WIRELESS COMPUTER COMMUNICATION IN LOCAL NETWORKS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of
Dokuz Eylül University
In Partial Fulfillment of the Requirements for
the Degree of Master of Science in Electrical and Electronics
Engineering, Program

by

Haldun ULUGÜN

July, 2002

İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We certify that we have read this thesis and **"WIRELESS COMPUTER COMMUNICATION IN LOCAL NETWORKS"** completed by **Haldun ULUGÜN** under supervision of **Asst.Prof.Dr. Zafer DİCLE** and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Yrd. Doç. Dr. Zafer DİCLE

Supervisor

Prof. Dr. Mustafa GÜNDÜZALP

(Committee Member)

Doç. Dr. Yalçın ÇEBİ

(Committee Member)

Approved by the
Graduate School of Natural and Applied Sciences

Prof.Dr. Cahit HELVACI
Director

ACKNOWLEDGMENTS

I would like to thank my supervisor Asst.Prof.Dr. Zafer DİCLE for his valuable guidance and support during the course of this thesis.

Haldun ULUGÜN

ABSTRACT

In this thesis we explore some of the unique idiosyncrasies of the physical layer and MAC sublayer when implemented in wireless LANs. Also investigated is the current IEEE 802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer Specification standard. This standard is used primarily for examples of possible solutions to some of the identified problems associated with wireless networks. Key differences between the operation of a wired network and a wireless network are examined. And also we gave an initial security analysis of the IEEE standard for wireles LANs. The intent of this thesis is to leave the reader with an understanding of some of the obstacles that exist when engineering a wireless network and how these obstacles can be surmounted.

ÖZET

Bu tezde Fiziksel katmanın ve MAC alt katmanının kablosuz yerel ağlara uyarlandığında bazı kendilerine özgü mizaçlarını inceledik. Ayrıca şu anki IEEE 802.11 Kablosuz Yerel Ağlara ait olan bu katmanların tanımlama standartlarına da değindik. Bu standart temel olarak Kablosuz Yerel Ağlarla ilgili tanımlanmış bazı problemlerin olası çözümlerine örnektir. Kablolu ve kablosuz ağların çalışmasındaki temel farklar incelenmiştir. Ve kablosuz yerel ağların çalışmasındaki temel güvenlik ilkelerini anlattık. Bu tezin amacı kablosuz ağların geliştirilme süreci içinde karşılaşılan engelleri ve bu engellerin nasıl üstesinden gelinebildiğinin okuyucunun anlamasının sağlanmasıdır.

CONTENTS

Chapter One
INTRODUCTION

Chapter Two
PHYSICAL LAYER

Chapter Three
MEDIUM ACCESS CONTROL

Chapter Four

SECURITY

Chapter Five

CONCLUSION

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Wireless communication or communication without the need for physical contact between sender and receiver, is almost as old as man himself. From the first signaling of a tribal drum or the first smoke signal, man's desire to communicate over distances effortlessly has motivated many a communication methodology. As man evolved, different methods of wireless communications were developed. From the first wireless telegraph to today's wireless telephone the push ofcommunications technology has been to un-tether the user. Likewise, wireless computer networking can be seen as the next step in the evolution of computer network access.

In this thesis, the issues concerning wireless networks are discussed in the context of indoor local area network (LAN) applications. The issues concerning satellite networks and low earth orbit systems are not addressed. However many of the principles discussed in this thesis do apply to these systems as well. The main purpose of this introductory chapter is to provide an overview of the work that follows. The first section discusses the motivation and possible implementations of a wireless network. This is followed by a discussion of why wireless transmission is the correct solution for the models discussed in section 1.1. Next, the chapter focuses on a quick review of the network layers most affected by switching to a wireless medium. This is followed by a discussion of the desired features and traits in a wireless network leading into a discussion of network architecture issues. Finally, the chapter concludes with a discussion of the main wireless network concepts.

1.1 Motivation for wireless versus wired

Convenience is often cited as the primary motivation for using a wireless medium instead of a wired alternative. Although, with may networked multimedia software packages, transmission speed is a pressing concern, wireless networking usually involves the networking of portable computers. These portable units are traditionally slower than their stationary cousins. As a result, the megabit ranges achievable on a wireless medium fall well within the needed transmission rates.

However, convenience alone does not completely answer the question of why the choice would be a wireless instead of a wired medium. Many have argued that users have grown accustom to having to wire a computer network and will not perceive the need to physically wire their portable units as an inconvenience. Others, however, have pointed out that society is becoming increasingly accustomed to a wide range of wireless conveniences such as cellular phones and the next logical step is wireless networking of portable computers. The question that still remains, however, is whether or not there are sufficient applications or situations where users would require wireless as their means of network access.

One of the cases where a wireless medium could be considered advantageous is when used in conjunction with a meeting or presentation. Specifically, imagine a group of executives gathered together in a conference room. Most executives currently own portable computers and many will in fact have these computers with them. Imagine one of these executives making a presentation but instead of using charts or overheads, he displays the same graphics on all the portable computers in the room. Now the presenter can make his presentation interactive by utilizing technologies such as white boards and allowing all members of the meeting to contribute more fully. But these executives do not need to be located strictly in a conference room. They could be traveling together on a plane, train, ship, bus, or any form of public transportation. They could be in a hotel, or a restaurant, or other public establishment. They need not even be executives. They could be a group of students working together on a project or a group of friends playing an interactive computer game. The list is virtually endless.

Of course, wireless access need not be restricted to just a group of people interacting via their portable computers. Imagine the convenience of being able to access a higher level network via a portable device. Specifically, the executive who does work not only at the office but also utilizes home resources. Instead of maintaining files via disks carried back and forth and being concerned with which unit has the most current copy, they could use their portable computer to effortlessly access the company's network and continue working. Or imagine this same person watching television and spending time with their family but also wanting to efficiently utilize the time by also working on their computer. Or even the children wanting to play a computer game while watching their favorite show. Wireless networking would offer the ability to access files and programs stored on a desktop unit via a portable unit from the comfort of the living room.

But what of the desk top computers? Would these also benefit from wireless connections? There are many cases when networking a group of desk top computers is best accomplished with a wireless medium. In many of today's offices, the office layout is dynamic. Office structures are changed rather quickly with the movement of a few cubicle walls. No longer does it take days to restructure the interior of an office building, but merely hours. A network that can be broken down and reestablished just as swiftly would definitely be an advantage. But the ease of network relocation need not be limited to just the changing internal topology of an office structure. Wireless networking offers the ability to swiftly establish temporary offices to meet the demands of a rapidly changing society or even to continue operation during an emergency situation [Bates, R.J. (2000)]. Using wireless eliminates the need to invest in cables and contractors each time the network is moved eliminating an expense associated with a wired computer network alternative.

1.2 Protocol layers involved

The architecture of a wireless network requires the careful integration of the various pieces into one unified and transparent system. Ideally, a change in the physical medium would only require changing the physical layer. Limiting changes to the physical layer would allow the use of differing transmission mediums to become transparent to the layers above. However, this is rarely the case. With wireless networking, the idiosyncrasies of the transmission medium affects all three of the lower network layers (for network layered structure, refer to figure 1.1). An understanding of the lower three protocol layers and their general functions is necessary before understanding why changes are needed. Although this thesis only covers the engineering issues of the physical layer and MAC sublayer, a brief discussion of the three bottom layers follows in order to provide some insight into other possible concerns.

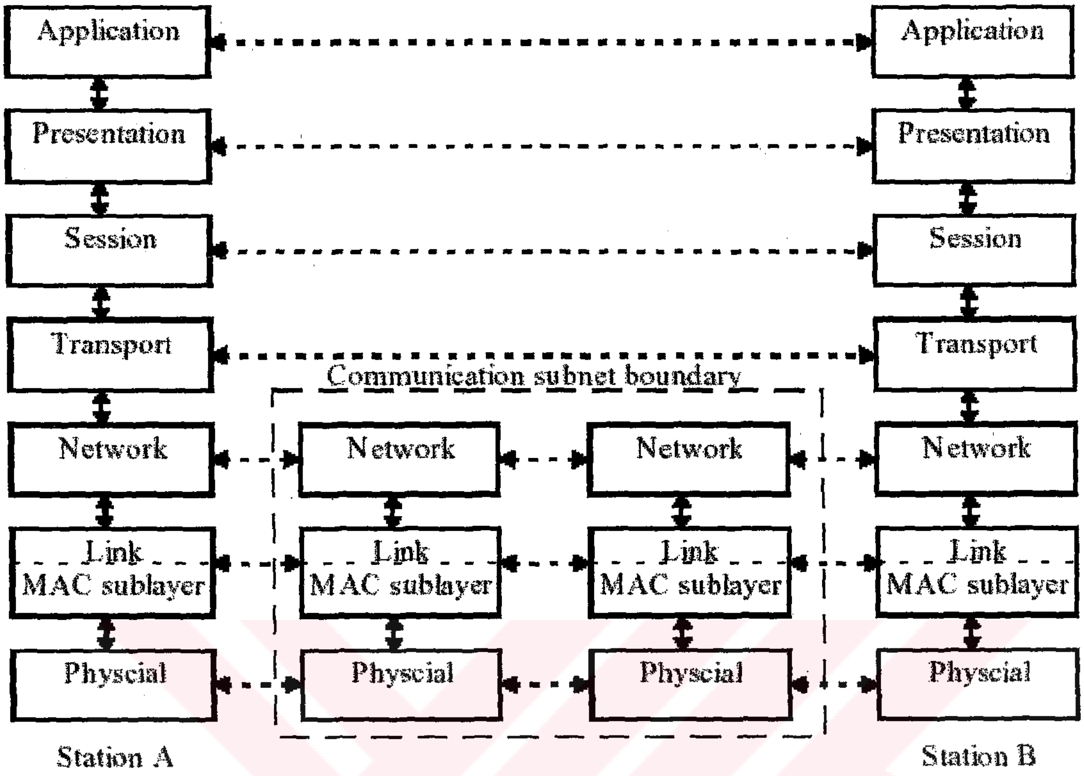


Figure 1.1: Network layer architecture

1.2.1 Physical layer

As pointed out by Andrew Viterbi, the physical layer is the most important layer in network design, and its importance is often overlooked by the network architect . Basically, the physical layer is concerned with the transmission of the individual bits across the chosen medium and the assurance, to within a reasonable measure, that they do arrive at their destination correctly . Design issues within this layer deal with the choice of signaling medium, signaling techniques , transmitter design and cost, receiver design and cost, technology limits on speed and accuracy, and especially in wireless networks, the physical environment in which the network will be used. Desired qualities of any design include the efficient use of the available resources, the minimization of channel noise [Viterbi, A. (1997)], and the acceptable minimum data transmission speed that the market will allow. Many of these issues and how they are affected by a switch to a wireless medium, are covered in detail in chapter 2 of this thesis.

