing routing in networks where functional distribution is not needed (e.g., for survivability) is to use a centralized routing server. (This, in fact, was the method used in the early DARPA packet radio network [3].) This technique has each node send its local connectivity information to a central location. At this location, routes are determined and the information required by each node to process and forward packets (such as the next node along the route) is sent to the individual network nodes on either a request basis or as a background operation which constantly updates tables in the nodes.

Use of a centralized routing server has several advantages over more distributed techniques. Because the server has all the connectivity information available (albeit not necessarily current), it can be quite efficient in the computation of routes. This can be a significant advantage in packet radio situations where both connectivity and congestion are more visible globally and where some nodes are typically collocated with mobile users as opposed to being located in some predetermined location. The centralized techniques can generally be extended to a small number of servers for load-sharing and/or backup, thus overcoming some of the problems of size and robustness inherent in a centralized method.

Perhaps the major disadvantage of a centralized technique is its limited ability to handle the rapid local topology changes that often are typical in a packet radio network. Because the connectivity information has to travel to the centralized server, and then the resulting routing information has to be disseminated to the required nodes, centralized methods are inherently limited in their ability to deal with rapid changes in topology. Distributed techniques can, if so designed, often deal with such changes on a local basis.

One method for distributing the routing process is to provide enough information to each node so that each node can simply compute for itself the best total route and then take action locally that is commensurate with that global optimum. For example, based on the computed best total route, a node may determine which is the best node to forward the packet. At the next node, the route may be recomputed or the entire route (or portion) could be included in the packet. (The latter is considerably less robust in the

---

"Sometimes use of the connection might cause the information to be updated more rapidly as a side effect."
face of changing topology.) This form of distributed routing can be accomplished by having each node transmit its local connectivity information explicitly to every other node. Typically a form of flooding is used to disseminate the information.

The method is quite robust (except for errors in tables or transmissions) and, in fact, is the (new) algorithm used in the Arpanet [21] and is planned for use in the gateways of the DARPA Internet system [22], [23]. However, if the network has a relatively high rate of topology changes, the amount of traffic on the network could be very high, as every substantial topology change can produce a number of packets roughly equal to the number of nodes in the network times the number of nodes directly affected by the change. Thus this method of routing is well-suited to a network like the Arpanet or a packet radio network consisting of fixed locations where topology changes are infrequent.

Another interesting routing structure occurs when packet radio networks are hierarchically organized. If the network is assumed to consist of clusters of packet radios that are interconnected, the topology between clusters is likely to change at a slower rate than that between radios, and therefore hierarchical techniques may be applicable. We see this applied to packet radio in [24] and [6].

An even more distributed method relies on each node only knowing information relative to local routing decisions. One method for accomplishing this is for each node to inform its neighbors (and only its neighbors) about the current state of its routing table (the table associating destinations or connections with the next node to be used). If that table contains an additive routing metric (such as number of hops to the destination), a neighboring node can determine, based on the contents of the table, the metric for its own table and the next node it should use to route the packet.

Such a routing technique is inherently well-suited to deal with the rapid topology changes that can occur in a packet radio network. However, explicit mechanisms may be necessary to deal with robustness issues (such as route loops). This, in fact, was the algorithm used originally in the Arpanet [25] and it was discarded because, in the Arpanet environment where local broadcast is not convenient, it was difficult to avoid some of the problems. The packet radio environment, where the radio channel affords easy broadcast, is more suitable for this algorithm. A more detailed discussion of the above two techniques and their tradeoffs is contained in [26].

Analyses of the tradeoffs between the various routing strategies have indicated considerable sensitivity to the particular assumptions about topology and topological changes. In addition, different routing techniques may be preferable at different levels in the network organization. For example, the current DARPA packet radio algorithms use the last technique described above to do routing inside clusters of radios, and uses a multiple routing server concept to make routing decisions for routing between clusters [24].

