Wireless LANs and Mobile Networking: Standards and Future Directions

May 31, 1996

Richard O. LaMaire[1†], Arvind Krishna[†], James Panian[‡], and Pravin Bhagwat[†]
[†]IBM T. J. Watson Research Center, Yorktown Heights, NY.
[‡]Ericsson Inc., Research Triangle Park, NC.

1 Introduction

The field of wireless Local Area Networks (LANs) is expanding rapidly as a result of advances in digital communications, portable computers, and semiconductor technology. The early adopters of this technology have primarily been vertical applications that place a premium on the mobility offered by such systems. Examples of these types of applications include inventory control in store and warehouse environments, point of sale terminals, and rental car check-in. Wireless LANs are also increasingly being used in the hospital and university environments in which users are highly mobile and may only require moderate bandwidths. In addition to the mobility that becomes possible with wireless LANs, these systems have also been used in environments where cable installation is expensive or impractical. Such environments include manufacturing floors, trading floors on stock exchanges, conventions and trade shows, and historic buildings. With the increasing proliferation of wireless LANs comes the need for standardization so as to allow interoperability for an increasingly mobile workforce. In this paper, we discuss several emerging standards that relate to wireless LAN systems. These standards include two physical and link layer standards, IEEE 802.11 and ETSI HIPERLAN, as well as a mobile networking standard, Mobile IP, and some developing standards for wireless link management.

In this paper, we focus on the use of radio frequency wireless LANs, as opposed to infrared wireless systems. For radio frequency wireless LANs, the availability of *unlicensed* spectrum is a significant enabler. In the U.S., it was the Federal Communications Commission's rule change in 1985 (first publish, rules modified in 1990)

[1]Author for correspondence: IBM T. J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY, 10598, e-mail: lamaire@watson.ibm.com

allowing unlicensed spread spectrum use of the three Industrial, Scientific and Medical (ISM) frequency bands, that encouraged the development of a number of wireless technologies. Today, unlicensed wireless LAN products are available in all three of the ISM bands at 902-928 MHz, 2.400-2.4835 GHz, and 5.725-5.850 GHz. As we shall describe later, the IEEE 802.11 committee makes use of the 2.4 GHz ISM band.

The discussion that follows treats several types of emerging standards that impact wireless LAN systems. We begin the paper with a description of two influential physical and data link layer standards, the IEEE 802.11 standard and the European HIPERLAN standard. Following this we briefly examine some developments concerning the U.S. Personal Communication Services (PCS) bands, future spectrum allocations, and wireless ATM systems. After describing these physical and link layer developments, we focus on the network layer. We discuss the extensions that are being made to the widely-used Internet Protocol (IP) to deal with mobility (wired or wireless). Finally, in the last technical section of the paper, we describe some emerging standards for wireless link management in which interfaces are specified to provide wireless link information to protocol stacks and applications on the mobile client. In the conclusion of the paper, we speculate on future directions of wireless LAN systems.

2 IEEE 802.11 Wireless LAN Standard

The IEEE 802.11 committee has been working on the establishment of a standard for wireless LANs. Having begun its work in 1990, the 802.11 committee is nearing the completion of the standard, which is expected to be finalized in mid 1996. Much of the standard appears to have reached final form at the current time (early 1996), so that we can describe the main features of the architecture, the multiple physical layers, and the common Medium Access Control (MAC) sublayer [1].

2.1 Architecture

We introduce the general architecture and terminology that was defined by the 802.11 committee [1]. As shown in Fig. 1, two primary topologies are supported by the

802.11 standard: one in which the stations access the backbone network[2] (*Distribution System* in 802.11 nomenclature) via access points (i.e., base stations), and one in which a group of stations communicate directly with each other in an ad hoc network, independent of any infrastructure or base stations. The first topology is useful for providing wireless coverage of building or campus areas by deploying multiple access points whose radio coverage areas overlap to provide complete coverage. The stations that are associated with a given access point are referred to as its Basic Service Set (BSS) in the 802.11 standard, but are more commonly referred to as the members of the access point's cell. The second topology, the one for ad hoc networks, is useful for applications such as file sharing in a conference room scenario. The Medium Access Control (MAC) protocol of the 802.11 standard was developed to allow these two types of topologies to coexist as illustrated by the overlap in the coverage range of the ad hoc network and access point B in Fig. 1.

As a prelude to the following discussion on the HIPERLAN standard, we mention that the IEEE 802.11 draft standard does not provide a mechanism for multihop routing, with the exception of the case discussed in the previous footnote. That is, in an ad hoc network a station can only communicate directly with another station, and in the access point topology a station can only send packets (i.e., frames) through the access point or directly to another station. No station can be used as a relay to the access point without the use of mechanisms that go beyond those currently defined in the standard.

2.2 Physical Layers

The 802.11 draft standard provides for three different types of physical layers to be used: 1) 2.4 GHz ISM band Frequency Hopping (FH) spread-spectrum radio, 2) 2.4 GHz ISM band Direct Sequence (DS) spread-spectrum radio, and 3) Infrared (IR) light. Note that in Europe, the same 2.4 GHz band (as the U.S. ISM band) has been allocated to allow wireless LAN operation, whereas in Japan only the frequencies from 2.471 to 2.497 GHz have been allocated (requiring special provisions in the IEEE 802.11 draft standard). In addition to having three types of physical (PHY) layers,

[2]This backbone network is typically wired, but can also be wireless. For the case of a wireless backbone, the 802.11 standard makes use of a special frame format that effectively tunnels the original frame over the 802.11 wireless network.

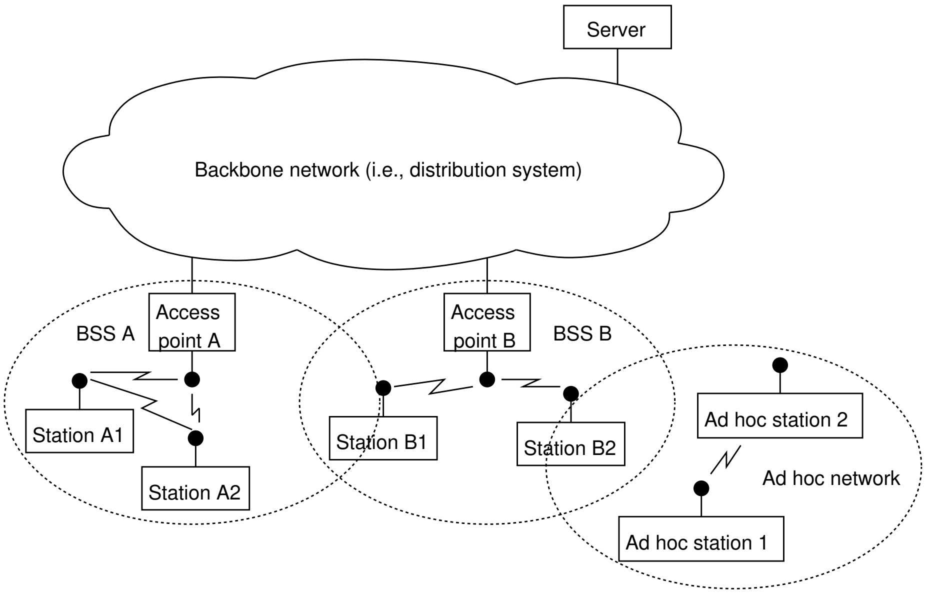

Figure 1: Wireless architecture.

two different data rates (1 Mb/s and 2 Mb/s) have been specified for each of the above PHY layers[3]. At this point in time, most of the attention has been directed towards the radio physical layers so we will only consider these here. Note that the Infrared Data Association (IrDA), a consortium of leading U.S. and Japanese manufacturers of computers, communications equipment and semiconductors, has been developing standards for infrared based attachment. While current IrDA standards focus on the replacement of the point-to-point serial/parallel cables that connect computers to peripherals [2], future activities of the IrDA will focus on multipoint protocols as are used in LAN systems.

