Singh, S. “Wireless LANs”
Mobile Communications Handbook
Ed. Suthan S. Suthersan
Boca Raton: CRC Press LLC, 1999

Wireless LANs

Suresh Singh
Oregon State University

32.1 Introduction

A proliferation of high-performance portable computers combined with end-user need for communication is fueling a dramatic growth in wireless **local area network** (LAN) technology. Users expect to have the ability to operate their portable computer globally while remaining connected to communications networks and service providers. Wireless LANs and cellular networks, connected to high-speed networks, are being developed to provide this functionality.

Before delving deeper into issues relating to the design of wireless LANs, it is instructive to consider some scenarios of user mobility.

1. A simple model of user mobility is one where a computer is physically moved while retaining network connectivity at either end. For example, a move from one room to another as in a hospital where the computer is a hand-held device displaying patient charts and the nurse using the computer moves between wards or floors while accessing patient information.
2. Another model situation is where a group of people (at a conference, for instance) set up an ad-hoc LAN to share information as in Fig. 32.1.
3. A more complex model is one where several computers in constant communication are in motion and continue to be networked. For example, consider the problem of having robots in space collaborating to retrieve a satellite.

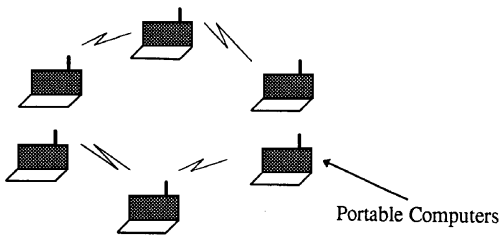

FIGURE 32.1: Ad-hoc wireless LAN.

A great deal of research has focused on dealing with physical layer and **medium access control (MAC)** layer protocols. In this chapter we first summarize standardization efforts in these areas. The remainder of the chapter is then devoted to a discussion of networking issues involved in wireless LAN design. Some of the issues discussed include routing in wireless LANs (i.e., how does data find its destination when the destination is mobile?) and the problem of providing service guarantees to end users (e.g., error-free data transmission or bounded delay and bounded bandwidth service, etc.).

32.2 Physical Layer Design

Two media are used for transmission over wireless LANs, infrared and radio frequency. RF LANs are typically implemented in the industrial, scientific, and medical (ISM) frequency bands 902–928 MHz, 2400–2483.5 MHz and 5725–5850 MHz. These frequencies do not require a license allowing the LAN product to be portable, i.e., a LAN can be moved without having to worry about licensing.

IR and RF technologies have different design constraints. IR receiver design is simple (and thus inexpensive) in comparison to RF receiver design because IR receivers only detect the amplitude of the signal not the frequency or phase. Thus, a minimal of filtering is required to reject interference. Unfortunately, however, IR shares the electromagnetic spectrum with the sun and incandescent or fluorescent light. These sources of modulated infrared energy reduce the signal-to-noise ratio of IR signals and, if present in extreme intensity, can make the IR LANs inoperable. There are two approaches to building IR LANs.

1. The transmitted signal can be focused and aimed. In this case the IR system can be used outdoors and has an area of coverage of a few kilometers.
2. The transmitted signal can be bounced off the ceiling or radiated omnidirectionally. In either case, the range of the IR source is 10–20 m (i.e., the size of one medium-sized room).

RF systems face harsher design constraints in comparison to IR systems for several reasons. The increased demand for RF products has resulted in tight regulatory constraints on the allocation and use of allocated bands. In the U.S., for example, it is necessary to implement spectrum spreading for operation in the ISM bands. Another design constraint is the requirement to confine the emitted spectrum to a band, necessitating amplification at higher carrier frequencies, frequency conversion using precision local oscillators, and selective components. RF systems must also cope with environmental noise that is either naturally occurring, for example, atmospheric noise or man made, for example, microwave ovens, copiers, laser printers, or other heavy electrical machinery. RF LANs operating in the ISM frequency ranges also suffer interference from amateur radio operators.

Operating LANs indoors introduces additional problems caused by multipath propagation, Rayleigh fading, and absorption. Many materials used in building construction are opaque to IR radiation resulting in incomplete coverage within rooms (the coverage depends on obstacles within the room that block IR) and almost no coverage outside closed rooms. Some materials, such as white plasterboard, can also cause reflection of IR signals. RF is relatively immune to absorption and reflection problems. Multipath propagation affects both IR and RF signals. The technique to alleviate the effects of multipath propagation in both types of systems is the same use of aimed (directional) systems for transmission enabling the receiver to reject signals based on their angle of incidence. Another technique that may be used in RF systems is to use multiple antennas. The phase difference between different paths can be used to discriminate between them.