1.2.2 MAC sublayer

The Medium Access Control (MAC) sublayer is primarily concerned with controlling the channel provided by the physical layer. In any multiple-access system, it is important to have some form of control over the limited resources thereby ensuring fair access. The MAC sublayer must reduce, if not completely eliminate, the chance of a data collision on the channel . To effectively and yet still efficiently divide the channel amongst the stations accessing it, the MAC sublayer should provide some form of arbitration as well as some form of allocation ensuring fairness and maximizing throughput [Chen, K. (1998)].

As will become more apparent in chapter 3 of this thesis, wired MAC protocols previously used for the allocation of resources do not easily translate into effective methods for accessing a wireless medium. For instance, the idea of trying to implement some form of token passing in an environment where the links between stations can be considered extremely dynamic and subject to failure requires overhead that could completely eliminate any benefits. These and other MAC design issues are currently being addressed by the IEEE standards board. The current standards from IEEE concerning MAC layer protocols as well as other issues and concerns of this sublayer, are covered in chapter 3 of this thesis.

1.2.3 Link layer

The primary purpose of the link layer is to recover from transmission errors that may have occurred while transmitting over the lower physical layer. This means that the link layer must provide the appearance of an error free transmission medium to the network layer above it . Achieving this level of reliability while keeping redundant data transmissions to a minimum, requires careful attention to not only the types of error correcting codes used but also the type of retransmission scheme employed. The design of the link layer requires a careful integration of flow and error control [Tanenbaum, A.S. (1998)]. Some of the error handling strategies used in a wired network do not translates well into a wireless network architecture. Although these design issues are not discussed in this thesis, they are still key issues in the architecture and engineering of a wireless network.

1.2.4 Network (IP) layer

The main purpose of the network layer is the correct routing of packets. This requires the network layer to provide the appearance of address consistency to the transport layer, even if the computer being addressed is changing its point of network attachment This layer must also manage the routes used, even when crossing heterogeneous networks [Tanenbaum, A.S. (1998)]. This requires a form of congestion control and a way of selecting optimum routes so that the routed packets are not lost or needlessly delayed. Also the Internet protocol (IP) addressing schemes and routing methods have to be adjusted to accommodate the needs of mobile computing. Currently, an IP address refers to a physical location which may not be accurate once the point of attachment of the unit changes. Several methods of handling the problems associated with mobile IP have been proposed and a committee of the Internet Engineering Task Force (IETF) is currently working on protocol specifications, referred to as an RFC, which will handle these problems .

1.3 Desired traits or features in a wireless local area network

A look at the desired features in a wireless local area network will provide further insight into the architectural issues of this form of networking. Some of the features that would be beneficial in a wireless network can be derived by looking at the possible utilization of these networks. In all of the scenarios mentioned in section 1.1, there are certain features which stand out as common. Namely there is a need for easy set-up or establishment of the network requiring little or no user interaction. This feature then translates into another desired trait namely that the network be easy to maintain. By this, user movement while wirelessly accessing the company's network via a portable unit, should appear seamless. Once a network connection has been established, the user should be able to maintain it regardless of where in a given office environment they may wander. This requirement, however, does not need to also include the lack of interruption. Users will understand that changing position may produce a momentary pause in the application they are running. However, the application should not stop and require a complete cold restart simply because a user has moved.

Another area of concern is speed. As portable computers become as powerful and fast as their desk top cousins, users will also expect the level of performance in networking these units to be the same as wired networks. Speed, however, is measured by most workers in how fast they can continue with an application and not in the actual bandwidth of the network 18]. This means that some tasks can be moved to the background allowing the user to return to their work without a noticeable degradation of performance. In addition, hardware advances in combination with more robust signaling techniques have made realizable speeds rivaling some of the popular current wired network products .

Another feature considered desirable in adding wireless as a networking alternative, is that of compatibility. Most users would prefer to add wireless access as an enhancement to their existing systems and not replace these systems with an entirely new network. This level of compatibility will require any wireless network product to be transparent above the network layer. All users level applications down to the transport layer functions should not be affected by the choice of medium used for the network.

Another trait needed in a wireless network is security or privacy. Security must not be compromised simply because a company has decided to move from a wired medium to a wireless one. Unlike their wired cousins where access to the network means that the person needs access to the transmission medium, any compliant 802.11 network entity will be able to access information on an 802.11 compliant network . The need for uniformity among compliant interfaces supplied by different manufactures is one of the reasons for requiring universal accessibility to these wireless network transmissions. But at the same time, some mechanism to achieve at least the same level of security offered by physical wiring must be addressed. This issue is explored further in chapter 3 of this thesis. However, security in a wireless medium is not just as simple as interception. Security concerns must also include denial of service. The network must be robust enough to withstand interference attacks from deliberate as well as incidental sources. Interference suppression and how it affects the design of the physical layer is addressed in chapter 2 of this thesis.

Of course, the actual equipment used to access these networks has properties that require system design consideration. For instance, with notebook computers power conservation and battery life are issues that must not be ignored [Viterbi, A. (1997)]. A notebook computer must be allowed to enter a sleep state without any loss in connectivity to the network. Also, wireless networks will be expected to support a whole new realm of services, including cellular telephone, video, teleconferencing, etc. . This requires the network architecture to take into account not only the different power requirements of the different units used to access the network , but also how to adapt to the varying traffic needs. These issues are discussed further in chapter 3 of this thesis.

1.4 Wireless LAN Concepts

Wireless LAN technology is becoming increasingly popular for a wide variety of applications. After evaluating the technology, most users are convinced of its reliability, satisfied with its performance and are ready to use it for large-scale and complex wireless networks.

Originally designed for indoor office applications, today's Wireless LANs can be used for both indoor peer-to-peer networks as well as for outdoor point-to-point and point-to-multipoint remote bridging applications.

Wireless LANs can be designed to be modular and very flexible. They can also be optimized for different environments. For example, point-to-point outdoor links are less susceptible to interference and can have higher performance if designers increase the "dwell time" and disable the "collision avoidance" and "fragmentation" mechanisms described later in this section.

1.4.1 Topology

Wireless LANs allow workstations to communicate and to access the network using radio propagation as the transmission medium. The wireless LAN can be connected to an existing wired LAN as an extension, or can form the basis of a new network. While adaptable to both indoor and outdoor environments, wireless LANs are especially suited to indoor locations such as office buildings, manufacturing floors, hospitals and universities.

The basic building block of the wireless LAN is the *Cell.* This is the area in which the wireless communication takes place. The coverage area of a cell depends on the strength of the propagated radio signal and the type and construction of walls, partitions and other physical characteristics of the indoor environment. PC-based workstations, notebook and pen-based computers can move freely in the cell.

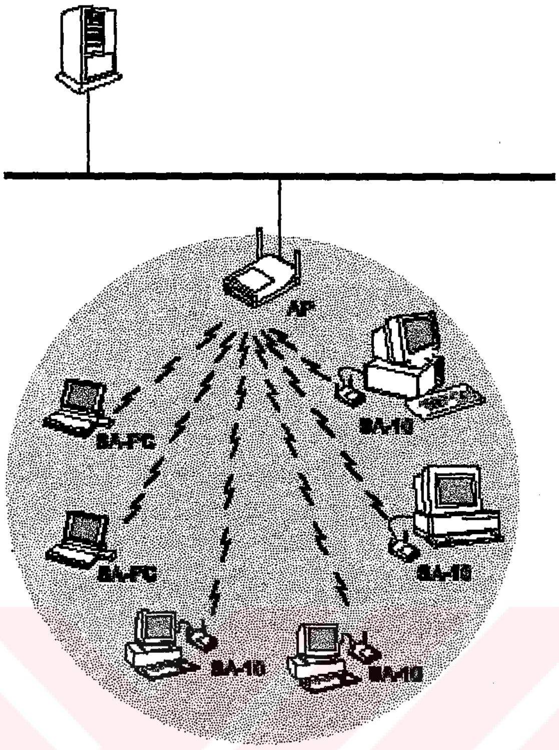


Figure 1.2 The Basic Wireless LAN Cell

Each Wireless LAN cell requires some communications and traffic management. This is coordinated by an Access Point (AP) which communicates with each wireless station in its coverage area.