The IEEE 802.11 committee allowed the definition of multiple PHY layers, in part, because the members of the committee had some interest in each of the aforementioned PHY layers and hence they sought to accommodate all of them. The benefit of this approach is that the various advantages of each of the PHY layers can be exploited by users who want an 802.11 compliant wireless LAN (e.g., see [3]). The

[3]The 802.11 PHY standards specify the 2 Mb/s data rate as being optional, but all PHYs are required to support the lower 1 Mb/s rate. Further, provisions are made to allow both speeds to coexist in the same channel.

disadvantage of this approach is that two users need to specify additionally the type and data rate of their wireless LAN system to permit interoperability, for example, an 802.11 FH 1 Mb/s system. Thus, the advantages of interoperability that we experience with say wireline modem technology are lost as is the cost advantage of large volumes that would accompany the choice of a single PHY layer.

In FH systems, the frequency at which the data is transmitted is varied among a set of frequencies (i.e., 79 frequencies in the U.S./European version of the 802.11 standard, and 23 in the Japanese version). That is, the transmitter sends data on a given frequency for a fixed length of time (i.e., the *dwell time* in 802.11) and then switches to the next frequency for another fixed length of time. The frequency hopping pattern is known to the receiver so that the receiver's frequency synthesizer can hop in synchronism and recover the original data signal. The FH systems defined in the 802.11 PHY are slow frequency hopping systems since they transmit multiple consecutive symbols at the same frequency. In FH systems, adjacent or overlapping cells (i.e., BSSs) use different hopping patterns. For hopping patterns with many frequencies (e.g., 79 in the U.S./European 802.11 standard), it is unlikely that the same frequency will be used at the same time by two adjacent cells. The Jan. 1996 draft standard specifies three different sets of hopping patterns, each of which is composed of 26 patterns (i.e., 26 logical channels). The patterns within a given set have been chosen to exhibit good properties, e.g., the consecutive frequencies in a given pattern are spectrally separated by at least 6 MHz so as to avoid a narrowband interferer.

In DS systems, the original data signal is modulated by a wideband spreading signal. This spreading signal is known to the receiver, which can then recover the original data signal. Note that in the 802.11 DS PHY, unlike multi-code CDMA systems, only one predefined spreading signal is used. The factor by which the bandwidth of the signal is expanded is known as the *processing gain* of the DS system. In 802.11, the processing gain is 11 (10.4 dB), which permits some resilience to narrowband noise and permits the 83 MHz U.S. band to be segmented into a few channels (i.e., 11 DS center frequencies are defined in 802.11 for the U.S., but only three of these channels can be used without overlap).

In summary, we note that since a FH system can offer a larger number of channels (i.e., frequency-hopping patterns) than a DS system, a FH system may be more useful for dense environments in which cells have overlap with many adjacent cells.

Further, FH and DS systems have somewhat different types of resilience to narrowband interference. FH systems experience the interference only for a fraction of time whereas DS systems experience a fraction of the interference power all of the time. Thus, FH systems have the performance advantage if the interference is high and DS systems have the advantage if the interference is low. Currently, both types of radio systems, FH and DS, have some manufacturers backing them. It remains to be seen whether there will be a market winnowing to a dominant PHY layer or whether both types of PHYs will maintain significant market shares. Both of these types of radio systems aim to transmit at power levels of 100 mW or less, which will enable them to achieve ranges of up to 100 m indoors, depending upon the data rate and the building geometry and composition.

2.3 Medium Access Control

The IEEE 802.11 draft standard defines a single Medium Access Control (MAC) protocol for use with all of the aforementioned physical layers. The use of a single MAC protocol better enables chip vendors to achieve high-volume production, which will help keep the costs low for these systems. There was considerable debate and compromise preceding the adoption of the current 802.11 MAC protocol. The MAC protocol defined in the 802.11 draft is sophisticated and entails considerable complexity. The protocol has a few options, as well as several features that can be turned on and off, and combines most of the functionality that was contained in the dozen or so MAC proposals that were considered by the committee [4].

The important characteristics of the 802.11 MAC protocol, which are likely to remain unchanged in the final standard, are its ability to support: 1) the access point oriented and ad hoc networking topologies, 2) both asynchronous and time critical traffic (called *time-bounded services* in 802.11), and 3) power management. The primary access method (called the Distributed Coordination Function (DCF)) that is used in the protocol is drawn from the family of Carrier-Sense Multiple Access/Collision Avoidance (CSMA/CA) protocols. Since the radio medium does not permit the use of a Collision Detection (CD) mechanism, as is used in the CSMA/CD protocol of Ethernet, the CSMA/CA protocol uses a random backoff to reduce the likelihood of two frames colliding. Collisions are most likely to occur during the time period immediately following the transmission of some frame, since two or more sta-

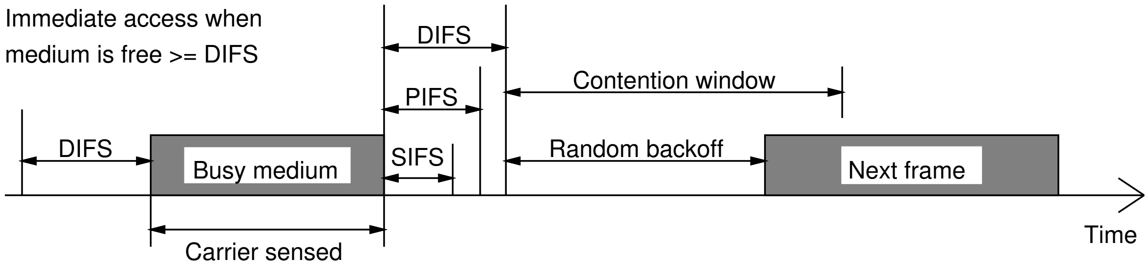

Figure 2: Primary access mechanism.

tions may be listening to a busy medium and hence transmit when it becomes free. In the CSMA/CA protocol of 802.11, the random backoff time is distributed according to a uniform distribution (in discrete slot times) where the maximum extent of the uniform range is called the *Contention Window (CW)* in 802.11. The CW parameter, i.e., the range of this uniform distribution is doubled (up to a maximum limit) each time a frame transmission is unsuccessful, as determined by the absence of an acknowledgment (ACK) frame. This exponential backoff mechanism helps reduce collisions in response to increasing numbers of contending stations. Further, as shown in Fig. 2, there is an initial Interframe Space (IFS) that can take on three different values representing priorities for transmission. The highest priority frames are transmitted using the Short IFS (SIFS). For example, the immediate acknowledgment that a receiving station sends back to the transmitting station makes use of the SIFS to guarantee that no other station intervenes. The next longest IFS, the Point coordination function IFS (PIFS) is used to provide a priority mechanism by which time critical frames can be transmitted before asynchronous data frames, which use the longest IFS, the Distributed coordination function IFS (DIFS).