3) Packet Forwarding Issues: Once routes are established, packets are then forwarded from node to node. Packet forwarding techniques are intimately tied to the selection of the routing establishment mechanisms and type of routing. In particular, the selection of the routing mechanism in large part governs the method for forwarding of packets through the network. However, there are a number of issues that need to be dealt with explicitly.

WHAT INFORMATION SHOULD BE PASSED FROM NODE TO NODE IN THE PROCESS OF FORWARDING A PACKET? In addition to the user data, a considerable amount of information associated with network management and control flows through the network. Much of this information is associated with the user data packets. For example, in a connection oriented network, each node must retain a pairing of the virtual circuit identifier and the next node, and each packet must contain the virtual circuit identifier to permit routing to take place. For connectionless routing methods, an indication must be given in the packet of the intended destination. Most if not all routing methods need some unique packet identifier so that duplicates can be identified and eliminated. To ensure valid data, some form of error control information (such as a checksum) must be included in the packet. Often some indication of special requirements for the particular packet must be forwarded (such as priority and delay).

Although the tradeoffs seem simple at the surface (virtual circuit methods require less overhead on a per packet basis than connectionless techniques), the cost per information packet must be balanced against the overhead cost in the network and the resulting functionality. In mobile packet radio environments where topologies are changing rapidly, it is often more effective and less wasteful of overall bandwidth to carry more information in the information packets themselves and have less out-of-band control information flowing.

HOW SHOULD ALTERNATE STRATEGIES BE USED LOCALLY TO TRY TO RAPIDLY CORRECT FOR LOCALIZED TOPOLOGY CHANGES? This issue demonstrates the tight coupling between the various algorithms operating in the dynamic packet radio environment. Most routing strategies are designed to work primarily with a relatively fixed route. For example, in a virtual circuit style network, each node has information telling it what next node should receive each packet on a particular virtual circuit. In a connectionless network, a node might have the information to tell it which should be the next node for a particular destination node. However, suppose the connectivity fails between the node and the desired next node. Since this is local information, it is likely that the node will discover this well before global routing tables are updated. Furthermore, it is most likely to discover this while in the process of trying to forward a packet. Therefore there is a possibility of trying to make a local correction to the route based only on local information.

As an example, a strategy that could be used is the following [19]. When a node A discovers it cannot reach the desired next node B, it sends the user data out in a special packet. This packet is marked “Any node which has connectivity to node B, please forward these data to B.” Thus if a localized rerouting can fix the route, the user data packet can still be delivered in the interim period while the global routing is being repaired. Alternately, no special strategy could be used to reroute the packet, and the node which cannot forward the packet successfully simply would notify the source node as to route failure. The latter has the advantage of being less complex (since it is desirable to fix the global route even if local rerouting can get a packet through).
4) Summary of Routing and Forwarding: As we have seen, there are a large number of tradeoffs involved in the design of the routing and packet forwarding algorithms. Underlying these tradeoffs is a major overall issue; the tradeoff between overhead and responsiveness to changes. Most of the difficulties in the design of the routing algorithms in particular arise from trying to deal with changing topologies, both changes in connectivity and node availability. This is inherent in the packet radio network environment. Certain information has to flow in the network in order to track these changes and respond to them. Yet, the goal of the network is to carry user traffic, not control traffic. Thus the challenge is to balance the need to minimize overhead against the need to track changes in the particular environment of interest.

C. Other Network Management Issues

The above issues dealt with network management and control directly related to the routing of packets. There are a number of other issues that must be dealt with in the management of a network, particularly one with the degree of dynamics associated with a packet radio network.

1) Congestion and Flow Control: HOW CAN TRAFFIC BE LIMITED AT ENTRY TO THE NETWORK AND WITHIN THE NETWORK SO THAT NETWORK CONGESTION IS CONTROLLED? Congestion and flow control are difficult issues to deal with in most networks. The dynamics of packet radio combined with the channel sharing provides additional challenges. Because the topology of a packet radio network is constantly changing, it is very difficult to do “traffic engineering” on the network to assure that it is capable of supporting the network traffic load at all times, even when the traffic originating at each node is limited to a predetermined value (unless, of course, that value is set way below the network capability). Even determining the capacity of a network (the maximum total throughput that a network can pass) given the topology and traffic patterns is difficult [12].