Stations also communicate with each other via the AP, so communicating stations can be hidden from one another. In this way, the AP functions as a relay, extending the range of the system.

The AP also functions as a bridge between the wireless stations and the wired network and the other wireless cells. Connecting the AP to the backbone or other wireless cells can be done by wire or by a separate wireless link, using wireless bridges. The range of the system can be extended by cascading several wireless links, one after the other.

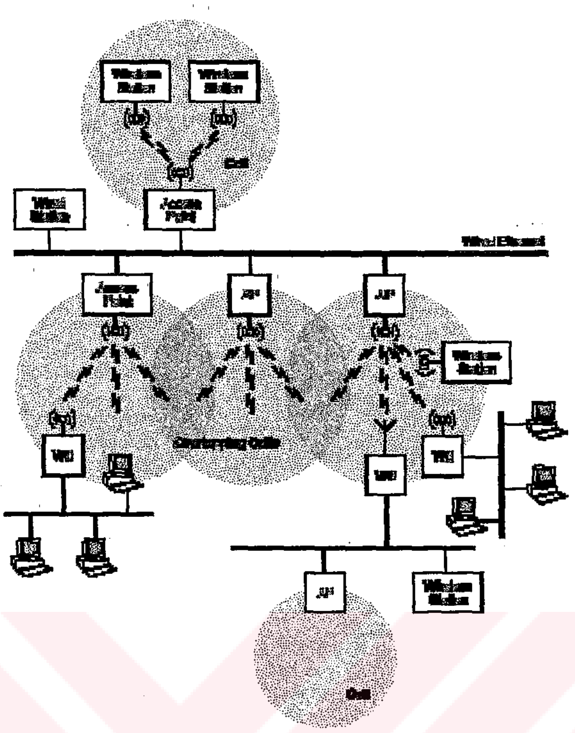


Figure 1.3 Wireless LAN Connectivity

1.4.2 Roaming

When any area in the building is within reception range of more than one Access Point, the cells' coverage is said to overlap. Each wireless station automatically establishes the best possible connection with one of the Access Points. Overlapping coverage areas are an important attribute of the wireless LAN setup, because it enables seamless roaming between overlapping cells.

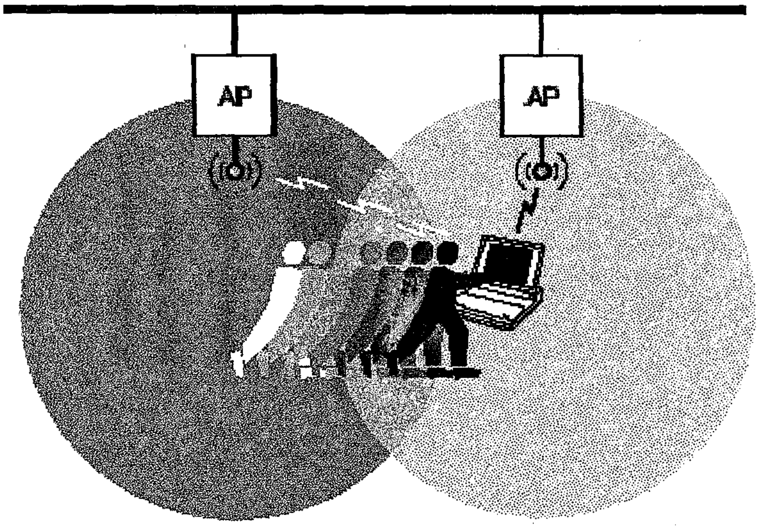


Figure 1.4: Roaming Through Overlapping Cells

Roaming allows mobile users with portable stations to move freely between overlapping cells, constantly maintaining their network connection. Roaming is seamless, a work session can be maintained while moving from one cell to another. Multiple access points can provide wireless coverage for an entire building or campus. When the coverage area of two or more APs overlap, the stations in the overlapping area can establish the best possible connection with one of the APs, continuously searching for the best AP. In order to minimize packet loss during switchover, the "old" and "new" APs communicate to coordinate the process.

1.4.3 Load Balancing

Congested areas with many users and heavy traffic load per unit may require a multi-cell structure. In a multi-cell structure, several co-located APs "illuminate" the same area creating a common coverage area which increases aggregate throughput. Stations inside the common coverage area automatically associate with the AP that is less loaded and provides the best signal quality. The stations are equally divided between the APs in order to equally share the load between all APs. Efficiency is maximized because all

APs are working at the same low level load. Load balancing is also known as load sharing.

1.4.4 Dynamic Rate Switching

The data rate of each station is automatically adjusted according to the received signal quality. Performance (throughput) is maximized by increasing the data rate and decreasing re-transmissions. This is very important for mobile applications where the signal quality fluctuates rapidly, but less important for fixed outdoor installations where signal quality is stable.

1.4.5 Media Access

When many users are located in the same area, performance becomes an issue. To address this issue, Wireless LANs use the Carrier Sense Multiple Access (CSMA) algorithm with a Collision Avoidance (CA) mechanism in which each unit senses the media before it starts to transmit. If the media is free for several microseconds, the unit can transmit for a limited time. If the media is busy, the unit will back off for a random time before it senses again. Since transmitting units compete for air time, the protocol should ensure equal fairness between the stations.

1.4.6 Fragmentation

Fragmentation of packets into shorter fragments add protocol overhead and reduce protocol efficiency when no errors are expected, but reduce the time spent on re-transmissions if errors are likely to occur. No fragmentation or longer fragment length add overhead and reduce efficiency in case of errors and re-transmissions (multi-path).

1.4.7 Collision Avoidance

To avoid collisions with other incoming calls, each station transmits a short RTS (Request To Send) frame before the data frame. The Access Point sends back a CTS (Clear To Send) frame with permission to start the data transmission. This frame includes the time that this station is going to transmit. This frame is received by all the stations in the cell, notifying them that another unit will transmit during the following Xmsec, so they can not transmit even if the media seems to be free (the transmitting unit is out of range).

1.4.8 Channelization

Using Frequency Hopping Spread Spectrum (FHSS), different hopping sequences are assigned to different co-located cells. Hopping sequences are designed so different cells can work simultaneously using different channels. Since hopping sequences and hopping timing of different cells cannot be synchronized (according to FCC regulations), different cells might try to use the same channel occasionally. Then, one cell uses the channel while the other cell backs off and waits for the next hop. In the case of a very noisy environment (multiples and interference), the system must hop quickly. If the link is quiet and clean, it is better to hop slowly, reducing overhead and increasing efficiency.

CHAPTER TWO
PHYSICAL LAYER

The purpose of this chapter is to focus on the concerns of the physical layer design in a wireless network. We start this chapter with a brief history of radio frequency leading into a discussion of the advantages and disadvantages of this technology. Next, we discuss different signaling techniques used in a radio frequency environment focusing mainly on frequency hopping spread spectrum (FHSS) and direct sequence spread spectrum (DSSS). A discussion of the current state of receiver and transmitter technology follows. We then provide a brief discussion of infrared signaling and technology before progressing into the next chapter. Also covered in detail are the FCC rules and regulations in part 15.247 of the FCC code.

2.1 Radio frequency (RF)

Mankind's knowledge and experimentation with radio frequency can trace its roots back to the early 1800's and the work of Michael Faraday and Joseph Henry. Continued investigation of this then new propagating electromagnetic wave phenomena continued through most of the 1800's and included the work of such notable scientists as Hans Christian Oersted, James Clerk Maxwell, Heinrich Hertz, and the father of radio communications Guglielmo Marconi . The first communications system utilizing this new technology was Guglielmo Marconi's 1895 invention of the first practical wireless telegraph system . This new invention would be patented the following year in 1896 by the British government, making it the first ever wireless patent .

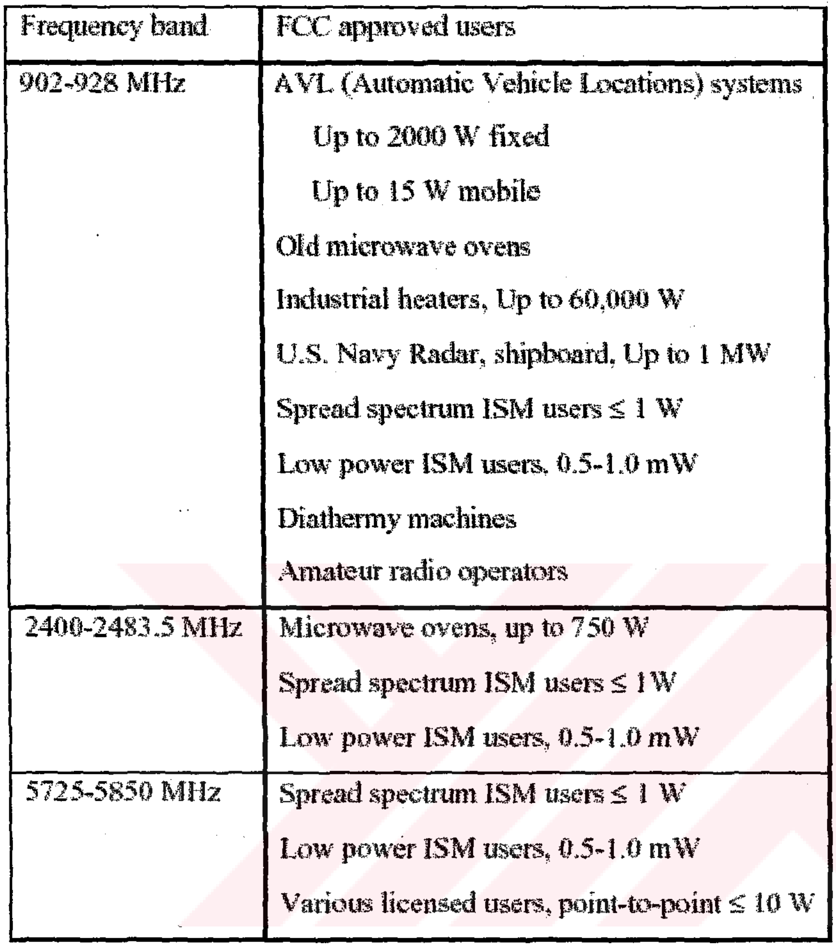
One hundred years latter we are still fascinated with radio frequency and wireless communication. As we enter the next stage in the information age, we turn once again to radio frequency as our method of conveying information from one station to another. Not far from their one hundred year old cousin, the wireless telegraph, today's wireless communication systems are once again turning to coding the information that transverses this wireless channel. Over the past one hundred years since the first wireless communications system was developed, however, we have acquired an extensive knowledge of the characteristics of this transmission medium. Today, instead of having people listening on the ends of our communications system to decipher the sent code, we have connected computers with advanced signal processing devices. We have also evolved new and more efficient signaling methods than those used at the close of the 1800's. One of the most successful methods of signaling in a noisy and interference prone radio frequency environment is spreadspectrum (SS) and its properties and unique characteristics are discussed in section 2.1.3 of this chapter.