In radio systems that depend on the physical sensing of the carrier, a problem arises (called the *hidden node* problem [5]) in which a single receiving station can hear (i.e., is in radio range of) two different transmitters, but the two transmitters cannot hear the carrier signals of one another. In this type of topology, the transmitters send frames without performing a random backoff (because the carrier signal of the other transmitter is never heard). This results in a high likelihood of collision. The 802.11 MAC protocol includes, as an option, a well-known mechanism to solve this hidden node problem. The protocol makes use of two control frames: 1) a Request To Send

(RTS) frame that a potential transmitter issues to a receiver, and 2) a Clear To Send (CTS) frame that a receiver issues in response to a transmitter's RTS frame. The CTS frame grants the requesting station permission to transmit while at the same time notifying all of the stations within radio range not to initiate any transmissions for a given time, which is called the Net Allocation Vector (NAV) in 802.11. Because of the signalling overhead that is involved, the RTS/CTS feature is not used for short packets for which the collision likelihood and cost (in terms of retransmission time) are both small anyway.

In order to support time-bounded services, the 802.11 standard specifies the optional use of the aforementioned Point Coordination Function (PCF) in which a point coordinator (or PCF station)[4] has priority control of the medium. That is, when PCF is active, the PCF station allows only a *single* station in each cell to have priority access to the medium at any one time. This is implemented through the use of the previously mentioned PIFS and a beacon frame (see Fig. 3) that notifies all[5] of the other stations in the cell not to initiate transmissions for the length of the Contention Free Period (CFP). Having silenced all of the stations, the PCF station can then allow a given station to have contention free access through the use of an (optional) polling frame that is sent by the PCF station. Note that the length of the contention free period can vary within each CFP repetition interval according to the system load. A typical wireless LAN installation would use different channels for adjacent cells so as to prevent two PCF stations (i.e., access points) from using (and hence colliding) on the same channel during the contention free period. This would allow coexistence, even on the same channel, with an ad hoc network that is using DCF only (see Fig. 1).

Most of the devices in which the 802.11 standard will be used have power limitations (e.g., small handheld personal digital assistants), so options for power conservation were included in the MAC protocol. When a station is in the power saving mode (i.e., the doze state) it cannot transmit or receive frames, however, it does keep some timers operating. The 802.11 standard defines power management procedures for cases with and without infrastructure (i.e., access points). In the presence of infrastructure, a dozing station periodically wakes up and listens to selected beacons

[4]The PCF station is always an access point, so the use of the PCF and hence support for time-bounded services is limited to networks with infrastructure.

[5]If one of the stations does not hear the expected beacon, it sets its NAV to a known maximum value for the length of the CFP.

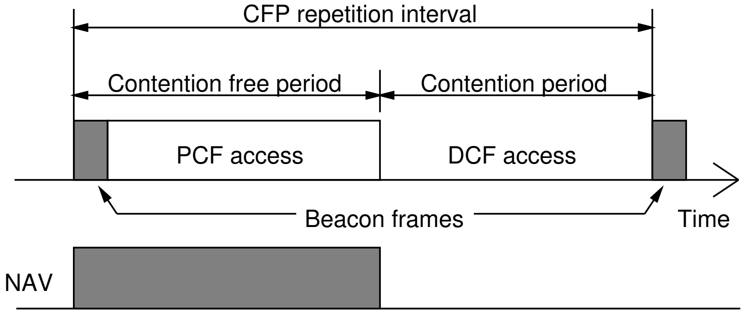

Figure 3: Alternation of contention free and contention periods.

that are sent by the access point. If the station hears a control frame indicating that the access point has queued data for that station, the station sends a special poll frame that tells the access point to send the data. In the absence of infrastructure, the power conserving stations in the ad hoc cell wake-up for only short predefined periods of time to hear if they should remain on so as to receive a frame.

A final issue to consider for a wireless LAN standard is that of security to guarantee both privacy of the wirelessly transmitted data and to verify the authenticity of the wireless station or user. The 802.11 draft standard specifies an (optional) data encryption algorithm called the Wired Equivalency Privacy (WEP) algorithm. The WEP algorithm is based on the RC4 PRNG algorithm developed by RSA Data Security, Inc. [6]. The 802.11 standard describes a couple of mechanisms for supporting authentication, however, the *Shared Key* mechanism is the only one that is fully defined at this time. As its name suggests, in this mechanism, the authentication of stations/users is based on the communicating stations having knowledge of a shared secret key.

3 HIPERLAN

The European community decided to pursue the goal of a wireless LAN that would be indistinguishable in performance from wired LANs such as Ethernet, and also have some support for isochronous services. A committee was set up in 1991 under the auspices of the European Telecommunications Standards Institute (ETSI) to formulate a high-performance radio LAN (HIPERLAN) standard. Unlike the IEEE 802.11

standard, this committee was not driven by existing products or regulations. A set of functional requirements was defined and the committee set out to satisfy the requirements. The standards work was confined to the lowest two OSI layers [7]. A draft standard [8] was released in July 1995 for imminent ratification. The high bit-rate requirements, coupled with the low-power requirements for safety and other reasons, imply that each radio will have a short range (10-100 m). Scenarios for usage and the choices considered for different aspects of the standard are described in [9, 10]. In brief, the standard allows for a radio LAN system operating at 23.529 Mb/s with support for multihop routing, time bounded services, and power saving.

The high data rate, together with the need for a number of channels require a reasonably large amount of spectrum, on the order of 150 MHz or more. The committee identified two bands, from 5.15-5.30 GHz and 17.1-17.2 GHz. Currently, the standard addresses mainly the 5 GHz band, which has been ratified[6] for HIPERLAN use by the Conference of European Posts and Telecommunications Administration (CEPT). The band is divided into five channels, with the lower three channels available in Pan-European countries and the upper two channels only available in some countries. The channel center frequencies start at 5.176468 GHz and the center frequencies are separated by 23.5294 MHz. Gaussian Minimum Shift Keying [11] is the chosen modulation method, mainly for reducing the adjacent channel interference and for amplifier efficiency considerations. The goal is to reach packet error rates below 10^{-3}. A (31,26) BCH code [12] is used on the bulk of the data packet, interleaved across 16 codewords. This leads to a block of 416 user data bits encoded to 496 bits. The coding scheme offers protection, in the sense of error-correction per block, from at least two random errors and burst errors less than 32 bits long. Data packets consist of multiple blocks of user data. Each block has 416 bits of user data, and there are at most 47 blocks per packet. The high bit rate and proposed indoor use of HIPERLAN will require equalization to mitigate the effects of intersymbol interference. The standards define the use of a particular 450 bit training sequence in every data packet, but stop short of defining the equalizer precisely, leaving that to each implementation.