Rayleigh fading is a problem in RF systems. Recall that Rayleigh fading occurs when the difference in path length of the same signal arriving along different paths is a multiple of half a wavelength. This causes the signal to be almost completely canceled out at the receiver. Because the wavelengths used in IR are so small, the effect of Rayleigh fading is not noticeable in those systems. RF systems, on the other hand, use wavelengths of the order of the dimension of a laptop. Thus, moving the computer a small distance could increase/decrease the fade significantly.

Spread spectrum transmission technology is used for RF-based LANs and it comes in two varieties: direct-sequence spread spectrum (DSSS) and frequency-hopping spread spectrum (FHSS). In a FHSS system, the available band is split into several channels. The transmitter transmits on one channel for a fixed time and then hops to another channel. The receiver is synchronized with the transmitter and hops in the same sequence; see Fig. 32.2(a). In DSSS systems, a random binary string is used to modulate the transmitted signal. The relative rate between this sequence and user data is typically between 10 and 100; see Fig. 32.2(b).

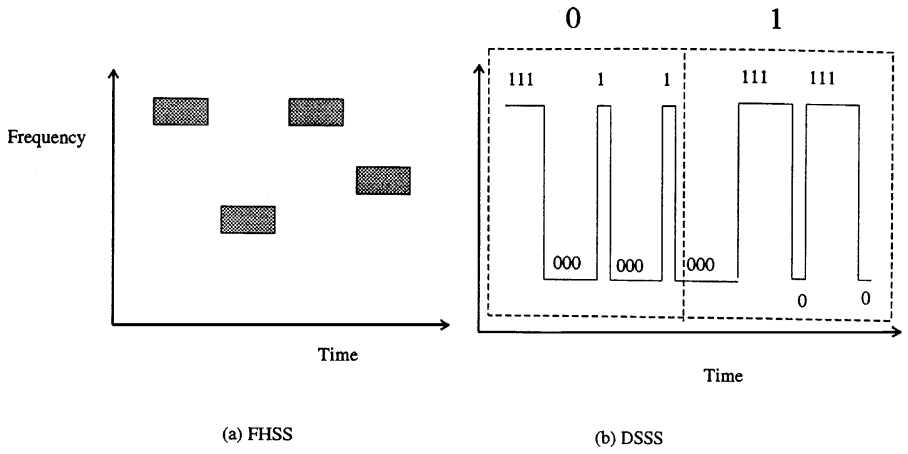

FIGURE 32.2: Spread spectrum.

The key requirements of any transmission technology is its robustness to noise. In this respect DSSS and FHSS show some differences. There are two possible sources of interference for wireless LANs: the presence of other wireless LANs in the same geographical area (i.e., in the same building, etc.) and interference due to other users of the ISM frequencies. In the latter case, FHSS systems have a greater ability to avoid interference because the hopping sequence could be designed to prevent potential

interference. DSSS systems, on the other hand, do exhibit an ability to recover from interference because of the use of the spreading factor [Fig. 32.2(b)].

It is likely that in many situations several wireless LANs may be collocated. Since all wireless LANs use the same ISM frequencies, there is a potential for a great deal of interference. To avoid interference in FHSS systems, it is necessary to ensure that the hopping sequences are orthogonal. To avoid interference in DSSS systems, on the other hand, it is necessary to allocate different channels to each wireless LAN. The ability to avoid interference in DSSS systems is, thus, more limited in comparison to FHSS systems because FHSS systems use very narrow subchannels (1 MHz) in comparison to DSSS systems that use wider subchannels (for example, 25 MHz), thus, limiting the number of wireless LANs that can be collocated. A summary of design issues can be found in [1].

32.3 MAC Layer Protocols

MAC protocol design for wireless LANs poses new challenges because of the in-building operating environment for these systems. Unlike wired LANs (such as the ethernet or token ring), wireless LANs operate in strong multipath fading channels where channel characteristics can change in very short distances resulting in unreliable communication and unfair channel access due to capture. Another feature of the wireless LAN environment is that carrier sensing takes a long time in comparison to wired LANs; it typically takes between 30 and 50 μs (see [4]), which is a significant portion of the packet transmission time. This results in inefficiencies if the CSMA family of protocols is used without any modifications.

Other differences arise because of the mobility of users in wireless LAN environments. To provide a building (or any other region) with wireless LAN coverage, the region to be covered is divided into cells as shown in Fig. 32.3. Each cell is one wireless LAN, and adjacent cells use different frequencies to minimize interference. Within each cell there is an access point called a **mobile support station** (MSS) or base station that is connected to some wired network. The mobile users are called **mobile hosts (MH)**. The MSS performs the functions of channel allocation and providing connectivity to existing wired networks; see Fig. 32.4. Two problems arise in this type of an architecture that are not present in wired LANs.