Virtual circuit techniques lead to somewhat direct flow control techniques, as resource reservation can take place in the process of setting up the route. Therefore, if allocated resources are never allowed to exceed that available, there is some assurance that network resources will not be overtaken. Connectionless and distributed approaches have more technical challenges here. One approach is to detect the presence of increasing congestion (by detecting delay in packet forwarding) and delay packet forwarding based on such detected congestion [27]. Thus the delay tends to propagate through the network back towards the traffic sources and reduce the traffic through the network. This technique is particularly useful in packet radio because of the local broadcast property, and therefore the ability of neighboring radios to detect the current state of packet forwarding delay (particularly if the unit’s delay parameter is included in the packet header). Considerable research is still needed in this area, though.

2) Management of Supported Users: In addition to management and control of the internal network units, it is also necessary to manage the interface to the attached user devices.

HOW CAN THE ASSOCIATION OF USER DEVICE TO PACKET RADIO UNIT BE MAINTAINED THROUGHOUT THE NETWORK? When a packet arrives from a user device (or gateway to another network), it typically is marked with the desired destination device. Thus an association must be made between that destination device and the actual packet radio unit to which the device is attached. With this it is a problem in all networks, the dynamics of packet radio networks plus their typical operating environment makes this problem particularly severe. As nodes fail, users must have the capability of connecting their devices to replacement radios, and that implies a dynamic name to address association.

One way to do this is to use a static association that can be changed through a manual process. That is, at installation time, the user device is associated with the packet radio unit, and that information is made known to all interested user devices (or a centralized table storing such information). If the device has to be moved to a different packet radio unit (because the unit failed, for example), the process is repeated. The obvious disadvantage to such an approach is the delay in propagating the information to all user devices and, therefore, delay before the user devices are able to communicate.

A preferred approach is to form the association dynamically. When a user device is attached (or detached) from a packet radio unit, the unit detects the fact, determines the user device identification, and passes that information on to either a centralized server or other units using a distributed algorithm [4]. Thus within minutes (or seconds) of moving a user device to another packet radio unit, communications are again possible.

IV. THE OPERATION AND MANAGEMENT OF A PACKET RADIO NETWORK

In designing a packet radio network, compatible operation within the data transport and electromagnetic environments must be assured. Some of the issues to achieve this involve design options important to the network users. Some are imposed by constraints such as the radio spectrum and still other issues stem from the ease with which the network is to be operated and maintained.

The data transport environment comprises the various interconnections that join the subject packet radio network to other networks with which connectivity is desired. These collective interconnections are often loosely referred to as the internetwork community or simply the Internet. The electromagnetic environment consists of other radio and noise emissions that might adversely impact the packet radio network, and likewise the way in which the packet radio network may have impact on other radio systems.

The perspective in this section is usually that of the network implementor or operator. Occasionally, the role of the user will be examined and some viewpoints will be those of the public interest or the concern of other parties who might be impacted by the operation of such a network.

A. Network Deployment and Maintenance

Critical aspects of packet radio network operation are deployment and maintenance. Deployment is the process of defining the initial topology in a way that meets coverage and capacity requirements, an especially important aspect of mobile operation. Maintenance, as with any distributed
1) Network Deployment: The ease of deployment of a packet radio network is more critical in a military context than a domestic one (although national emergencies often generate requirements similar to those of military deployments). In a military system where deployments may have to occur quickly, some of the resources in each node are devoted to siting aids. But some deployment parameters are common to any system; for example, the level of homing (the number of nodes within radio contact of a given node) and the area covered by each node.

Selecting these parameters constitutes a tradeoff between adequate coverage, redundancy, the spectral conflict imposed by the density of nodes, and, of course, cost. So, an early design issue in deployment is WHAT IS THE APPROPRIATE LEVEL OF HOMING?