2.1.1 Advantages/Disadvantages of using radio frequency for WLANs

One of the main advantages of radio frequency is our long association with it. Someone would be hard pressed to find a person who in their lifetime had not come in contact with a radio communication system. This common use of radio waves by the general public has translated into a perception of safety. This is definitely an advantage when introducing a wireless technology based on radio frequency transmission. However, another side effect of the longevity of radio is the demands placed on the limited available spectrum. This has translated into the need to explore higher frequency ranges and move into microwaves as the preferred frequency for transmission. Unfortunately, microwaves have been much maligned in recent years and some studies have shown a strong correlation between health problems and exposure to microwave frequencies even lower than the acceptable United States limits . This need for the safer low power level when using microwaves, especially in an indoor environment, has introduced fading and other coverage problems. This chapter discusses, however, established methods and techniques for handling this situation.

The limited spectrum has also sparked interest in signaling techniques such as spread spectrum to efficiently utilize any part of a frequency band. It has also sparked interest in the industrial, medical, and scientific (ISM) bands and the Federal Communications Commission (FCC) regulations covering license free spread spectrum communications utilizing these bands. The bands set aside for industrial, medical, and scientific use as identified in [Dixon, R.C. (2000)], are listed in table 2.1 along with other services allowed to operate in this range. Section 15.247 of part 15 of the FCC rules and regulations set forth a list of guidelines that must be adhered to for use of spread spectrum signaling in the ISM bands. A summary of the rules is given in table 2.2 in section 2.1.3 of this chapter.

Table 2.1: FCC ISM band users



Frequency band	FCC approved users
902-928 MHz	AVL (Automatic Vehicle Locations) systems Up to 2000 W fixed Up to 15 W mobile Old microwave ovens Industrial heaters, Up to 60,000 W U.S. Navy Radar, shipboard, Up to 1 MW Spread spectrum ISM users ≤ 1 W Low power ISM users, 0.5-1.0 mW Diathermy machines Amateur radio operators
2400-2483.5 MHz	Microwave ovens, up to 750 W Spread spectrum ISM users ≤ 1W Low power ISM users, 0.5-1.0 mW
5725-5850 MHz	Spread spectrum ISM users ≤ 1 W Low power ISM users, 0.5-1.0 mW Various licensed users, point-to-point ≤ 10 W

One of the problems with utilizing a spread spectrum signaling technique is that it requires both complex transmitters and receivers . This can be detrimental in that the complexity of these devices currently translates into additional cost which could push the product out of a viable cost range.

An advantage of radio frequency, especially with frequencies lower than 5 GHz, is that it is not seriously attenuated by wall, floors, etc.. This makes it an ideal medium for a building-wide wireless LAN. Stationary computers can utilize wireless as a method of communicating between floors or even buildings . Also, because most cellular telephone devices utilize radio frequencies, a data network could be integrated with such a communications network, thus increasing the services offered within the one network.

One of the disadvantages of radio frequency is its susceptibility to interference. Many sources of noise exist within an office environment that can in fact interfere with the network transmissions , making reception of the signal difficult. One of these often overlooked in designing indoor systems is the corona discharge associated with copiers, printers, fax machines and other similar office equipment . This interference is the same type of interference associated with higher power transmission lines. Although the voltage levels used in the interfering office equipment is only a small fraction of the voltage used in a power line transmission system, the distance from the device is also substantially reduced. Therefore, these interfering sources must be taken into account.

Interference within a wireless network can also be caused by what Viterbi labels as the four "multiples," namely: multipath, multiple media, multiple cell-sites, and multiple user access [Viterbi, A. (1997)]. Each of these four sources provide unique problem which must be addressed both at the physical layer and in some case in the MAC sublayer as well.

Multipath interference occurs in an indoor environment on a larger scale than it occurs in an outdoor wireless network environment. Once the network is enclosed, many objects within the enclosure can act as reflecting plates. Metal objects will reflect the signal with little to no attenuation. Since in the 2.4 GHz range is only 12 cm, these bouncing waves can add to the original transmitted signal making the actual signal hard to detect. Multipath effects can be mitigated with the use of spread spectrum signaling techniques and diversity in the receivers . Spread spectrum signaling is a method of signaling that allows the signal energy to be spread across the frequency band. This is accomplished either by changing frequencies or by a series of phase shifts. The amount of time spent on one frequency or a single phase shift is referred to as a chip. The number of chips used to send one bit of information is referred to as a chipping rate. The shifting of frequencies allows for the avoidance of interference since a slight change in frequency will sometimes shift the wave lengths enough to reduce the multipath interference to an acceptable level. The use of phase shifts and multiple chips per bit is an averaging approach to interference suppression. Here, the information is gathered by averaging the chips received to reconstruct the information bit with improved accuracy. Spread spectrum signaling is described in further detail in section 2.1.3 of this chapter.

Diversity has a slightly different meaning than its dictionary definition when used in the context of signal processing. According to [Jensen, M. & Abidi, A. (1999)], diversity is the "...reception of different versions of the same information, with different fading levels." A receiver that employs diversity techniques is a receiver that uses these different versions of the same information to aid in eliminating ambiguity in the information received. The higher the order of the diversity, the further it reduces the probability of errors and the better the reception. In a multipath environment or an environment where the signal can bounce off surfaces arriving at the receiver via several different paths, diversity can actually improve reception by using the different propagation paths to increase the signal to noise ratio (SNR). However, for the techniques used in suppressing multipath interference to work effectively, the interfering multipath signal should be more than one chip out of phase with the current received signal . In the high frequency ranges currently being used, many sources of multipath propagation in an indoor environment have caused the one chip-distance mentioned above to become unachievable. In spite of everything, methods exist for dealing with this problem, including learning periods for the receivers to help suppress the noise. Also, a slight adjustment of the station can sometimes move the receiver enough to avoid a spot with severe multipath interference.

An even more interesting challenge is that presented by multiple user access. Traditionally, multi-user problems have been handled by sensing the carrier before sending. Moreover, the carrier sensing techniques traditionally utilized in the physical layer of wired networks do not translate well into the physical layer of wireless networks. A simple state transition diagram of the carrier sensing and collision detection technique used in most wired 802.3 compliant networks is given in figure2.1.

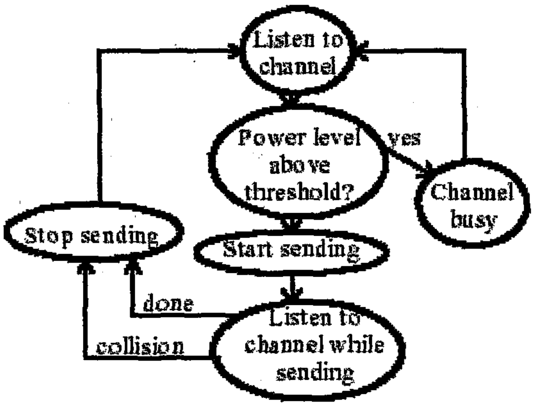


Figure 2.1: Wired network carrier sensing and collision detection state transition diagram

Several problems occur when trying to implement this method of carrier sensing and collision detection in a wireless network. First, the simple power level detection circuitry used in a wired network would fail in an environment where multipath signals exists. These signals could add, producing a power level at the receiver that would be above the threshold. This would serve to aggravate the exposed station problem. Another problem with power level detection occurs when the distance from the stations vary. A station nearby the receiving station will cause a higher received power level than a station far away. If the station determines whether or not the channel is clear to use based solely on power level, then the power level will need to be set low enough to accommodate far stations further aggravating the exposed station scenario.

To attempt collision detection during transmission based on the common wired network methods of looking for code violations and power level determinations would present problems as well. In the hidden station problem, even continuing to listen to the channel after the transmission has commenced will not detect the collision. Also, since multipath interference can cause what appears as code violations, this form of collision detection would become unreliable. A method of overcoming these difficulties by moving this service to the MAC sublayer is described in chapter 3 of this thesis.