The MAC protocol is based on a carrier sensing mechanism, but is quite different in its details from that used in the IEEE 802.3 standard (Ethernet) or the IEEE 802.11 standard discussed earlier. In case the medium has been sensed to be free

[6] HIPERLAN does not have exclusive use of either the 5 GHz or 17 GHz band and can use these bands on a non-protected basis and without interfering with other users.

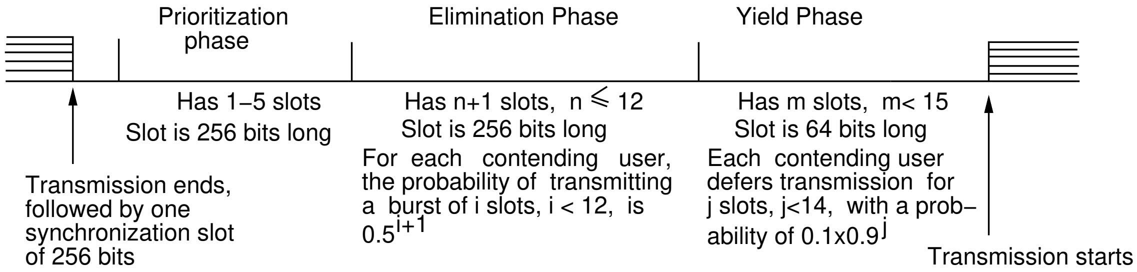

Figure 4: Channel access for HIPERLAN when the medium is sensed busy.

for a sufficient length of time, 1700 bit times in this case, immediate transmission is allowed. If not, then the channel access, in the terminology used by the HIPERLAN standard, consists of three phases: the prioritization phase, elimination phase, and yield phase. The actions of each node in these three phases are described below and in Fig. 4. The prioritization phase is aimed at allowing only nodes having packets of the highest available priority to contend further for channel access. This phase consists of a number of slots, with a node having a packet with priority p transmitting a burst[7] in slot $p + 1$ if it has heard no higher priority burst. At the end of the first burst on the channel, the prioritization phase ends and the elimination phase begins. During the elimination phase, nodes that transmitted a burst during the prioritization phase now contend for the channel. This is achieved by each node transmitting a burst for a geometrically distributed number of slots and then listening to the channel for one time slot. If another burst is heard while listening to the channel, the node stops contending for the channel. Thus, only the node(s) with the longest burst will, in the absence of the hidden node problem, be allowed to further contend for the channel. Immediately after the longest burst and listening period of the elimination phase is the start of the yield phase. In this phase, each of the surviving nodes defers transmission for a geometrically distributed number of slots, while listening to the channel. However, if they hear any transmission, they defer transmission altogether. The purpose of the elimination phase is to bring the number of contenders down to a small number, and then the yield phase tries to ensure that only one node eventually

[7] Roughly speaking, a burst consists of transmitting the carrier frequency. More precisely, there is a particular bit sequence that is repeated for the duration of a burst, but all receivers only respond to the received signal strength and not the particular bit sequence.

transmits. As a result, the chances of actual collisions for data are negligibly small (i.e., less than 3%).

The HIPERLAN technical committee wanted to explicitly support a Quality-of-Service (QoS) for packet delivery. The QoS support is provided via two mechanisms, the initial value in both cases being assigned by the application using the HIPERLAN services: the priority of a packet (high or normal), and the packet lifetime measured in integral milliseconds with a range of 0-32767 ms (default value is 500 ms). The residual lifetime of a packet together with its priority are used to determine its channel access priority. As described earlier and shown in Fig. 4, the channel access priority can fall into one of five categories, and this priority is used for the prioritization phase described above. No other explicit mechanism is used to support the desired QoS, unlike the time bounded services of the IEEE 802.11 standard. Since multihop routing is supported within the standard, the lifetime of a packet and the residual lifetime are transmitted along with the packet. Packets that cannot be delivered within the allocated lifetime are discarded. Even though the original aim of the committee was to support statistically independent rates for different traffic classes, the choice of the MAC protocol together with the support for ad hoc networks and multihop routing, allow only a *best-effort* type service.

The committee envisioned that a pure cellular architecture will not be sufficient for the system, hence allowing HIPERLAN nodes to forward packets destined for other nodes. This, of course, requires the need to maintain routing databases at nodes and the dynamic updating of these databases. Methods for this topology maintenance have also been addressed in the standard, both for the databases at each node and for broadcasting the information to other nodes. However, it is optional for a node to *forward* packets, and hence a node can also choose to forego this function, becoming a *non-forwarder* in the terminology. An interesting discussion of some of the issues involved can be found in [13].

Power saving through both hardware specific features and protocol design have been addressed in HIPERLAN. The first method of power saving is via the *p-saver* method. In essence, a node can announce that it only listens periodically, with a short duty cycle for remaining powered on. This allows the node to power down most of its circuits at all other times. Other nodes wishing to transmit to it, namely *p-supporters*, only send packets to the p-saver when they expect it to be listening. Further, since there are broadcasts and multicasts on the air, there is support for deferred multicasts.

Nodes that relay multicasts announce their schedule for doing so, allowing other nodes to power down except when they expect to hear multicasts. The final step towards power saving is through an innovative two-speed transmission method. Packets have a short low bit rate (LBR) header, at 1.4706 Mb/s, which contains enough information to inform a node whether it needs to listen to the rest of the packet or not. Thus, even if the node is listening it can keep the error-correction, equalization, and other circuits powered off unless the LBR header informs it otherwise.

There is support for packet encryption in the HIPERLAN packet transmission mechanism. The standard stays away from defining the particular encryption method used, but defines methods to inform the receiver which of a particular set of encryption keys has been used to encrypt the packet. The standard defines a small set of such keys and how they are kept at nodes. It does not, however, define any key distribution strategy, which would be a management function on top of the basic services. Another ETSI committee is working on a security standard for HIPERLAN, which will be required for conformance.

The standard clearly defines a common air interface and packet exchange mechanisms. However, there are interesting questions which will be answered only through building prototype systems and trial deployments. The first issue is that of channel selection. How will all of the nodes belonging to a logical HIPERLAN decide on a common channel? The channel access method strongly depends on carrier-sensing. What impact will this have on the hidden-node situation [5] and how will the throughput be affected in such situations? Also, since the standard assumes that all of the nodes belonging to a HIPERLAN use only one channel, what user and traffic density can be tolerated, especially for services requiring guaranteed delays? A final issue that we raise concerns power consumption for doing all the functions. This has been considered by the committee and other authors [9] and was a factor in not choosing other modulation schemes that might have different power consumption profiles than the current choices. Some efforts aimed at building HIPERLAN systems and the technological factors involved are described in [14].