Another important aspect of deployment is the degree of automation employed. This can range from the use of built-in siting aids to assist in manual deployment to the total automation of deployment with the DARPA packet radio system [4]. A very convenient siting aid is an output (or input) from the radio that, when the radio is being situated, enumerates the nodes that are its neighbors. In this way, the degree of connectivity can be gauged at the time of network deployment. So another deployment issue is the DEGREE OF AUTOMATION NEEDED IN THE NETWORK DEPLOYMENT PROCESS.

For example, how should new software be disseminated to the packet radio units? To avoid having to physically contact each unit to upload new software, it is desirable to support downline loading of new software over the network. This is particularly important in a network configuration having unattended repeaters (to overcome obstacles such as mountains). There are several methods for doing this. One is to have a "software distribution server" deliver new software to auxiliary memory in each unit via a network connection. It is assumed that sufficient software would exist in the unit's ROM (read only memory) to permit network software delivery.

A more distributed approach is to allow units to load software from a neighboring unit. The latter approach has the advantage of not requiring any overt action when a new node appears (for example, after being out of radio connectivity). It is more difficult in this case, though, to deal with software integrity and related issues. In either case, assurance has to be obtained that the nodes are running the most current software. This is usually done by using a version number embedded in the software. This version number also can permit nodes to recognize that a neighboring node is running a more recent version of software and to request a downline load of the new version.

HOW SHOULD THE VARIOUS PARAMETERS IN A PACKET RADIO UNIT BE SET AND CHANGED? Some parameters in the network are of primarily local concern (such as transmitted power and coding rate) and, therefore, can be set in coordination with the other local radio units. Other parameters, such as frequency, have to be set in coordination with the entire network. Again, a centralized approach can be used with a network control center making any changes from the default settings. Distributed approaches are more difficult, but may lead to increased network functionality.

2) Design Issues Relevant to Maintenance: The topic of maintenance will only be addressed in a limited way here and aspects that cover the design of hardware to make it intrinsically reliable will not be mentioned at all. Two novel approaches that rely on other network resources to assist in the maintenance process are briefly mentioned.

One design characteristic that leads to greater ease of maintenance is the use of common hardware and software in each network node. This assumes that the required range of network functions can be embodied in a single piece of nodal hardware and that the concomitant software differs only in its execution and not in the line-by-line comparison of code. Having this capability not only eases the practice of repair-by-replacement, but also opens the way for internodal network assistance—common resident automatic error detection and cross-net downline loading from a neighbor node. The relevant issue, then, is WHAT ARE THE PERFORMANCE AND COST INEFFICIENCIES OF USING COMMON SOFTWARE AND HARDWARE IN EACH NODE?

The notion of downline loading to repair a software fault or an intermittent or temporary hardware error was mentioned above. This "repair" can be effected automatically or by manual intervention. As in the initialization phase, attention has to be paid to the impact that such downline loading has on normal network traffic. However, the use of such techniques adds to the resiliency of the nodal operation and to the ease of network maintenance.

A network that has been given a wide range of functionally, including adapting to the loss or gain of nodes, complements the hardware in its reliability role. Thus the designer must decide HOW TO ACHIEVE THE HIGHEST LEVEL OF OVERALL NETWORK RELIABILITY FOR A GIVEN COST BY TRADING OFF HARDWARE RELIABILITY AND THE USE OF NETWORKING FEATURES SUCH AS DISTRIBUTED FUNCTIONALITY AND REDUNDANCY?

Evident by this time is that the networking in packet radio systems is not limited to simple store-and-forward transport. Networking defines a more complete, collaborative role among the nodes to accomplish a variety of goals. Obviously, the above issue is a central one. The ability to diagnose faults in a given or neighbor node and to repair such problems or take remedial action enhances the reliability of the system as long as the network does not get "captured" by such internal servicing. Clearly, along with the intelligence to recognize such faults and to attempt repair, comes also the intelligence to devote only so much resource to the task, cauterizing the problem after a specified level of effort.