The third interference problem identified by Viterbi is multiple media. This problem occurs because of the varying bit rates of different LANs that may be interfaced to the wireless LAN. The multiple media problem can be handled the same way it is handled with other heterogeneous network interfaces. The IEEE 802.11 standard defines a device called a portal. This device is similar to the bridge but because of the mobile nature of the stations attached to a wireless network, it handles only the interface between IEEE 802.11 LANs and other 802.xx LANs . A portal is not used to extend a 802.11 compliant network, therefore it does not posses the same level of functionality as a bridge. A portal does, however, serve as an interface between heterogeneous 802.xx compliant networks. A full description of the architectural layout of the current model for a wireless LAN described in the IEEE standard, is covered in chapter 3 of this document.

Finally, the multiple cell-site problem presents an interesting challenge. In a cellular network, there is a need to balance the requirement for a seamless and continuous communication environment with the suppression of multiple cell-site interference. This problem exists primarily in a soft hand-off scenario where it is possible for a station to be co-located between two cell sites and registered with both. When this occurs, it is possible for the station to receive simultaneous but not identical transmissions from base stations in both cells. This problem is usually dealt with by allowing a station to register with only one cell site at a time. However, especially where mobile units utilizing multimedia services are concerned, this could cause an unacceptable interruption of service. In data networks, the problem of seamless coverage is not as critical and a momentary interruption of service is usually not perceivable by the user. Different methods for handling and overcoming this multiple cell-site problem are discussed in chapter 3, section 3.3.2.1 of this document. Another area often overlooked in considering interference sources of an indoor environment is the source of shadowing losses. These shadowing losses are related primarily to the building construction. They can be attributed to metallic partitions, metal studs in plaster walls, metal lath in plaster walls, metal reinforced concrete floors, and other similar styles of construction used in both new and older office buildings. Also contributing to this problem are metal file cabinet, metal desks, metal waste baskets, and other metal office equipment both stationary and mobile. In this document, we discuss this problem only to the extent that a receiver learns the environment and compensates for it prior to data reception. This method of receiver diversity and learning is effective in overcoming a multitude of interference problems. Even with the aforementioned problems, radio frequency transmission in an indoor networking environment still appears as the most viable method of communicating. The use of innovative signaling techniques both in the transmitters and the receivers have helped to alleviate some of the difficulties without forfeiting many of the benefits. Those problems that can not be overcome in the physical medium can often be mitigated at a higher layer in the protocol stack.

2.1.2 Signaling techniques

The demand for radio frequency and the limitations on spectrum have created a demand for innovative signaling techniques. Fortunately, since Shannon first introduced his concepts of information theory, this field has received much attention and study. Some of the most successful techniques in dealing with the unique interference problems of a noisy medium are spread spectrum signaling methods. Two of these methods will be discussed in detail in the following sections.

2.1.3 Spread spectrum (SS)

Spread spectrum, as the name suggests, is a technique for spreading the signal's energy over the full bandwidth of the channel. According to George Cooper and Clare McGillem, a signaling technique can be considered spread spectrum if it satisfies both of the following conditions . "First, the bandwidth of the transmitted signal must be greater than the message bandwidth (p. 268)." Second, "...that the transmitted bandwidth must be determined by some function that is independent of the message and is known to the receiver (p. 268)." Much of the discussion in this section is taken from the book by Cooper and McGillem.

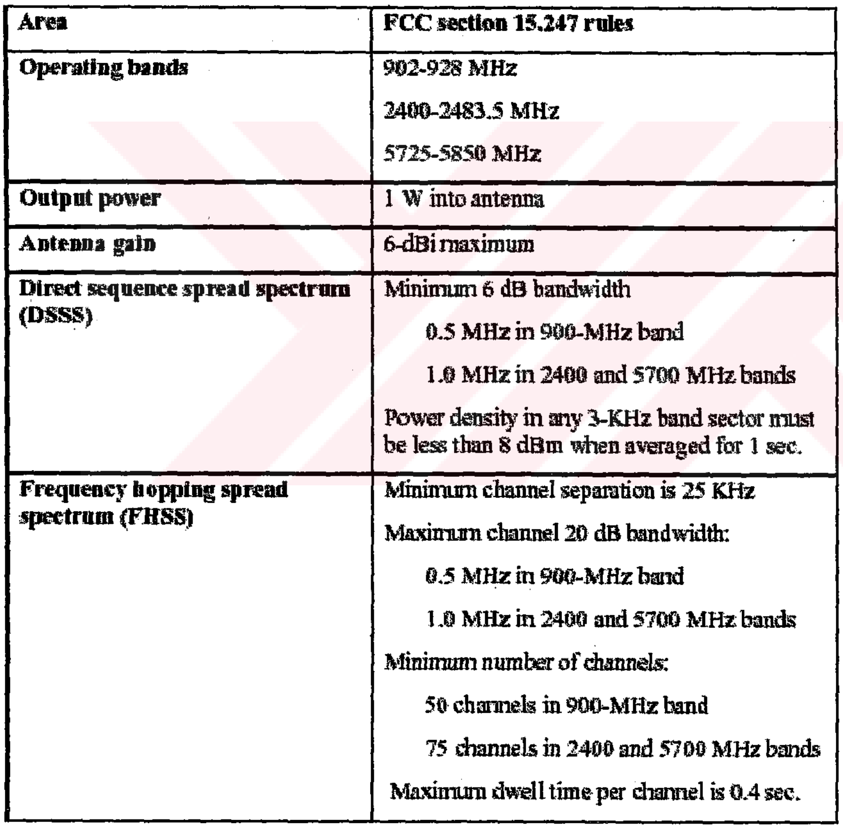
One of the main advantages of spread spectrum communications for wireless indoor networking is its ability to suppress interference. In fact, according to Viterbi, spread spectrum transmission is the only way to make interference harmless . This ability to suppress interference is the reason spread spectrum signaling has become the preferred signaling method on a wireless network. In fact, the types of spread spectrum communications are often distinguished by how they deal with interference. Spread spectrum transmission can really be described as coming in two basic flavors. They are namely avoidance and averaging . In the avoidance class, the most commonly known form is frequency hopping and in the averaging class the most commonly known is direct sequence. Both of these classes have distinct advantages and disadvantages concerning their ability at suppressing certain types of noise. One of the side effects of being able to combat interference is that it can also help reduce the distance between simultaneously broadcasting stations . However, especially with direct sequence signaling, care must be taken not to fall victim to the near-far problem or allowing the transmission from a near station to block the transmission from a far station.

Spread spectrum by its nature provides security to the network without the need for further encryption. This level of security is achieved by using a pseudorandom sequence as the coding sequence. If each station uses a different random sequence, then only a receiver with a prior knowledge of the sequence can receive the signal. Unfortunately, if a device is used in this way to provide security, its export is restricted by the U.S. federal government. This may be the main motivating factor for the uniformity of signaling code in the IEEE 802.11 standard. This standard has the restriction that all IEEE 802.11 compliant systems must be able to receive a physical layer transmission . This has forced security measures into a higher layer protocol and away from physical layer devices.

Other problems associated with spread spectrum relate to the type of spreading used. Direct sequence systems are susceptible to phase distortion on the channel as well as requiring longer acquisition times. Also, the need for a fast code generator can limit the speed of the output. Frequency hopping systems also experience problems. For instance, the speed of the system is primarily determined by the speed of the frequency synthesizer. Also, if the hopping speed is slowed to the bit rate or slower, then error correcting techniques need to be incorporated in higher layers of the protocols.

The FCC has approved the use of spread spectrum signaling in the ISM bands as long as they meet the criteria outlined in section 15.247 of the FCC rules and regulations. Table 2.2, taken from [Dixon, R.C. (2000)], recaps these rules

Table 2.2: Summary of FCC part 15.247 spread spectrum rules for FHSS and DSSS



Area	**FCC section 15.247 rules**
Operating bands	902-928 MHz 2400-2483.5 MHz 5725-5850 MHz
Output power	1 W into antenna
Antenna gain	6-dBi maximum
Direct sequence spread spectrum (DSSS)	Minimum 6 dB bandwidth 0.5 MHz in 900-MHz band 1.0 MHz in 2400 and 5700 MHz bands Power density in any 3-KHz band sector must be less than 8 dBm when averaged for 1 sec.
Frequency hopping spread spectrum (FHSS)	Minimum channel separation is 25 KHz Maximum channel 20 dB bandwidth: 0.5 MHz in 900-MHz band 1.0 MHz in 2400 and 5700 MHz bands Minimum number of channels: 50 channels in 900-MHz band 75 channels in 2400 and 5700 MHz bands Maximum dwell time per channel is 0.4 sec.

The next two sections take an in depth look at two types of spread spectrum namely frequency hopping and direct sequence.

2.1.3.1 Frequency hopping spread spectrum (FHSS)

Frequency hopping in spread spectrum communication refers to the process of signaling whereby the signal is spread over a wide frequency band. Primarily, the bandwidth of any one chip or hop interval is much smaller than the full frequency band. Several authors define two types of frequency hopping systems, namely slow frequency hopping and fast frequency hopping. The distinction made between these two methods is in how rapidly the frequency is changed. Slow frequency hopping is defined as when the rate of changing frequencies is less than the bit rate and conversely fast hopping refers to a system where the hopping rate is faster than the bit rate. As would be expected, slow frequency hopping is susceptible to near-far problems and other non-Gausian interference problems. Also, slow hopping can become susceptible to interference on one of the channels. As a result, slow hopping requires a more robust error correction technique to recover bits lost during a hopping interval . The key to reducing interference at the receiver in a frequency hopping system come from the assumption that the system transmits on a channel with high interference only a fraction of the time . Therefore, when utilizing frequency hopping, spread spectrum as an interference avoidance device in a noisy environment, fast hopping is preferable. A view of the transmitter block diagram for a frequency hopping spread spectrum system is shown in figure 2.2.