1. The number of nodes within a cell changes dynamically as users move between cells. How can the channel access protocol dynamically adapt to such changes efficiently?
2. When a user moves between cells, the user has to make its presence known to the other nodes in the cell. How can this be done without using up too much bandwidth? The protocol used to solve this problem is called a handoff protocol and works along the following lines: A switching station (or the MSS nodes working together, in concert) collects signal strength information for each mobile host within each cell. Note that if a mobile host is near a cell boundary, the MSS node in its current cell as well as in the neighboring cell can hear its transmissions and determine signal strengths. If the mobile host is currently under the coverage of MSS M1 but its signal strength at MSS M2 becomes larger, the switching station initiates a handoff whereby the MH is considered as part of M2's cell (or network).

The mode of communication in wireless LANs can be broken in two: communication from the mobile to the MSS (called *uplink* communication) and communication in the reverse direction (called *downlink* communication). It is estimated that downlink communication accounts for about 70–80% of the total consumed bandwidth. This is easy to see because most of the time users request files or data in other forms (image data, etc.) that consume much more transmission bandwidth than

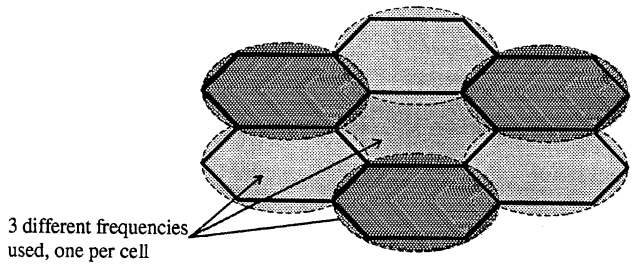

FIGURE 32.3: Cellular structure for wireless LANs (note frequency reuse).

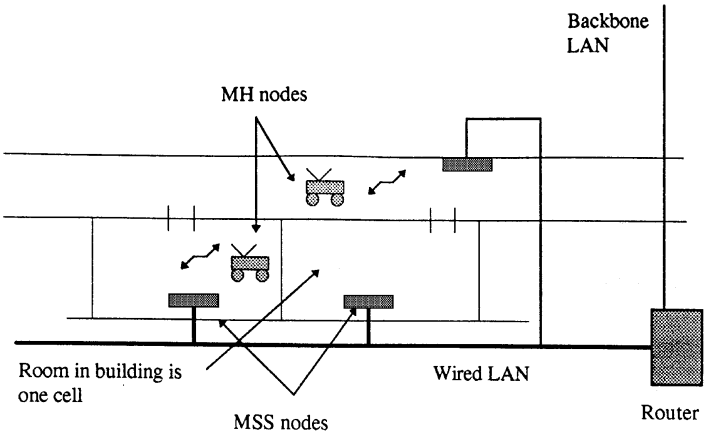

FIGURE 32.4: In-building LAN (made up of several wireless LANs).

the requests themselves. In order to make efficient use of bandwidth (and, in addition, guarantee service requirements for real-time data), most researchers have proposed that the downlink channel be controlled entirely by the MSS nodes. These nodes allocate the channel to different mobile users based on their current requirements using a protocol such as **time division multiple access** (TDMA). What about uplink traffic? This is a more complicated problem because the set of users within a cell is dynamic, thus making it infeasible to have a static channel allocation for the uplink. This problem is the main focus of MAC protocol design.

What are some of the design requirements of an appropriate MAC protocol? The IEEE 802.11 recommended standard for wireless LANs has identified almost 20 such requirements, some of which are discussed here (the reader is referred to [3], for further details). Clearly any protocol must maximize throughput while minimizing delays and providing fair access to all users. In addition to these requirements, however, mobility introduces several new requirements.

1. The MAC protocol must be independent of the underlying physical layer transmission technology adopted (be it DSSS, FHSS or IR).
2. The maximum number of users can be as high as a few hundred in a wireless LAN. The MAC protocol must be able to handle many users without exhibiting catastrophic degradation of service.
3. The MAC protocols must provide secure transmissions because the wireless medium is

easy to tap.

4. The MAC protocol needs to work correctly in the presence of collocated networks.
5. It must have the ability to support ad-hoc networking (as in Fig. 32.1).
6. Other requirements include the need to support priority traffic, preservation of packet order, and an ability to support multicast.

Several contention-based protocols currently exist that could be adapted for use in wireless LANs. The protocols currently being looked by IEEE 802.11 include protocols based on **carrier sense multiple access (CSMA)**, polling, and TDMA. Protocols based on **code division multiple access** (CDMA) and **frequency division multiple access** (FDMA) are not considered because the processing gains obtained using these protocols are minimal while, simultaneously, resulting in a loss of flexibility for wireless LANs.