Redundancy-of-coverage (multiple homing) enhances reliability regardless of the reason for nodal failure. Normally such redundancy mitigates against temporary propagation outages but any nodal failure is compensated for. Too much redundancy in a broadcast system, of course, leads to inordinate collisions and network inefficiency [10].

3) Diagnostics and Monitoring: Diagnostics and monitoring of operation are necessary to the successful operation of a packet radio network. Both functions constitute measurements and can be active, as in the case of probes, or passive, as in traffic monitoring. Both have their impact on network performance through the use of processor time in the switches and the use of air time in the transmitting of probe or reporting packets. The issue is, therefore, concern the number of measurements needed, how frequently they are made, and the degree of their passivity.

Monitoring can conveniently be divided into character-
izing the operation of individual nodes (switches), and depicting the performance of the nodes collectively. In the first case, functions like buffer occupancy, processing delays, and node throughput are important. In the second case, functions such as routing and the components of network delay are obtained. Monitoring of this type is used in routing and congestion control and was addressed earlier.

But that same monitoring is also important to network operation and maintenance. Short-term problems at a node are dealt with by network management methods such as the temporary halting of routes through a congested node. Longer term congestion or reliability problems must naturally fall out of the monitoring process to be able to invoke the correct actions of repair and restoration.

**How Should Nodes Detect Whether They and Their Neighbors are Operational and Who Should Be Notified in Case of Failure?** This is a particularly tricky problem in a mobile packet radio network. A unit can certainly run self-diagnostics to determine its own status. The problem comes in determining if a node has failed totally or has simply moved out of range. Similarly to the routing issues above, there are two strategies for dealing with this. The first is to have a centralized node responsible for keeping track of the existence and status of all nodes in the network. The second is to take a distributed approach, where each node keeps track of either all or a subset of the network nodes.

Neither of these approaches solves the problem of detecting the difference between a failed node, for which some repair action might be needed, and a node which has simply moved out of range of all other nodes in the network, for which the action required is to either relocate the node or put in place additional repeater nodes. Note, however, that in both of these cases the action required must take place by means outside the normal network operation. (If a node has moved out of range, someone must move it back in range. If a node has failed, someone must repair it.) Thus it would appear to not be unreasonable to relegate this to a local and manual operation, having the operator simply run a diagnostic package when his unit is out of contact with the network.

Other than the manual probing done at the recognition of a problem, the use of diagnostics in the packet radio network occur at power-up. Each node can run self-diagnostics at that time and take appropriate action if tests fail. Examples of built-in testing are the scanning of memory, the parity checking of code as it loads, the cycling of transmit frequencies, and the stepping through the available power levels. In spite of the usefulness of internal tests, the value of cross-net diagnostics and debugging cannot be overemphasized. It too should be part of the network design.

**B. Connecting the Packet Radio Network to the External World**

Packet radio networks can be operated autonomously or can be connected with other packet-switched networks. The two major areas of concern in the latter case are the specifics of the interconnections process, normally embodied in gateways, and the addressing scheme.

**1) Gateways:** Gateways can perform many functions but, as far as addressing is concerned, they are packet translation devices that interpret addresses at the Internet level and impose headers (addresses) appropriate both to the

local networks to which they are attached as well as other networks. They are host-level devices and to work correctly must have some relationship with not only the other gateways of the internet but the network-attached hosts themselves.