4 Other Standards

In September 1993, the FCC allocated unlicensed bands for new Personal Communication Services (PCS). Subsequently, in a June 1994 ruling, the FCC reduced the allocated band for unlicensed PCS to a 20 MHz band from 1.910 to 1.930 GHz that is segmented into a 1.910-1.920 GHz subband for asynchronous applications such as wireless LANs and a 1.920-1.930 GHz subband for isochronous applications such as cordless telephones. A device operating in the asynchronous subband must follow a special Listen-Before-Talk (LBT) etiquette, which is designed to allow multiple systems to coexist in the same vicinity. It is important to note that these new PCS bands are currently occupied by point-to-point microwave links and that it may take as long as several years to fully clear these bands for wireless LAN users. The reclamation of the unlicensed PCS bands is being conducted by UTAM, Inc. (Unlicensed PCS Ad Hoc Committee for 2 GHz Microwave Transition and Management), a non-profit coalition of equipment manufacturers. UTAM intends to collect fees from the manufacturers of unlicensed PCS equipment in order to fund the relocation of the current microwave users. In addition, the FCC recently allocated (early 1995) an unlicensed Data PCS band from 2.390-2.400 GHz in which the aforementioned LBT etiquette also must be used. The Data PCS band is currently clear, except for some government users that are authorized on a secondary, unprotected basis.

While the Europeans (i.e., CEPT) have ratified the use of spectrum for high-bandwidth wireless applications (e.g., 5.150-5.300 GHz for HIPERLAN, as described in the previous section), the regulatory situation in the U.S. is still evolving. In mid-1995, the FCC received two proposals for future allocations near 5 GHz. The Wireless Information Networks Forum (WINForum) submitted a proposal called the Shared Unlicensed PErsonal Radio Network (SUPERNET), which requested 250 MHz of spectrum to support multimedia computer applications of up to 20 Mb/s. This proposal requested the band from 5.100 to 5.350 GHz. A second proposal, submitted by Apple Computer Inc., with support from some other companies, requested a total of 300 MHz of spectrum, in two bands, to support National Information Infrastructure (NII) wideband applications. The so-called *NII Band* proposal requests the bands from 5.150 to 5.300 GHz (i.e., the HIPERLAN band) and from 5.725 to 5.875 GHz. Both of these proposals plan to adopt aspects of the HIPERLAN standard.

In addition to these future directions, efforts are underway to demonstrate wireless systems that are designed specifically for operation with Asynchronous Transfer Mode (ATM) networks. One example of these efforts is the recently initiated work of a European project called the Wireless ATM Network Demonstrator (WAND). The WAND project was formed in answer to a call, from the European Union, in the Advanced Communication Technologies and Services (ACTS) program that was launched in 1995. The project is run by a consortium of six European communications and computer companies. The WAND project aims to research and demonstrate the feasibility of using ATM over a high-speed radio interface and seeks to achieve a data rate similar to that of HIPERLAN (i.e., around 20 Mb/s). The HIPERLAN effort has itself been extended into a family of standards, with the one described in the last section being the first one to complete. A second HIPERLAN standard will be aimed at mobile wireless ATM (in the 5 GHz band), and its standardization efforts will be symbiotic with the WAND activities. Other HIPERLAN standards will be aimed at higher speeds and at the 17 GHz bands.

5 Mobile Networking

Using wireless network interfaces, mobile devices can be connected to the Internet in the same way as desktop machines are connected using Ethernet, Token-Ring or point-to-point links. The major difference, however, is that mobile devices can move while in operation, which means that their point of attachment to the network can change from time to time. From a network's viewpoint, host movement constitutes a change in the network topology. It is natural that mobile users desire uninterrupted access to all networking services even while moving. Unfortunately, neither the Internet protocol suite nor the OSI network architecture can provide this functionality. The assumption that end systems are stationary lies at the very foundation of the Internet and OSI network architectures. This is a serious problem, since it is not possible to deploy a new *mobility-aware* protocol stack in the Internet, which already consists of tens of millions of hosts. The challenge lies in finding a solution that allows mobile nodes to function efficiently within the Internet architecture without requiring modifications to the existing infrastructure and host software.

Over the past three years, many proposals have been made for supporting host mobility on datagram-based internetworks [15, 16, 17, 18, 19]. The vast majority of these proposals have been designed to be compatible with today's TCP/IP-based Internet. Except for the scheme proposed in [19], which operates at the link layer, the rest of the proposals provide support for mobile networking at the network layer. To consolidate these efforts, the Internet Engineering Task Force (IETF) has created a Mobile IP working group to come up with a standard for near term deployment within the Internet. The proposed modification to the Internet Protocol (IP), which is referred to as Mobile IP [20], enables mobile nodes to change their network attachment points without disrupting any active network sessions. The key feature of the Mobile IP design is that all required functionalities for processing and managing mobility information are embedded in well-defined entities, called Home Agent (HA), Foreign Agent (FA), and Mobile Node (MN) (see Fig. 5). The new functions defined by the standard allow a mobile node to roam on the Internet, without changing its IP address. Since Mobile IP exploits existing mechanisms available within IP, it is completely transparent to the transport and higher layers and does not require any changes to existing Internet hosts and routers.

The Internet routing system routes a datagram to a host based on the network number contained in the node's Internet address. If a node changes its point of attachment and moves to a new network, IP datagrams destined for it can no longer be delivered correctly. The Mobile IP solution allows mobile nodes to retain their addresses regardless of their point of attachment to the network. When the mobile node visits a foreign network, it is associated with a *care-of-address*, which is an Internet address associated with the mobile node's current point of attachment. The care-of-address either identifies the mobile host directly (if the address is acquired through Dynamic Host Configuration Protocol (DHCP) [21]) or identifies a Foreign Agent that is responsible for providing access to visiting mobile nodes. The Home Agent, which is located at the mobile node's home network, maintains the binding between the mobile node and its care-of-address. When away from home, the mobile node registers its care-of-address with the Home Agent; the Home Agent is responsible for intercepting datagrams addressed to the mobile node's home address and tunneling (encapsulating) them to the associated care-of-address. The Foreign Agent decapsulates the incoming packets and relays them to the mobile node.

In this scheme, all datagrams addressed to a mobile node are always routed via the Home Agent. However, the packets in the reverse direction, i.e., those originating from the mobile node and addressed to a stationary host, are relayed along the shortest path by the Internet routing system. This gives rise to what is known as the triangle routing problem. Route optimization is possible if the location information (the association between the mobile node and its care-of-address) is allowed to be cached at the stationary host [22]. The stationary host can use it to directly tunnel traffic to the care-of-address. Unless the location information is properly authenticated, there