It is important to highlight an important difference between networking requirements of ad-hoc networks (as in Fig. 32.1) and networks based on cellular structure. In cellular networks, all communication occurs between the mobile hosts and the MSS (or base station) within that cell. Thus, the MSS can allocate channel bandwidth according to requirements of different nodes, i.e., we can use centralized channel scheduling for efficient use of bandwidth. In ad-hoc networks there is no such central scheduler available. Thus, any multiaccess protocol will be contention based with little explicit scheduling. In the remainder of this section we focus on protocols for cell-based wireless LANs only.

All multiaccess protocols for cell-based wireless LANs have a similar structure; see [3].

1. The MSS announces (explicitly or implicitly) that nodes with data to send may contend for the channel.
2. Nodes interested in sending data contend for the channel using protocols such as CSMA.
3. The MSS allocates the channel to successful nodes.
4. Nodes transmit packets (contention-free transmission).
5. MSS sends an explicit acknowledgment (ACK) for packets received.

Based on this model we present three MAC protocols.

32.3.1 Reservation-TDMA (R-TDMA)

This approach is a combination of TDMA and some contention protocol (see PRMA in [7]). The MSS divides the channel into slots (as in TDMA), which are grouped into frames. When a node wants to transmit it needs to reserve a slot that it can use in every consecutive frame as long as it has data to transmit. When it has completed transmission, other nodes with data to transmit may contend for that free slot. There are four steps to the functioning of this protocol.

a. At the end of each frame the MSS transmits a feedback packet that informs nodes of the current reservation of slots (and also which slots are free). This corresponds to steps 1 and 3 from the preceding list.
b. During a frame, all nodes wishing to acquire a slot transmit with a probability ρ during a free slot. If a node is successful it is so informed by the next feedback packet. If more than one node transmits during a free slot, there is a collision and the nodes try again during the next frame. This corresponds to step 2.
c. A node with a reserved slot transmits data during its slot. This is the contention-free transmission (step 4).

d. The MSS sends ACKs for all data packets received correctly. This is step 5.

The R-TDMA protocol exhibits several nice properties. First and foremost, it makes very efficient use of the bandwidth, and average latency is half the frame size. Another big benefit is the ability to implement power conserving measures in the portable computer. Since each node knows when to transmit (nodes transmit during their reserved slot only) it can move into a power-saving mode for a fixed amount of time, thus increasing battery life. This feature is generally not available in CSMA-based protocols. Furthermore, it is easy to implement priorities because of the centralized control of scheduling. One significant drawback of this protocol is that it is expensive to implement (see [2]).

32.3.2 Distributed Foundation Wireless MAC (DFWMAC)

The CSMA/CD protocol has been used with great success in the ethernet. Unfortunately, the same protocol is not very efficient in a wireless domain because of the problems associated with cell interference (i.e., interference from neighboring cells), the relatively large amount of time taken to sense the channel (see [6]) and the hidden terminal problem (see [12, 13]). The current proposal is based on a CSMA/collision avoidance (CA) protocol with a four-way handshake; see Fig. 32.5.

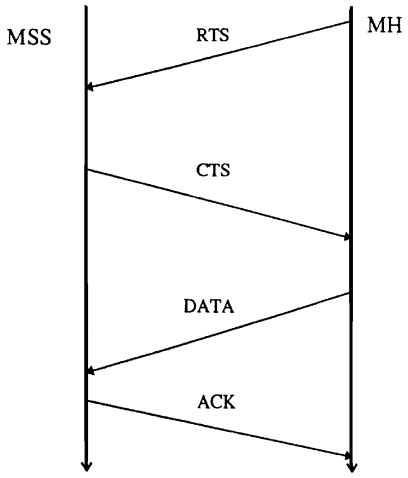

FIGURE 32.5: CSMA/CA and four-way handshaking protocol.

The basic operation of the protocol is simple. All MH nodes that have packets to transmit compete for the channel by sending ready to transmit (RTS) messages using nonpersistent CSMA. After a station succeeds in transmitting a RTS, the MSS sends a clear to transmit (CTS) to the MH. The MH transmits its data and then receives an ACK. The only possibility of collision that exists is in the RTS phase of the protocol and inefficiencies occur in the protocol, because of the RTS and CTS stages. Note that unlike R-TDMA it is harder to implement power saving functions. Furthermore, latency is dependent on system load making it harder to implement real-time guarantees. Priorities are also not implemented. On the positive side, the hardware for this protocol is very inexpensive.

32.3.3 Randomly Addressed Polling (RAP)

In this scheme, when a MSS is ready to collect uplink packets it transmits a READY message. At this point all nodes with packets to send attempt to grab the channel as follows.

a. Each MH with a packet to transmit generates a random number between 0 and P.
b. All active MH nodes simultaneously and orthogonally transmit their random numbers (using CDMA or FDMA). We assume that all of these numbers are received correctly by the MSS. Remember that more than one MH node may have selected the same random number.
c. Steps a and b are repeated L times.
d. At the end of L stages, the MSS determines a stage (say, k) where the total number of distinct random numbers was the largest. The MSS polls each distinct each random number in this stage in increasing order. All nodes that had generated the polled random number transmit packets to the MSS.
e. Since more than one node may have generated the same random number, collisions are possible. The MSS sends a ACK or NACK after each such transmission. Unsuccessful nodes try again during the next iteration of the protocol.