Gateways may have an additional role in highly mobile networks such as packet radio where topological partitioning may occur dynamically. Under these circumstances, the gateways, normally internet devices, may take on a role of intranetwork addressing and routing. Specifically, the internet may become the trajectory over which an intranet packet gets delivered when a single network temporarily diverges [22]. Whenever gateways play important roles such as this in mobile packet radio networks, the following issue arises: **Should Addressing and Routing Be Network-Or Gateway-Based?** Network-based addressing means that each network has a unique name and address of which all relevant gateways are aware. In this case, all points within a single network share some portion of their address in common. In contrast, if gateway-based addressing is used, then internet packets are routed from gateway to gateway and each gateway attached to a network must have some means to route packets to destinations within that network. Furthermore, in this case, hosts must have a means to bind themselves dynamically to at least one gateway. Gateway-based routing, while somewhat less intuitive, provides a solution to the problem of what to do when a single network becomes partitioned.

**2) Network Access—Methods and Administration:**

Network access means the functional entry of a network by a person or device capable of using resources within the network or its attached devices. As a general rule it is prudent, depending upon the threat, to exercise access control at the periphery of the network rather than at some centralized (or interior) point or points. Exercising access control at some internal point means that the network must offer a petitioner transport to that point without knowledge as to whether he is entitled to entry or not.

Packet radio with mobile nodes means that access can occur virtually anywhere within the topology. If so, how can access control best work? Most packet networks provide access through either a connected host or directly through a network-based device (such as a dial-up port). The combination is very convenient, principally for the traveling user who might find it difficult to gain access to a host when not in his normal area. Earlier conventions, wherein network access was not critically controlled and control of host access was invoked at the host only, led to considerable vulnerability to both the network and the attached hosts. Because of the wide host accessibility once network access had been gained, network-based access points have characteristically been a weak point in protecting networks from unwanted host entry. Network log-on hosts are increasingly the rule where network-based access is afforded and they may be practical depending on how close to the actual point of network entry access gets controlled.

In mobile packet radio network entry can occur at any node. Mobile users may request entry (connection service) at different places at different times or different places at the same time. Obviously, it becomes more difficult to distribute the access authorization in this situation than if entry were at a fixed location. If access control is decentralized, all nodes may require all authorizations for all mobile users at all times.
Thus the issue is HOW SHOULD ACCESS BE CONTROLLED, AT NETWORK ENTRY OR AT SOME MORE CENTRALIZED POINT? A distributed access control system has some problems of concurrency but minimizes resource utilization in the access process. A centralized one suffers from single-point vulnerability and needs to have some means to prevent someone from tying up the network with access requests.

While the above deals only with single network access it is assumed that access from the internet is also possible. Gateways may then hold the same access role as the collection of hosts, depending on how important access control to the network itself is viewed.

II. ADDRESSING AND NAMING: WHAT ADDRESSING AND NAMING CONVENTIONS ARE APPROPRIATE FOR PACKET RADIO NETWORK INTERNAL TRAFFIC? Increasingly it is assumed that addressing in a packet radio network should be compatible with some internet addressing convention. If so then the question naturally arises as to whether purely internal packet radio network traffic should carry the internet address (header). This is principally a question of efficiency. If internal packet radio network nodes carry full internet addresses then each packet has extraordinary overhead. On the other hand, if it is desirable to load or debug a given node from an internet host not on the radio net, then internet addressing may be very convenient.

While an address of a switching node can also serve as the name of an attached host, there is, in mobile networks, some advantage to keeping name and address separate. A host that moves topologically can retain a constant name while having its address change. As mentioned earlier, DARPA's packet radio network dynamically binds name and address as part of its continuous assessment of topology. Whether dynamic name-address binding is needed, then, is an operational issue.

A very useful service to have on any network is a directory of users and hosts. These network servers, among other things, may give the name and address of the entity sought. They may serve the user community or even the internal switches as nodes come and go. Therefore, an issue is whether to use name and address servers, where should they be located, and how are they to be maintained?