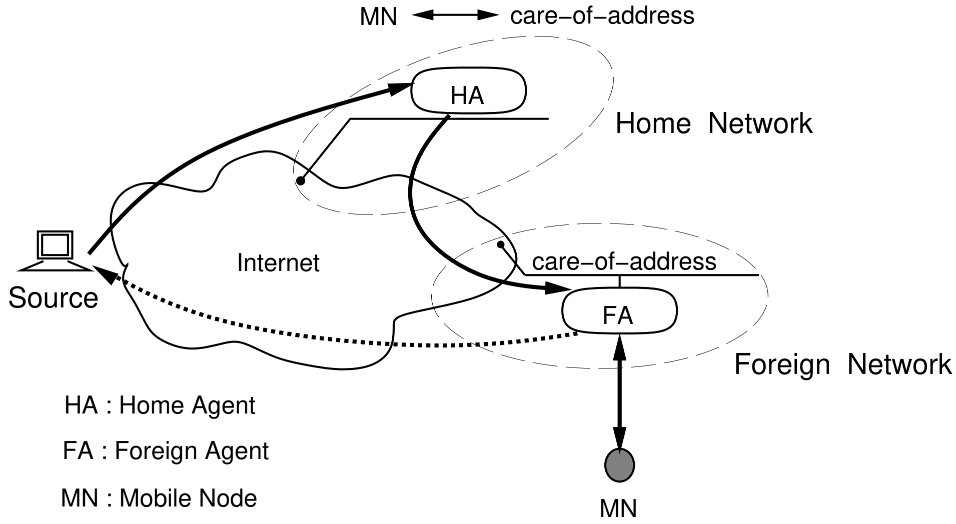

Figure 5: Mobile IP architecture components.

is a potential security risk involved in performing route optimization. Currently, there is a disagreement within the Mobile IP working group on whether it is possible to support such an authentication mechanism within the existing Internet. Therefore, the current Mobile IP proposal does not permit route optimization.

The Mobile IP working group of IETF, which was formed in the summer of 1992, is now in the final stages of releasing a standard RFC. The working group is now focusing on defining an architecture for supporting mobility within Internet Protocol version 6 (IPv6). Since the standards and protocols for IPv6 are still evolving, and since there is no existing installed base of IPv6 hosts or routers with which the proposed solution must be compatible, IPv6 provides a unique opportunity and an unconstrained platform for developing the next generation of mobile internetworking protocols and applications. The first working group draft [23], therefore, improves on Mobile IP design in several ways. For example, the new design does not require Foreign Agents; mobile nodes dynamically acquire a care-of-address using the IPv6 neighbor discovery protocol. Second, by making all IPv6 nodes *mobile aware*, the new protocol provides for direct tunneling of traffic to mobile node's care-of-address. The advantage of this design is that the load on Home Agents is significantly reduced and the triangle routing problem is resolved.

6 Wireless Link Management

There is a need for applications to be made aware of the characteristics of the wireless link. For example, wireless devices experience intermittent connectivity as a normal part of operating in a wireless network. A wireless user may experience a fade where the network connection is momentarily lost during a long file transfer. A mobile aware application need not react by aborting the file transfer. The mobile aware application can instead suspend its application layer time outs and notify the user of the fade condition. When the mobile link is re-established, the application can resume the file transfer [24].

Applications need to have access to status information to decide upon the optimum wireless network to use. Information such as radio link speed, battery level, network type (CDPD, Mobitex, etc.), network ID (the name of the service provider), and tariffing schedules can be used to algorithmically select the best network to run the application over.

6.1 Wireless Link Awareness

There are industry groups defining standards and specifications for providing wireless link information to protocol stacks and applications on the mobile client: the Personal Computer Communications Association (PCCA) [25], the Windows Sockets 2.0 Wireless Extensions Workgroup [26], and the group of companies defining the Windows Sockets (Winsock 2) specifications [24]. The standards and specifications are being created for the Windows 95 and Windows NT operating system environments.

As shown in Fig. 6, the PCCA is defining enhancements to the Network Driver Interface Specification (NDIS 3.1) device driver model to provide wireless specific information to wireless and mobile-enabled protocol stacks. The Windows Sockets 2.0 Wireless Extensions Group is defining the Network Device (NetDev) management interface to expose the wireless specific information to wireless and mobile-enabled applications. Winsock 2 is providing a common Application Programming Interface (API) where wireless protocol stacks (Mobitex, RD-LAP, etc.) can be selected by specific applications as well as traditional protocol stacks such as TCP/IP.

The PCCA is defining wireless enhancements to the NDIS 3.1 device driver interface to make it possible for wireless-aware transports to get status information from the wireless network adapter so that they can *tune* themselves and the wireless device

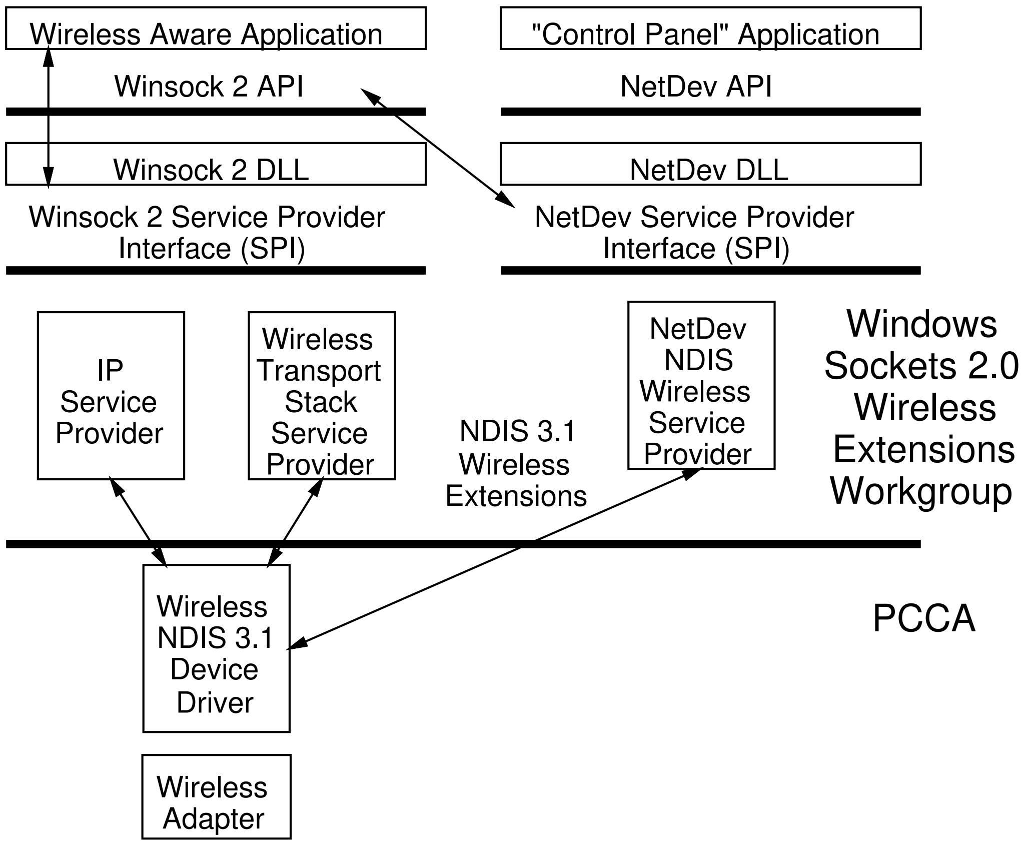

Figure 6: Structure for application and protocol stack awareness of wireless links.

for best performance. For example, the wireless-aware transport could be an existing user datagram protocol over the internet protocol (UDP/IP) stack that is modified to handle the special characteristics of a wireless network. The NDIS 3.1 enhancements are designed to also provide the functionality needed to support a wireless-aware API such as NetDev. Recently (March 1996), the PCCA issued the first version of their standard for wireless extensions to NDIS [25].