The protocol is discussed in detail in [4] and a modified protocol called GRAP (for group RAP) is discussed in [3]. The authors propose that GRAP can also be used in the contention stage (step 2) for TDMA- and CSMA-based protocols.

32.4 Network Layer Issues

An important goal of wireless LANs is to allow users to move about freely while still maintaining all of their connections (network resources permitting). This means that the network must route all packets destined for the mobile user to the MSS of its current cell in a transparent manner. Two issues need to be addressed in this context.

- How can users be addressed?
- How can active connections for these mobile users be maintained?

Ioanidis, Duchamp, and Maguire [8] propose a solution called the IPIP (IP-within-IP) protocol. Here each MH has a unique **internet protocol** (IP) address called its home address. To deliver a packet to a remote MH, the source MSS first broadcasts an address resolution protocol (ARP) request to all other MSS nodes to locate the MH. Eventually some MSS responds. The source MSS then encapsulates each packet from the source MH within another packet containing the IP address of the MSS in whose cell the MH is located. The destination MSS extracts the packet and delivers it to the MH. If the MH has moved away in the interim, the new MSS locates the new location of the MH and performs the same operation. This approach suffers from several problems as discussed in [11]. Specifically, the method is not scaleable to a network spanning areas larger than a campus for the following reasons.

1. IP addresses have a prefix identifying the campus subnetwork where the node lives; when the MH moves out of the campus, its IP address no longer represents this information.
2. The MSS nodes serve the function of routers in the mobile network and, therefore, have the responsibility of tracking all of the MH nodes globally causing a lot of overhead in terms of message passing and packet forwarding; see [5].

Teraoka and Tokoro [11], have proposed a much more flexible solution to the problem called virtual IP (VIP). Here every mobile host has a virtual IP address that is unchanging regardless of the location of the MH. In addition, hosts have physical network addresses (traditional IP addresses) that may change as the host moves about. At the transport layer, the target node is always specified by its VIP address only. The address resolution from the VIP address to the current IP address takes place either at the network layer of the same machine or at a gateway. Both the host machines and the gateways maintain a cache of VIP to IP mappings with associated timestamps. This information is in the form of a table called *address mapping table* (AMT). Every MH has an associated *home gateway*. When a MH moves into a new subnetwork, it is assigned a new IP address. It sends this new IP address and its VIP address to its home gateway via a *VipConn* control message. All intermediate gateways that relay this message update their AMT tables as well. During this process of updating the AMT tables, all packets destined to the MH continue to be sent to the old location. These packets are returned to the sender, who then sends them to the home gateway of the MH. It is easy to see that this approach is easily scaleable to large networks, unlike the IPIP approach.

32.4.1 Alternative View of Mobile Networks

The approaches just described are based on the belief that mobile networks are merely an extension of wired networks. Other authors [10] disagree with this assumption because there are fundamental differences between the mobile domain and the fixed wired network domain. Two examples follow.

1. The available bandwidth at the wireless link is small; thus, end-to-end packet retransmission for transmission control protocol (TCP)-like protocols (implemented over datagram networks) is a bad idea. This leads to the conclusion that transmission within the mobile network must be connection oriented. Such a solution, using virtual circuits (VC), is proposed in [5].
2. The bandwidth available for a MH with open connections changes dynamically since the number of other users present in each cell varies randomly. This is a feature not present in fixed high-speed networks where, once a connection is set up, its bandwidth does not vary much. Since bandwidth changes are an artifact of mobility and are dynamic, it is necessary to deal with the consequences (e.g., buffer overflow, large delays, etc.) locally to both, i.e., shield fixed network hosts from the idiosyncrasies of mobility as well as to respond to changing bandwidth quickly (without having to rely on end-to-end control). Some other differences are discussed in [10].

32.4.2 A Proposed Architecture

Keeping these issues in mind, a more appropriate architecture has been proposed in Ghai and Singh [5], and Singh [10]. Mobile networks are considered to be different and separate from wired networks. Within a mobile network is a three-layer hierarchy; see Fig. 32.6. At the bottom layer are the MHs. At the next level are the MSS nodes (one per cell). Finally, several MSS nodes are controlled by a **supervisor host (SH)** node (there may be one SH node per small building). The SH nodes are responsible for flow control for all MH connections within their domain; they are also responsible for tracking MH nodes and forwarding packets as MH nodes roam. In addition, the SH nodes serve as a *gateway* to the wired networks. Thus, any connection setup from a MH to a fixed host is broken in two, one from the MH to the SH and another from the SH to the fixed host. The MSS nodes in this design are simply connection endpoints for MH nodes. Thus, they are simple devices that implement the MAC protocols and little else. Some of the benefits of this design are as follows.