4) Security—Compatibility and Constraints: Because of the flexible routing of packets in packet switched networks, such networks are more suited to end-to-end rather than link encryption. End-to-end encryption means the user-supplied contents of a given packet are not available at any intermediate switching node. But, for the packets to be routed correctly, the header must be readable by all switching nodes. Thus one of the interesting issues in packet radio networks is SHOULD THE HEADER REMAIN IN THE CLEAR EVEN IF THE TEXT OF THE PACKET IS ENCRYPTED? If the security demands on network operation are severe, then the entire packet must be encrypted, with corresponding requirements of header decryption in each node. There are many advantages, of course, to keeping the interior of the network, the subnet, free of encryption and decryption equipment. Not only is it then unnecessary to deal with the movement of keys, but a very complex network management system for encryption can be avoided as well.

Another encryption strategy for packet radio networks that might be interesting to explore is the use of public key encryption for the packet header only. In fact, a sequential set of n pairs of addresses, as in a route, could be encrypted in the corresponding public key of each relevant node. This technique could thus define a unique route that could not be entirely decoded, even by the participating nodes. Neighbor nodes could still assist in the routing if congestion arose but only if the source that specified the route enabled that feature. Distributed routing decisions, in other words, would not be possible.

C. Impact on the Radio Spectrum

Electromagnetic compatibility is a two-way street. Like all radio systems, packet radio networks are vulnerable to intentional or accidental impact by external signals or noise and they, in turn, can have impact on others. The spectrum is a carefully regulated resource in most countries and therefore the radio frequency, the bandwidth, the power, modulation, and sometimes the direction and directivity of the signal are subject to review and approval.

1) Electromagnetic Compatibility: While not necessary, some packet radio systems use a large time–bandwidth product through the use of techniques such as spread spectrum; that is, the modulation bandwidth is much greater than the information bandwidth. There are several reasons for this: the opportunity for encoding for code division multiple access (CDMA), less fading because of the ability to resolve and process multipath components, a lower probability of interference from narrow-band signals within the bandwidth occupied by decorrelating them, and for low radio emission visibility. However, this consideration of the spectral distribution of energy is but a part of one of the most important issues in any packet radio system: HOW SHOULD A GIVEN PACKET RADIO SYSTEM BE ACCOMMODATED IN THE EXISTING RADIO SPECTRUM? This issue is not one of simply being compatible with the allocations of the spectrum imposed by some governing body. The choice of frequency also has to do with whether the propagation characteristics that determine coverage and reliability are suitable and if the cost, portability, and placement requirements of the radio frequency components are likewise acceptable, as discussed in Section II.

2) Electronic Counter-Countermeasures and Noise Immunity: Because of the highly capable network management in a packet radio network, it is possible to give the network some countermeasures of its own, ones intrinsically different from those offered by individual nodes. For example, the knowledge of a jammer targeting one or two nodes can be used to direct traffic around these threatened parts of the network. Power control on a link-by-link basis to hide the presence of the network from a known enemy is another example. But as with any of the other overhead functions of the network the major issue ultimately becomes THE AMOUNT OF NETWORK-BASED COUNTERMEASURES VERSUS THE MAINTENANCE OF THROUGHPUT. Since the value of countermeasures is a purely military question, it cannot be answered apart from some specific mission statements.

While this issue surrounding countermeasures is military-oriented, there are also some similar but less hostile threats on the domestic scene. Ignition noise and other varieties of noise are inimical to digital data radio systems and signal processing is one of the best means to cope with them. The coherent spreading of the radio frequency wave-
form, a widely practiced ECCM technique, is one effective means of combating impulsive interference. Likewise, adaptive routing techniques can be used to route around an area of the network experiencing a high degree of interference.

3) Efficiency: Several efficiency factors may be defined for a network, but ultimately one must consider the number of users that can effectively share a given set of resources, say bandwidth and area. (The sharing of time is intrinsic to packet switching.) A typical measure might be the number of users accommodated in a given area over a given time for a given bandwidth. Unfortunately, the shared collective resource in packet radio has many dimensions, including not only time-bandwidth but space, antenna directivity, and conceivably polarization. There are other aspects of network functionality that must also be considered in any examination of efficiency. These functions deal mostly with topological change: the use of resource, including bandwidth, to measure, assess, and control the network and to maintain individual connections, even under the threat of countermeasures.