An additional benefit of providing a device driver interface to wireless devices is that multiple protocol stacks can bind to a single wireless device and simultaneously transmit and receive data over the device. This enables the wireless user to run multiple applications over multiple protocol stacks with the wireless device just like a LAN based user does today running over Ethernet or Token-Ring.

The Windows Sockets 2.0 Wireless Extensions Workgroup is defining NetDev as a high level API for applications to use to manage wireless devices. NetDev allows the calling applications to enumerate installed devices, react to plug and play events, query and set network parameters, and set network event triggers to enable asynchronous indications to be made to the applications. A Winsock 2 application uses a quality of service structure and device status information from NetDev to select the optimum wireless transport. Winsock 2 provides a mechanism that enables the application to select the specific wireless device that the wireless transport uses by mapping socket handles to NetDev device handles. The Winsock 2 quality of service structure also enables applications to be informed about a change in the network availability status when the mobile user experiences a fade or a disconnection. The application is also informed when the connection is re-established. Additionally, Winsock 2 provides a mechanism that allows individual wireless transport providers to pass up wireless link status information directly through Winsock 2 with transport specific command codes.

Table 1 shows some of the wide area wireless status information that has been defined by the PCCA. This status information is common across a number of wide area wireless network devices. Additional network specific status information has also been defined for a number of networks such as DataTac, Ardis, Mobitex, and CDPD. Though the current definitions are for wide area networks, they can easily be extended for wireless LAN systems. In this case, it is likely that only a subset of the defined status information fields will be used.

6.2 Local and Remote Management

The Mobile Management Task Force (MMTF) has issued a draft mobile management information base (MIB) [27] for remotely managing many aspects of mobile communications using the Simple Network Management Protocol (SNMP). The MMTF is a coalition of companies that are working with existing standards organizations such as the Internet Engineering Task Force and the Desktop Management Task Force (DMTF) to propose extensions and modifications to existing standards and to encourage the development of new standards as needed. The MMTF is currently examining an approach where DMTF management information formats (MIFs) are defined first, and then converted to SNMP MIBs using an algorithm standardized within the DMTF [28]. The MIFs and MIBs would include characteristics of both the mobile computing device and the mobile link for both wired and wireless mobile communications.

The use of the DMTF Desktop Management Interface (DMI) offers similar functionality to NetDev and can also supply mobile link awareness to applications and protocol stacks. Unlike NetDev, which is being specified for Windows 95 and Windows NT, the DMTF DMI is available on a variety of mobile operating systems (OS/2, DOS, Windows 3.1, AIX, Windows 95, Windows NT).

The use of SNMP over the wireless link has to be managed in an efficient manner. SNMP is very polling intensive. Left unchecked, SNMP can introduce a large number of flows over the wireless link. The wireless link is bandwidth and in many cases tariff constrained. Proxy agents can be used on the wired portion of the network to filter SNMP flows to the mobile. A proxy agent responds to SNMP requests for static information or information about the mobile link. A proxy agent passes SNMP requests directly through to the mobile that are related to dynamic system status information.

Standards bodies and industry groups that are defining open wireless protocol standards are also defining MIBs for managing the wireless communications layers. Two examples are the aforementioned IEEE 802.11 wireless LAN standard [1] and the CDPD specification [29].

Status information	*Description*
network type	The current network being used by the wireless device. Examples are AMPS, DataTAC, Ardis.
header format	Specifies the frame format passed across NDIS. Examples are DIX Ethernet frames, MPAK frames, RD-LAP frames, and MDC4800 frames.
indication request	Allows protocol stacks to register to be notified of changes in other status information.
device information	Supplies the manufacturer, model number, software version number and serial number of the wireless device.
operation mode	The wireless device's power mode. Possible values are normal mode and power savings mode.
lock status	Indicates whether the wireless device is locked or unlocked.
disable transmitter	Used to enable or disable the wireless transmitter.
network ID	Returns the ID of the network that the device is currently communicating with. An example is "XYZ Cellular Services".
permanent address	Returns the device's permanent network address.
suspend	Used to suspend or make operational the NDIS device driver. When suspended the device driver releases the serial port. This is useful if another application would like to access the device through the same serial port that the NDIS device driver was using.
base station ID	Returns the ID of the base station that the device last contacted. Base station ID can also be monitored to determine when a hand-off has occurred.
channel ID	Returns the ID of the channel currently in use.
channel quality	Supplies the connection quality of the wireless link between the wireless device and the network. Also denotes whether the wireless device is in or out of range.
registration status	Indicates whether the mobile has registration pending, registration denied, or registered.
radio link speed	Returns the radio link speed in bits/second for the current network.
battery level	Returns the current battery level and whether or not external power is connected to the wireless device.

Table 1: Wide area wireless status information defined by the PCCA.

7 Conclusion

Fulfilling the promise of wireless LANs, i.e., the convenience of tetherless access and the maintenance of network sessions for mobile clients, has impacts across all network protocol layers. We have described some of the activities that impact some of these layers in the previous sections. Of course, the impact is greatest in the lowest two layers, the physical and data link layers, since the wireless medium is quite different from the traditional wired media. It is also in these two layers that the technical community has spent the greatest effort in the pursuit of wireless networking. Many advances have already been made, but in conclusion we point out that the work in this area is far from done. Two techniques that have offered great benefits in the wide area wireless networking arena are smart antennas and coded modulation schemes. An instance of smart antennas is the use of antenna diversity [30] to alleviate fading and other channel effects. Further use of smart antennas can be in both the transmission and reception of directional radio signals to improve the signal to interference ratio [31]. Recent advances in modulation, such as coded modulation which is used in telephone modems, can also be used to advantage in local area wireless systems. To date, the prohibitive signal processing requirements have not allowed sophisticated coded modulation schemes to be used at the high bit rates of indoor wireless systems, but these prohibitions are expected to lessen with technological advances. In this case, schemes such as those described in [32] should be considered for use in wireless LANs.

The next two layers of the networking stack, the network and transport layers, are also impacted by wireless/mobile networking. Note that the network layer, such as IP, is used to *glue* together disparate physical media and applications. As such, the desire to interoperate with existing hosts on the Internet is going to allow only incremental changes in this layer, as opposed to the development of a fresh new standard, which was possible in the lower two OSI layers. There has also been much work done in understanding the impacts of both wireless and mobility on the transport layer and on flow control for data networks [33, 34, 35, 36]. However, this work is still ongoing and is yet to enter a standardization phase. Once we are past the network and transport layers and the incremental changes therein, interoperability is more or less guaranteed. Applications and network management will have to evolve in the future, as discussed in the last section, to both take full advantage of the mobility offered and also be aware of the wireless link.

Acknowledgments

We thank the reviewers for their careful reading of the paper. Their many useful suggestions have greatly improved the accuracy and clarity of this paper.