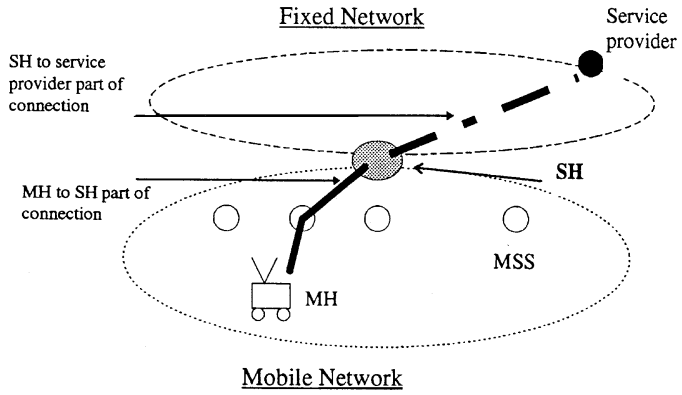

FIGURE 32.6: Proposed architecture for wireless networks.

1. Because of the large coverage of the SH (i.e., a SH controls many cells) the MH remains in the domain of one SH much longer. This makes it easy to handle the consequences of dynamic bandwidth changes locally. For instance, when a MH moves into a crowded cell, the bandwidth available to it is reduced. If it had an open ftp connection, the SH simply buffers undelivered packets until they can be delivered. There is no need to inform the other endpoint of this connection of the reduced bandwidth.
2. When a MH node sets up a connection with a service provider in the fixed network, it negotiates some quality of service (QOS) parameters such as bandwidth, delay bounds, etc. When the MH roams into a crowded cell, these QOS parameters can no longer be met because the available bandwidth is smaller. If the traditional view is adopted (i.e., the mobile networks are extensions of fixed networks) then these QOS parameters will have to be renegotiated each time the bandwidth changes (due to roaming). This is a very expensive proposition because of the large number of control messages that will have to be exchanged. In the approach of Singh [10], the service provider will never know about the bandwidth changes since it deals only with the SH that is accessed via the wired network. The SH bears the responsibility of handling bandwidth changes by either buffering packets until the bandwidth available to the MH increases (as in the case of the ftp example) or it could discard a fraction of real-time packets (e.g., a voice connection) to ensure delivery of most of the packets within their deadlines. The SH could also instruct the MSS to allocate a larger amount of bandwidth to the MH when the number of buffered packets becomes large. Thus, the service provider in the fixed network is shielded from the mobility of the user.

32.4.3 Networking Issues

It is important for the network to provide connection-oriented service in the mobile environment (as opposed to connectionless service as in the internet) because bandwidth is at a premium in wireless networks, and it is, therefore, inadvisable to have end-to-end retransmission of packets (as in TCP). The proposed architecture is well suited to providing connection-oriented service by using VCs.

In the remainder of this section we look at how virtual circuits are used within the mobile network and how routing is performed for connections to mobile hosts. Every connection set up with one or more MH nodes as a connection endpoint is routed through the SH nodes and each connection

is given a unique VC number. The SH node keeps track of all MH nodes that lie within its domain. When a packet needs to be delivered to a MH node, the SH first buffers the packet and then sends it to the MSS at the current location of the MH or to the predicted location if the MH is currently between cells. The MSS buffers all of these packets for the MH and transmits them to the MH if it is in its cell. The MSS discards packets after transmission or if the SH asks it to discard the packets. Packets are delivered in the correct order to the MH (without duplicates) by having the MH transmit the expected sequence number (for each VC) during the initial handshake (i.e., when the MH first enters the cell). The MH sends ACKs to the SH for packets received. The SH discards all packets that have been acknowledged. When a MH moves from the domain of SH1 into the domain of SH2 while having open connections, SH1 continues to forward packets to SH2 until either the connections are closed or until SH2 sets up its own connections with the other endpoints for each of MH's open connections (it also gives new identifiers to all these open connections). The detailed protocol is presented in [5].

The SH nodes are all connected over the fixed (wired) network. Therefore, it is necessary to route packets between SH nodes using the protocol provided over the fixed networks. The VIP protocol appears to be best suited to this purpose. Let us assume that every MH has a globally unique VIP address. The SHs have both a VIP as well as a fixed IP address. When a MH moves into the domain of a SH, the IP address affixed to this MH is the IP address of the SH. This ensures that all packets sent to the MH are routed through the correct SH node. The SH keeps a list of all VIP addresses of MH nodes within its domain and a list of open VCs for each MH. It uses this information to route the arriving packets along the appropriate VC to the MH.