Perhaps the most important issue regarding efficiency in a mobile packet radio network is the value of this new network functionality. More quantitatively WHAT FRACTION OF TOTAL AVAILABLE RESOURCES SHOULD BE ALLOCATED FOR NETWORK OVERHEAD FUNCTIONS? Obviously, there is no one answer to this question but perhaps more importantly, there is no general means to answer it. The use of network resources for other than pure data transport varies from desirable to critical and evaluating the network on transport alone does not recognize that value. A well-designed network will allocate the maximum resource to meeting transport criteria consistent with connection establishment and maintenance. Conceivably, network overhead could take more time-bandwidth than user traffic and still be justified. In the final analysis, then, performance can only be judged in light of the overall mission of the network.

V. NETWORK PERFORMANCE EVALUATION AND COSTS

There is a great deal of complexity in the design of a packet radio network, but no aspect is more difficult than the gauging of network performance. Choosing the metrics and forming the criteria for optimum performance has not been accomplished except for the simplest of topologies. It is not only the topologies that make the analysis complicated, but the fact that the numerous functions the networks perform have not been given relative value. Internal network monitoring and diagnostics, dynamic hand-off, route establishment, alternate routing, delay minimization, and access control are just some of the features other than simple transport that impact network performance. How well a network performs these various functions are components of an overall figure of merit yet to be devised.

A. Metrics for Performance Evaluation

1) Metrics from a User Perspective: While most analyti-
cal expressions of performance deal only with throughput and delay, the user is generally interested in several more performance metrics. Table 1 lists some of these.

<table>
<thead>
<tr>
<th>Table 1 User-Oriented Performance Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
</tr>
<tr>
<td>Delay</td>
</tr>
<tr>
<td>Priority</td>
</tr>
<tr>
<td>Throughput</td>
</tr>
<tr>
<td>Coverage</td>
</tr>
<tr>
<td>Mobility</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
</tbody>
</table>

The costs of a packet radio network, as with any network, can be grouped into design, implementation, and operating costs. Cost cutting in the design phase may have financial repercussions on operations. In the case of the DARPA packet radio network the extra time and cost taken to provide remote diagnostics and maintenance have paid themselves many times over. Ultimately, the cost issue is the tradeoff of THE COST OF CREATING THE NETWORK VERSUS THE COST OF OPERATING IT. The outcome is often defined by the amount of each type of money available (capitalization costs versus operating costs) and by such issues as cash flow and payback. In both a military and commercial setting, advantage will accrue from keeping network operation from being very labor-intensive.

Another perspective on network costs comes from looking at the cost of components. In any large network the costs will be dominated by the unit cost of the switch, even if some network control or monitoring equipment is to be used. Another dominant equipment cost is the means of attaching terminal equipment to the network: interface devices and so-called digital termination equipment. So ultimately, the important cost issue becomes HOW TO MINIMIZE THE COST OF THE HIGHLY REPLICATED COMPONENTS OF THE NETWORK.

VI. SUMMARY AND CONCLUSIONS

We have presented in this paper a discussion of a number of issues that arise in the design, development, and deployment of a packet radio network. It is clear from this discussion that there is a large collection of design choices to be made. A number of packet radio networks have already
been developed and tested, and it is expected that more will be deployed in the future as the technology matures. It is hoped that this paper provides the context in which the lessons of past research can be applied to the systems of the future.

ACKNOWLEDGMENT

The list of contributors to packet radio technology, and the ideas in this paper, is clearly too long to include here. However, a special note of acknowledgment is due R. Kahn [8] for his vision in initiating the DARPA packet radio network program, upon which much of the other packet radio efforts have been based.

REFERENCES


Barry M. Leiner (Senior Member, IEEE), for a photograph and biography see page 4 of this issue.

Donald L. Nielson (Member, IEEE), for a photograph and biography see page 5 of this issue.

Fouad A. Tobagi (Fellow, IEEE), for a photograph and biography see page 5 of this issue.