References

[1] *IEEE P802.11D3: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.* Piscataway, NJ: IEEE Standards Dept., Jan. 1996.

[2] *Infrared Data Association Serial Infrared Link Access Protocol (IrLAP), Version 1.0.* Infrared Data Association. (see http://www.irda.org/irda/content2.html).

[3] L. Goldberg, "Wireless LANs: Mobile computing's second wave," *Electronic Design*, vol. 43, pp. 55–72, June 26, 1995.

[4] C. Links, W. Diepstraten, and V. Hayes, "Universal wireless LANs," *Byte*, vol. 19, pp. 99–108, May 1994.

[5] H. Ahmadi, A. Krishna, and R. O. LaMaire, "Design issues in wireless LANs," *Journal of High-Speed Networks (JHSN)*, vol. 5, no. 1, pp. 87–104, 1996.

[6] B. Schneier, *Applied Cryptography.* New York, NY: J. Wiley, 1994.

[7] A. S. Tanenbaum, *Computer Networks.* Englewood Cliffs, NJ: Prentice-Hall, second ed., 1988.

[8] ETSI TC-RES, *Radio Equipment and Systems (RES); HIgh PErformance Radio Local Area Network (HIPERLAN); Functional Specification.* ETSI, 06921 Sophia Antipolis Cedex – France, July 1995. Draft prETS 300 652.

[9] T. Wilkinson, T. G. C. Phipps, and S. K. Barton, "A report on HIPERLAN standardization," *International Journal of Wireless Information Networks*, vol. 2, pp. 99–120, April 1995.

[10] G. Halls, "HIPERLAN– the 20 Mbit/s radio LAN," in *IEE Electronics Division Colloqium on Radio Lans and Mans*, (IEE Colloqium (Digest), No. 071, Stevenage, England), pp. 1/1 – 1/8, 1995.

[11] R. Steele, *Mobile Radio Communications.* London: Pentech Press, 1992.

[12] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes.* North-Holland Mathematical Library, Amsterdam: The Netherlands: Elsevier Science Publishers B.V., 1986.

[13] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *Proc. of ACM SIGCOMM '94*, (London, U.K.), pp. 234–244, Sept. 1994.

[14] P. Minnet, "Hiperlan: a European standard for 5 GHz and 17 GHz wireless LAN," *Microwave Engineering Europe*, pp. 61–66, Oct. 1995.

[15] J. Ioannidis, D. Duchamp, and G. Q. Maguire Jr., "IP-based protocols for mobile internetworking," in *Proc. of ACM SIGCOMM '91*, (Zurich, Switerland), pp. 235–245, Sept. 1991.

[16] F. Teraoka, K. Uehara, H. Sunahara, and J. Murai, "VIP: A protocol providing host mobility," *Communications of the ACM*, vol. 37, pp. 67–75, Aug. 1994.

[17] H. Wada, T. Yozawa, T. Ohnishi, and Y. Tanaka, "Mobile Computing Environment Based on Internet Packet Forwarding," in *Proc. of Winter USENIX*, (San Diego, CA), pp. 503–517, Jan. 1993.

[18] P. Bhagwat and C. Perkins, "A mobile networking system based on internet protocol (IP)," in *Proc. of USENIX Symposium on Mobile and Location Independent Computing*, (Cambridge, MA), pp. 69–82, Aug. 1993.

[19] B. V. Patel, P. Bhattacharya, Y. Rekhter, and A. Krishna, "An architecture and implementation toward multiprotocol mobility," *IEEE Personal Communications*, vol. 2, pp. 32–42, June 1995.

[20] C. Perkins, "IP Mobility Support." Draft RFC - work in progress, Feb. 1996. (ftp://ds.internic.net/internet-drafts/draft-ietf-mobileip-protocol-15.txt).

[21] R. Droms, "Dynamic Host Configuration Protocol." RFC1541, Oct. 1993. (ftp://ds.internic.net/rfc).

[22] D. B. Johnson and C. Perkins, "Route Optimization in Mobile IP." Draft RFC - work in progress, Feb. 1996. (ftp://ds.internic.net/internet-drafts/draft-ietf-mobileip-optim-04.txt).

[23] C. Perkins and D. B. Johnson, "Mobility support in IPv6." Draft RFC - work in progress, Jan. 1996. (ftp://ds.internic.net/internet-drafts/draft-ietf-mobileip-ipv6-00.txt).

[24] *Windows Sockets 2 Protocol-Specific Annex, Version 2.0.2.* Intel, Jan. 1996. (http://www.intel.com/ial/winsock2/specs.htm).

[25] *PCCA STD-201, Extensions to NDIS for Wireless WANs, Revision 1.0.* Personal Computer Communications Association, March 1996. (ftp://ftp.airdata.com/pub/pcca/std-201.doc).

[26] *NetDev API for Win95 and WinNT.* Windows Sockets Wireless Extensions Workgroup, June 1995. (ftp://ftp.intel.com/pub/winsock2/netdev/netdev4.doc).

[27] "Mobile MIB draft mibs." Mobile MIB Task Force, (http://www.epilogue.com/mmtf/mmtf.html).

[28] S. Bostock, *DMTF SNMP to DMI Mapping Standard.* Novell, Inc., July 1995. RFC 950713 (see http://www-uk.hpl.hp.com/people/arp/dmtf/home.htm and snmp/draft/snmp02.txt in ftp server at ftp.dmtf.org).

[29] *Cellular Digital Packet Data System Specification - Release 1.1.* CDPD Forum, Inc., Chicago, IL, Jan. 19 1995.

[30] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Transactions on Communications*, vol. 42, pp. 1740–1751, February/March/April 1994.

[31] G. G. Raleigh, S. N. Diggavi, V. K. Jones, and A. Paulraj, "A blind adaptive transmit antenna algorithm for wireless communication," in *Proc. of IEEE In-*

ternational Conference on Communications '95 (ICC), (Seattle, WA), pp. 1494–1499, June 1995.

[32] F. Abrishamkar and E. Biglieri, "Trellis-coded CPM for wireless communications," in *Proc. of IEEE International Conference on Communications '93 (ICC)*, (Geneva, Switzerland), pp. 1439–1443, May 1993.

[33] H. Balakrishnan, S. Seshan, and R. H. Katz, "Improving reliable transport and handoff performance in cellular wireless networks," *Wireless Networks*, vol. 1, pp. 469–481, Dec. 1995.

[34] P. Bhagwat, P. Bhattacharya, A. Krishna, and S. K. Tripathi, "Enhancing throughput over wireless LANs using channel state dependent packet scheduling," in *Proc. of IEEE INFOCOM '96*, (San Francisco, CA), pp. 113–1140, March 1996. Also, IBM Research Report, RC 20093, 1995.

[35] R. Cáceres and L. Iftode, "Improving the performance of reliable transport protocols in mobile computing environment," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 850–867, June 1995.

[36] A. DeSimone, M. C. Chuah, and O.-C. Yue, "Throughput performance of transport-layer protocols over wireless LANs," in *Proc. of GLOBECOM '93*, pp. 542–549, Dec. 1993.