32.5 Transport Layer Design

The transport layer provides services to higher layers (including the application layer), which include connectionless services like UDP or connection-oriented services like TCP. A wide variety of new services will be made available in the high-speed networks, such as continuous media service for real-time data applications such as voice and video. These services will provide bounds on delay and loss while guaranteeing some minimum bandwidth.

Recently variations of the TCP protocol have been proposed that work well in the wireless domain. These proposals are based on the traditional view that wireless networks are merely extensions of fixed networks. One such proposal is called I-TCP [2] for indirect TCP. The motivation behind this work stems from the following observation. In TCP the sender times out and begins retransmission after a timeout period of several hundred milliseconds. If the other endpoint of the connection is a mobile host, it is possible that the MH is disconnected for a period of several seconds (while it moves between cells and performs the initial greeting). This results in the TCP sender timing out and transmitting the same data several times over, causing the effective throughput of the connection to degrade rapidly. To alleviate this problem, the implementation of I-TCP separates a TCP connection into two pieces—one from the fixed host to another fixed host that is near the MH and another from this host to the MH (note the similarity of this approach with the approach in Fig. 32.6). The host closer to the MH is aware of mobility and has a larger timeout period. It serves as a type of gateway for the TCP connection because it sends ACKs back to the sender before receiving ACKs from the MH. The performance of I-TCP is far superior to traditional TCP for the mobile networks studied.

In the architecture proposed in Fig. 32.6, a TCP connection from a fixed host to a mobile host would terminate at the SH. The SH would set up another connection to the MH and would have the responsibility of transmitting all packets correctly. In a sense this is a similar idea to I-TCP except that in the wireless network VCs are used rather than datagrams. Therefore, the implementation of TCP service is made much easier.

A problem that is unique to the mobile domain occurs because of the unpredictable movement of MH nodes (i.e., a MH may roam between cells resulting in a large variation of available bandwidth in each cell). Consider the following example. Say nine MH nodes have opened 11-kb/s connections in a cell where the available bandwidth is 100 kb/s. Let us say that a tenth mobile host M10, also with an open 11-kb/s connection, wanders in. The total requested bandwidth is now 110 kb/s while the available bandwidth is only 100 kb/s. What is to be done? One approach would be to deny service to M10. However, this seems an unfair policy. A different approach is to penalize all connections equally so that each connection has 10-kb/s bandwidth allocated.

To reduce the bandwidth for each connection from 11 kb/s to 10 kb/s, two approaches may be adopted:

1. Throttle back the sender for each connection by sending control messages.
2. Discard 1-kb/s data for each connection at the SH. This approach is only feasible for applications that are tolerant of data loss (e.g., real-time video or audio).

The first approach encounters a high overhead in terms of control messages and requires the sender to be capable of changing the data rate dynamically. This may not always be possible; for instance, consider a teleconference consisting of several participants where each mobile participant is subject to dynamically changing bandwidth. In order to implement this approach, the data (video or audio or both) will have to be compressed at different ratios for each participant, and this compression ratio may have to be changed dynamically as each participant roams. This is clearly an unreasonable solution to the problem. The second approach requires the SH to discard 1-kb/s of data for each connection. The question is, how should this data be discarded? That is, should the 1 kb of discarded data be consecutive (or clustered) or uniformly spread out over the data stream every 1 s? The way in which the data is discarded has an effect on the final perception of the service by the mobile user. If the service is audio, for example, a random uniform loss is preferred to a clustered loss (where several consecutive words are lost). If the data is compressed video, the problem is even more serious because most random losses will cause the encoded stream to become unreadable resulting in almost a 100% loss of video at the user.

A solution to this problem is proposed in Seal and Singh [9], where a new sublayer is added to the transport layer called the *loss profile transport sublayer* (*LPTSL*). This layer determines how data is to be discarded based on special transport layer markers put by application calls at the sender and based on negotiated loss functions that are part of the QOS negotiations between the SH and service provider. Figure 32.7 illustrates the functioning of this layer at the service provider, the SH, and the MH. The original data stream is broken into *logical segments* that are separated by markers (or flags). When this stream arrives at the SH, the SH discards entire logical segments (in the case of compressed video, one logical segment may represent one frame) depending on the bandwidth available to the MH. The purpose of discarding entire logical segments is that discarding a part of such a segment of data makes the rest of the data within that segment useless—so we might as well discard the entire segment. Observe also that the flags (to identify logical segments) are inserted by the LPTSL via calls made by the application layer. Thus, the transport layer or the LPTSL does not need to know encoding details of the data stream. This scheme is currently being implemented at the University of South Carolina by the author and his research group.