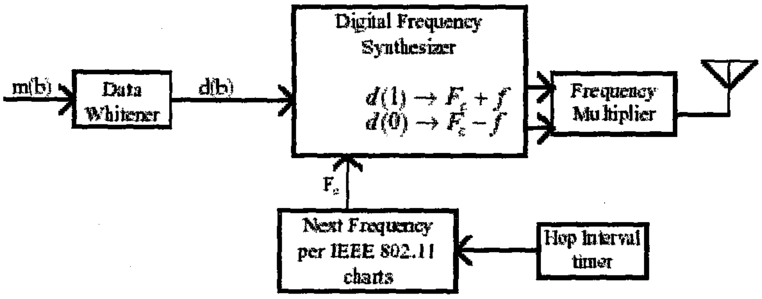

m(b)
Data Whitener
d(b)
Digital Frequency Synthesizer
$d(1) \rightarrow F_c + f$
$d(0) \rightarrow F_c - f$
Frequency Multiplier
F_c
Next Frequency per IEEE 802.11 charts
Hop Interval timer

Figure 2.2: FHSS transmitter block diagram

The actual signaling or sending of a bit at the carrier frequency, is handled in the

IEEE specifications as a frequency offset from the center frequency of the channel. As shown in figure 2.1, this center frequency is referred to as F_c, and the offset is represented by a lower case f. The specifications state that, for the United States, the center frequencies of the 79 channels in the 2.4 GHz band are 1 MHz sequential steps ranging from 2.402 GHz to 2.480 GHz. These bandwidths are measured at the -20 dB points and F_c sits midway at □500 kHz . The nominal peak deviation or f is specified at 160 kHz. For example, the symbol set { 1, 0, tristate} transmitted at 2.402 GHz would be transmitted as:

$$symbol(1) = \left(2.402\times10^{9} + 160\times10^{3}\right)Hz = 2402.16MHz \tag{2.1}$$

$$symbol(0) = \left(2.402\times10^{9} - 160\times10^{3}\right)Hz = 240184\,MHz \tag{2.2}$$

$$symbol(tristate) = 2.402GHz \tag{2.3}$$

The power levels of the transmission are defined in the current specifications with only a minimum and a maximum point. The minimum is defined as 1.0 mW and the maximum at 100 mW. The other restriction placed on the transmission system by the specifications is that the "...signal must maintain an E_b/N_o of 16.0 dB in the presence of Gaussian white noise at a BER of greater than or equal to 10_{-5} ." Most of the receiver specifications are still open, including the specification of the receiver type.

To handle the difference between the actual transceiver functions and the interpretation of the received signal, the IEEE 802.11 standard divides the physical layer into two sublayers as shown in figure 2.3. The physical layer frame is a function of the PLCP sublayer whereas the actual transmission specification listed above would be functions of the physical medium dependent (PMD) sublayer.

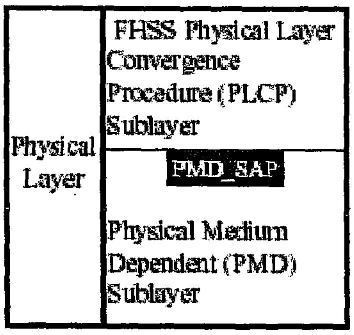

Physical Layer
FHSS Physical Layer Convergence Procedure (PLCP) Sublayer
PMD_SAP
Physical Medium Dependent (PMD) Sublayer

Figure 2.3: Physical layer and sublayers

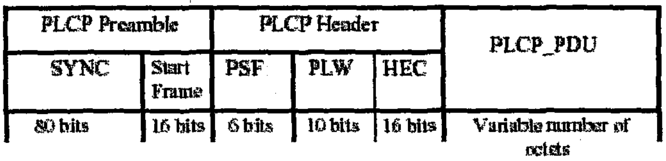
Detecting and tuning into a signal are handled in the IEEE 802.11 standard specification by the preamble section of the physical layer frame. The format of this frame is given in figure 2.4. This preamble section of the frame breaks into two parts: namely an 80 bit synchronization section (SYNC) followed by a 16 bit unique word. This 16 bit unique word is the same in all three methods of signaling currently specified in the standard. The 80 bit SYNC field is sent as an alternating pattern of ones and zeros at a rate of 1 Mbps. This pattern actually serves several functions. Besides helping the station detect the presence of a signal on a specific channel, it also helps with frequency synchronization between the sending and receiving stations. Once the signal is detected and the station has synchronized, the station need only watch for the start of a frame delimiter which is a unique word given by the primitive polynomial in equation 2.4. Care is taken so that this pattern is not repeated anywhere within the header of the frame.

$$x^{11} + x^{9} + x^{8} + x^{7} + x^{4} + x^{3} + x^{2} + 1 \qquad (2.4)$$

The PLCP header consist of a 6 bit PLCP signaling field (PSF) which has not yet been fully defined, the 10 bit PLCP_PDU length word (PLW) which specifies the number of octet contained in the PLCP_PDU, and finally the header error check (HEC). The IEEE 802.11 standards draft utilizes the CCITT 16 bit CRC for implementation of the HEC. This CRC is given by the generator function:

$$G(x) = x^{16} + x^{12} + x^{5} + 1 \qquad (2.5)$$

The check sum is performed as the standard ones complement of the remainder modulo two division of the PSF and PLW fields by the generator polynomial. The advantage of this method of calculating a check sum is its easy hardware implementation via invertors, a shift register, and a row of exclusive-or gates.



PLCP Preamble		PLCP Header			PLCP_PDU
SYNC	Start Frame	PSF	PLW	HEC	
80 bits	16 bits	6 bits	10 bits	16 bits	Variable number of octets

Figure 2.4: FHSS PLCP frame format

As can be seen from the above discussion the method of transmission utilizing FHSS is a good method for overcoming the noise problems associated with indoor wireless networks. One of the nice advantages of frequency hopping is that the hopping pattern can be adjusted to avoid frequencies where interference is too high. This technique of interference avoidance even adds an additional dimension to a cellular network's frequency reuse policy. If correctly arranged, the cells can share a frequency band as long as the hopping patterns used within adjacent cells do not overlap [Jensen, M. & Abidi, A. (1999)]. Also, if noninterfering hopping patterns are used, two wireless networks could independently coexist in the same room without affecting each others signal quality.

2.1.3.2 Direct sequence spread spectrum (DSSS)

Direct sequence spread spectrum is sometimes referred to as a pseudo-noise or PN system. The reason for this second title deals with how direct sequence spread spectrum is implemented. Unlike frequency hopping, direct sequence operates on one carrier frequency and utilizes an averaging method of interference suppression. The method utilizes a pseudo-random sequence consisting of what are referred to as chips. A symbol or information bit is sent using a fixed number of chips. This is referred to as a chipping rate. The bits are first whitened using a pseudo-noise generator and then sent using a chipping code . The carrier is most commonly modulated by utilizing biphase or quadphase modulation which is performed differentially to remove any ambiguity concerning starting phase. The current technology limitation on this chipping rate is 1.55 Gcps .

Most of the time direct sequence spread spectrum signaling is used on a multiple access channel with the understanding that any interfering signal can be modeled as Gausian White Noise and as such dealt with in the same fashion . In a direct sequence spread spectrum system, the probability of an error is then based on the phase and amplitude of the interfering signal . Although these systems do become interference limited, their performance is still better than time division or frequency division systems .

The physical layer in the IEEE 802.11 standard is broken into two sublayers utilizing the same division used in the FHSS specification of the physical layer and given in figure 2.3. There are however, some differences in the frame formats used in the Physical Layer Convergence Procedure (PLCP) sublayer for a DSSS system versus a FHSS system. The frame format is given in figure 2.5 for a DS PLCP sublayer frame. The differences lie in some of the fields of the header. The SYNC field is now 128 bits long and is sent as the chipped version of 128 one bits. It is not sent through the data scrambler. The unique word remains unchanged. The signal field is used to indicate the type of modulation used. The current standard supports two choices in this field namely differential binary phase shift keying (DBPSK) giving a data rate of 1 Mbps and differential quadrature phase shift keying (DQPSK) giving a data rate of 2 Mbps. The length field is the same as the FHSS header field and is used to indicate the data length. The CRC is the same as the FHSS HEC and is used to validate that the header was received correctly.

SYNC	Unique Word	Signal	Service	Length	CRC	MPDU
128 bits	16 bits	8 bits	8 bits	16 bits	16 bits	variable num. of octets

Figure 2.5: DSSS PLCP frame format

The most interesting part of any DSSS system is not the packet format, although important for control, but rather the way the information is sent across the channel. The data scrambler and descrambler used in the IEEE 802.11 is rather simple. It uses the polynomial:

$$1+x^{-4}+x^{-7} \qquad (2.6)$$

The block diagram of the scrambler can be found in figure 2.6 and the descrambler is shown in figure 2.7. In these diagrams, the blocks labeled x-1 denotes a time delay of one unit. These delayed signals are fed back and added using a MOD 2 or an exclusive-or gate to produce the scrambled output. The process is reversed for the descrambler. This feed-forward configuration makes the scrambler and descrambler self initializing.