32.6 Conclusions

The need for wireless LANs is driving rapid development in this area. The IEEE has proposed standards (802.11) for the physical layer and MAC layer protocols. A great deal of work, however,

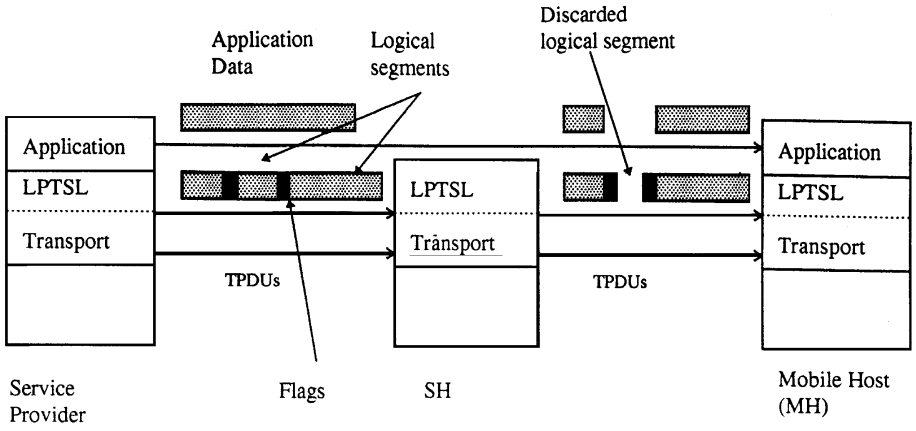

FIGURE 32.7: LPTSL, an approach to handle dynamic bandwidth variations.

remains to be done at the network and transport layers. There does not appear to be a consensus regarding subnet design for wireless LANs. Our work has indicated a need for treating wireless LAN subnetworks as being fundamentally different from fixed networks, thus resulting in a different subnetwork and transport layer designs. Current efforts are underway to validate these claims.

Defining Terms

Carrier-sense multiple access (CSMA): Protocols such as those used over the ethernet.

Medium access control (MAC): Protocols arbitrate channel access between all nodes on a wireless LAN.

Mobile host (MH) nodes: The nodes of wireless LAN.

Supervisor host (SH): The node that takes care of flow-control and other protocol processing for all connections.

References

[1] Bantz, D.F. and Bauchot, F.J., Wireless LAN design alternatives. *IEEE Network*, 8(2), 43–53, 1994.

[2] Barke, A. and Badrinath, B.R., I-TCP: indirect TCP for mobile hosts. Tech. Rept. DCS-TR-314, Dept. Computer Science, Rutgers University, Piscataway, NJ, 1994.

[3] Chen, K.-C., Medium access control of wireless LANs for mobile computing. *IEEE Network*, 8(5), 50–63, 1994.

[4] Chen, K.-C. and Lee, C.H., RAP: a novel medium access control protocol for wireless data networks. *Proc. IEEE GLOBECOM'93*, IEEE Press, Piscataway, NJ, 08854. 1713–1717, 1993.

[5] Ghai, R. and Singh, S., An architecture and communication protocol for picocellular networks. *IEEE Personal Comm. Mag.*, 1(3), 36–46, 1994.

[6] Glisic, S.G., 1-Persistent carrier sense multiple access in radio channel with imperfect carrier sensing. *IEEE Trans. on Comm.*, 39(3), 458–464, 1991.

[7] Goodman, D.J., Cellular packet communications. *IEEE Trans. on Comm.*, 38(8), 1272–1280, 1990.
[8] Ioanidis, J., Duchamp, D., and Maguire, G.Q., IP-based protocols for mobile internetworking. *Proc. of ACM SIGCOMM'91*, ACM Press, New York, NY, 10036 (Sept.), 235–245, 1991.
[9] Seal, K. and Singh, S., Loss profiles: a quality of service measure in mobile computing. *J. Wireless Networks*, 2, 45–61, 1996.
[10] Singh, S., Quality of service guarantees in mobile computing. *J. of Computer Comm.*, 19, 359–371, 1996.
[11] Teraoka, F. and Tokoro, M., Host migration transparency in IP networks: the VIP approach. *Proc. of ACM SIGCOMM*, ACM Press, New York, NY, 10036 (Jan.), 45–65, 1993.
[12] Tobagi, F. and Kleinrock, L., Packet switching in radio channels: Part I carrier sense multiple access models and their throughput delay characteristic. *IEEE Trans. on Comm.*, 23(12), 1400–1416, 1975a.
[13] Tobagi, F. and Kleinrock, L., Packet switching in radio channels: Part II the hidden terminal problem in CSMA and busy-one solution. *IEEE Trans. on Comm.*, 23(12), 1417–1433, 1975b.

Further Information

A good introduction to physical layer issues is presented in Bantz [1] and MAC layer issues are discussed in Chen [3]. For a discussion of network and transport layer issues, see Singh [10] and Ghai and Singh [5].