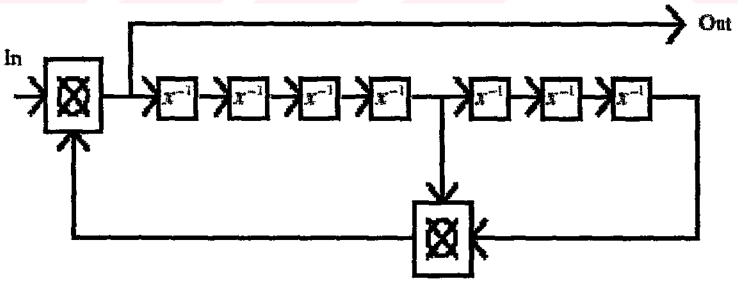


Figure 2.6: DSSS scrambler

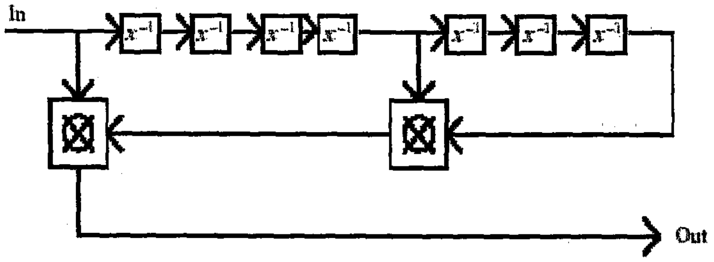


Figure 2.7: DSSS descrambler

Once the scrambled code leaves the scrambler it enters the spreader. The IEEE 802.11 standard suggests an 11 chip Barker spreading sequence given by:

+1, -1, +1, +1, -1, +1, +1, +1, -1, -1, -1

The spread code is then modulated using the truth tables given in table 2.3 and table 2.4. A better example of modulation truth tables are those given by Cooper and McGillem in their book . These modulation values are given in tables 2.5 and 2.6.

Table 2.3: DBPSK encoding table

Bit Input	Phase Change (+jω)
0	0
1	π

Table 2.4: DQPSK encoding table

Bit Pattern (d0, d1)	Phase Change (+jω)
00	0
01	π/2
11	π
10	3π/2

Table 2.5: Better DBPSK encoding table

	Message Bits	
Barker Code	1	0
1	0	π
-1	π	0

Table 2.6: Better DQPSK encoding table

	Message Bits	
Barker Code	1	0
1 1	π/4	5π/4
1 -1	7π/4	3π/4
-1 1	3π/4	7π/4
-1 -1	5π/4	π/4

An example of how data bits would be sent using the modulation scheme given in table 2.5 with the Barker code is given in figure 2.8.

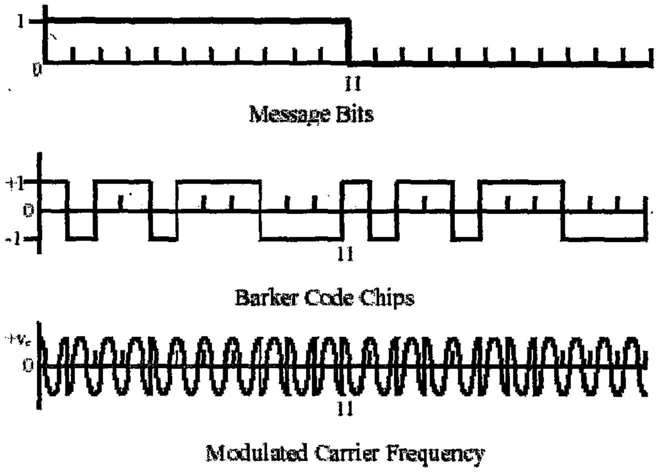


Figure 2.8: Sample of DSSS modulation

The receiver for this type of signal would need to detect the phase shifts in order to detect the signal. To avoid any ambiguity, the IEEE 802.11 standard uses the 128 bit SYNC section of the frame to not only help detect a signal but also to align the receiver to the transmitter. Once the receiver is aligned, it can start scanning for the unique word which marks the start of the header and subsequent packet. Note that in contrast to the FHSS SYNC section of the header, the SYNC section in the DSSS header is sent as 128 ones. The reason is obvious when observing figure 2.8. The detection of a signal is obtained by noticing the phase shifts corresponding to a Barker Coded one sent continuously. Detection and frequency alignment can both be achieved with this modulated signal. There is no need to alternate between one and zero, as is done in the FHSS header. In fact, such an alternating bit sequence would only hinder detection and acquisition not assist it.

Now that the two most commonly used forms of spread spectrum signaling have been discussed, it is time to turn our attention to the actual transmitters and receivers used.

2.1.3.3 Transmitters and receivers

Although both transmitters and receivers have been discussed in relation to an implementation of both FHSS and DSSS, there are still some unique design concerns that should be taken into account. The limitations on signaling comes primarily from physical limitations in the transmitters and receivers. As mentioned previously, in frequency hopping spread spectrum the limit is not the code generator but rather the frequency synthesizer . In direct sequence spread spectrum signaling, the limitation on chipping rates is currently limited to 1.55 Gcps .

Another hardware problem comes from the frequency offsets of the actual crystal oscillators inside the individual computers . This offset can cause a drift between the transmitter and the receiver that can make the signal hard to recover . In a real system, there are two causes of frequency offset. These causes are unintentional offsets caused by unstable oscillators and intentional offsets imposed by the design . The intentional offsets are used by the receivers to reconstruct the signal with a better overall accuracy. These figures show that with the right choice of oscillator, the effects of unintentional offsets can be minimized. The SYNC section of the IEEE 802.11 PLCP sublayer frames discussed above do allow for initial synchronization. Also, in the IEEE 802.11 implementation, both the data rates and the chipping rates are sufficiently below the carrier frequency as to not cause enough drift during the duration of a packet to lose synchronization. Since care is taken to synchronize and tune the frequency at the start of reception, the frequency drift between the transmitter and receiver can usually be ignored.

Another problem affecting transmitter and receiver design is the suppression of interference. Due to the fact that the interference level at the transmitter differs from that at the receiver, a power adjustment made at the transmitter to overcome an interfering signal may not be sufficient at the receiver . If the transmitter and the receiver are close, there is a strong correlation between the two interference levels but this correlation weakens with increasing distance . However, with the right kind of diversity implemented in the receiver, most of the interference can be mitigated.

In cellular or base station type architectures, there is an additional problem with the use of orthogonal codes. Orthogonal codes are codes that are uncorrelated meaning they do not interfere with each other. Orthogonal codes are used primarily in architectures such as code division multiple access (CDMA) designs. This type of system is discussed briefly in section 3.3.1 of this document. The problem with orthogonal codes comes from maintaining the orthogonal characteristics of the codes when the transmissions are not synchronized. In the uplink transmission of a cellular network, the users will not be synchronized and the codes could overlap at any point in the transmission. Therefore the use of orthogonal codes to send multiple unsynchronized signals simultaneously on an uplink channel, may result in a loss of orthogonality between the signals . However, on the downlink channel, synchronization can be maintained and several transmissions can be overlapped increasing the channel's bandwidth.

Viterbi suggests that the problem with the overlapping codes can be overcome at the receiver if the strongest signal is decoded, re-modulated, and subtracted, from the composite signal . He has proposed that the number of codes that maintain orthogonality under these conditions is quite large. However, other researchers have discovered that when the codes must remain orthogonal even in the un-synchronized environment discussed above, the number of codes that meet these conditions reduces to a small subset of the original group . Since it is assumed that the majority of the traffic in a cellular network will be to access the resources of the larger backbone network, the expectation is that uplink traffic will be significantly lower than downlink traffic . This means that collisions occurring in uplink traffic will not have as serious an impact on the performance of the system as downlink collisions. The impact of the uplink transmission collisions is even further reduced because most cellular systems employ a two channels structure using one for uplink traffic and one for downlink traffic. This two channel structure eliminates the up-down collision. Because the codes used in the downlink traffic can be perfectly synchronized maintaining the orthogonality of the codes, it is still possible to assign each station in the cell a unique code and overlap several transmissions without interference . Other techniques such as polling and contention periods can be employed on uplink traffic to improve its performance.

The actual hardware implementation of a transmitter requires only a transceiver core chip, an antenna switch chip and a low frequency synthesizer chip. If using microwave communication, then a resonant oscillator provides better performance when minimizing near-carrier noise. The rest of the processing is done digitally and can be implemented with an on-board processor.

When designing or choosing a receiver it is important to remember that its main purpose is to extract the data bits from the signaling information received . In order to accomplish this task it must be able to evaluate the multipath environment thereby determining not only the different possible propagation paths of the signal but also the relative phase and time delay of each of those paths . The functioning of a receiver can then be further defined to include not only acquisition of the signal but also the tracking of that signal .

With a spread spectrum system, the transmitted frame's preamble section helps the receiver detect the presence of a signal versus channel noise. Once a signal is detected on the channel, the receiver needs to synchronize with the sender and continue to align its synchronization during the full length of the transmission . This issue is addressed in the IEEE 802.11 standard and handled with the SYNC field. Also, the choice of data bit rates versus carrier frequency in the IEEE 802.11 standard allow for a certain amount of drift before the synchronization needs adjustment . However, for faster systems, this problem of maintaining synchronization can become an issue. The whole alignment and synchronization process is further hampered when it is noted that timing offsets of a half a chip can result in power losses of several decibels [Zeigler, R. & Cioffi, J. (1999)]. In a low power environment where the E_b/N_o ratio is already small, this reduction could be the difference between whether or not the signal is received. Remembering that not all stations in a wireless network are expected to be able to hear each other, it is possible for this power level problem to make the station unreachable.

Spread spectrum signaling is popular due to its mitigation of multipath effects. In fact multipath and multiple-access interference are attenuated by the processing gain when they are out of phase by more than one chip with the signal on which the receiver is tuned . The problem with this is that in a high frequency indoor environment, the multipath interference will not maintain the one chip limit and interference can occur. For example, in the IEEE draft implementation of DSSS signaling, the eleven chip Barker code is transmitted at 1 Mbps. This means each chip has a propagation length (assuming the speed at $3x10_8$ m/sec and rounding the result) of 27 meters. The returned signal from the numerous multipath sources in an office environment will not meet this one chip limit.

To mitigate the ISI problem caused by multipath, a directive antenna could be implemented . Of course this would require line of sight between transmitter and receiver, and the performance would degrade if the receiver were moving . However, movement during use in a portable computing environment can be considered the exception and not the norm; therefore, the implementation of a directive antenna is a viable solution in this type of environment. Also, most users are familiar with the properties of radio transmission systems and will not see some small adjustments in position of the device to improve reception as a handicap. However, if the position at which the unit can receive the signal correctly is too limited, most users will consider the need for exact positioning of the device a disadvantage.

Fortunately, several signal processing techniques exist in receiver filter implementations to help mitigate the effects of multipath. These techniques can help alleviate the need for line-of-sight transmission. One of the most common is the sequence detection or Viterbi detection filter [Zeigler, R. & Cioffi, J. (1999)]. It has been shown that this type of receiver is better than a traditional equalization receiver when the channel response is known and is time invariant . The problem with this is that in an indoor wireless environment, the components of the multipath signal can be changing dynamically making it difficult to know with any degree of certainty the channel response at any particular instant. Fortunately, this receiver can be made adaptive, thus handling a dynamically changing environment [Zeigler, R. & Cioffi, J. (1999)]. Viterbi even points out that, with wideband spread spectrum signaling, the multipath signals can be isolated and combined to improve the performance of the receiver [Viterbi, A. (1997)]. This is accomplished by allowing the receiver to learn the characteristics of the channel prior to reception of the data. The learning phase of such a receiver can be handled using the IEEE 802.11 draft specifications during the SYNC section of the frame and adjustments made only if needed . If the frame lengths are short enough, there will be a smaller probability of changes in the channel characteristics during reception of the frame.

Another popular receiver technique is the RAKE receiver. Here multipath signals with varying delays are used to improve reception . The receiver uses several taps to take readings of the channel along a one chip interval. During the initialacquisition period, the receiver tests for all possible delay positions and looks for the maximum recovered energy . The learned information is then uses to align the taps and recover additional signal strength from the delayed multipath components of the signal. In other implementations, the initial path parameters can be derived from a downlink pilot signal , and the receiver parameters can be adjusted as needed .

As can be seen from the above discussion, although problems exist in an indoor wireless network environment that affect the design of the transmitters and receivers, they are manageable.

2.2 Infrared

Infrared is the other form of wireless signaling that is often used in indoor wireless networks. The general public has become accustomed to having and utilizing infrared signaling devices for such purposes as changing the channels on their television or controlling their stereo or VCR. A pleasant result of this widespread utilization of infrared signaling is that this technology is viewed as safe by the general public . Also, most users are aware of the line-of-sight requirements of infrared transmission from using their remote control devices. Infrared signaling technology is discussed briefly below.

2.2.1 Advantage/Disadvantage of infrared networks

The advantages of infrared technology are numerous. Often mentioned is the way infrared is almost completely attenuated by walls, floors, and so on. This makes it ideal for implementing a network where security is maintained by confinement to one room . It is virtually impossible for anyone outside the room to intercept network transmissions. Also, because infrared technology utilizes noncoherent methods to detect signals, the transmitters and receivers are simpler and less expensive than their radio frequency counterparts [Bantz, D. & Bauchot, F. (1998)]. Another nice advantage of infrared transmission is that with the exception of limitations imposed on the permissible optical power densities, it is license free.

Some of the disadvantages of Infrared technology stem from the characteristics of light transmission. For example, the region of the spectrum used in infrared signaling is also shared by the sun making an office window a potential interference source. Also, because Infrared transmissions are seriously attenuated by walls, floors, and other surfaces, it is not a suitable choice for wide area transmission systems that would cover more than one room . However, if a network is to be implemented across room boundaries, a cellular structure-based system could be used. Simply by connecting the base station via another medium, infrared transmission could be effectively used in each of the separate rooms.

Another effect of the attenuation problems associated with infrared signaling is the need for line-of-sight transmission . This makes infrared signaling extremely susceptible to movement and can limit its applicability. Also because infrared is a calorimetric method, there is a strong possibility that collisions will go undetected since the optical power levels can vary from station to station .

Infrared transmission systems can be used as a complimentary system to enhance the coverage and performance of a radio frequency WLAN. The systems can coexist in the same room and not interfere with each other's transmissions. The signaling techniques used with infrared are base on power level detection and not phase or frequency detection like their radio frequency cousins. Some of these methods are covered next.

2.2.2 Signaling techniques

There are three commonly used signaling techniques with infrared signaling. The first and the one most people are familiar with is the aim and focus transmission technique . This is the method utilized in TV, VCR, and stereo remote controls. One advantage to this method is that the range is dependent on the power and degree of focusing therefore it can range in kilometers and be utilized for transmission between buildings . Of course, as most people have discovered from changing channels on their TV, this method is extremely susceptible to movement of the transmitter or receiver while transmitting. The next method of transmission is to bounce the signal off a ceiling reflector . These devices can be either active or passive depending on the throughput needs of the system . The final method is to radiate the signal omni-directionally [Bantz, D. & Bauchot, F. (1998)]. The obvious advantage of omni-directional transmission is that no line-of-sight tuning is required .

The system used most often is to locate a form of a reflecting device on the ceiling of the room and bounce signals off of it. As already mentioned, these devices come in two varieties, active and passive. Active ceiling devices can be used to implement some form of data control, channel access control, and collision detection. They fulfill a good deal of the same functionality supplied by a base station in a cellular radio frequency wireless LAN. One such device is described in and referred to as a satellite. This satellite amplifies the optical signal and fills the entire room, making it easier for the receiver to detect the presence of a signal. Also, because satellites break the path between any two stations into an uplink and a downlink, collisions can be controlled. This control is usually implemented by the satellite signaling the stations that a collision has occurred. This means that if a collision is detected by the satellite it can replace the downlink signal with a collision presence signal thus providing 100% collision detection on the downlink .

Like base stations in cellular networks, the satellites could be connected to each other via a high speed backbone LAN (i.e. COAX or fiber). Such a system is proposed in . When a satellite receives a packet at the COAX or fiber interface and the recipient is known to be in its cell, it repeats the packet only at the wireless interface. If a packet is received via the wireless interface and the station is not within the cell, the packet is repeated at the wired interface. If the packet is received at the wireless interface and the station is in the cell, the packet is repeated at the wireless interface only.

The collision handling of the satellite varies depending on which interface is the cause of the collision. If a collision is detected on the wired interface, the satellite replaces the downlink signal over the wireless interface with a collision present signal but takes no action at the wired interface. If the cell is idle when the collision is detected at the wired interface, then no action is taken at the wireless or wired interfaces. Collisions on the wireless interface are handled only on the wireless interface.

CHAPTER THREE

MEDIUM ACCESS CONTROL

This chapter is a discussion of the medium access control (MAC) sublayer issues involved in wireless networking. This discussion is followed with a brief discussion of code division multiple access (CDMA) already discussed in chapter 2 but continued here in how it relates to such issues as MAC layer addressing. Since time division multiple access (TDMA) protocols are popular, a brief discussion of these follows but are not covered in depth. Finally, the chapter covers in some depth the asynchronous transfer mode protocol.

3.1 Description

"The principle conclusion we can draw from these studies is that we can draw no conclusions from them. One can always find a set of parameters that make one of the LANs look better than the others."

Andrew Tanenbaum

Although Andrew Tanenbaum, in the above quote from his book on computer networks, is referring to the wired IEEE 802 family of MAC layer standards, it is an even more appropriate observation when considering the numerous architectural proposals for wireless networking. The design of a suitable MAC sublayer requires the need to handle access to the chosen medium. In wireless networking, simply by the characteristics of the transmission method, the medium is multiple access. This means there will be the inevitable contention for the limited resources. A MAC protocol must play the role of traffic cop, carefully and fairly controlling who has access to the channel and how long they are able to hold the channel.

Existing MAC sublayer standards, which are part of the IEEE 802.xx family, make certain assumptions concerning the detectability of channel use, the security of the channel, and the connectivity of the stations attached to the network. These assumption no longer hold in a wireless channel environment. For instance, in a wired network, stations are physically wired together and the wire needs to be severed or the station turned off in order to lose connectivity. Also, in a wired network, the stations rarely change their point of attachment, meaning if a station is attached at point A today, it is safe to assume it will be attached at point A tomorrow. On the other hand, in a wireless network not only are the stations not physically attached, but movement during use can cause the station to enter a dead spot, or a spot in the room where it can no longer be reached, breaking its connection to the network. Also, just because a station is attached at point A today, does not imply it will remain attached at this point tomorrow. Therefore, a new set of protocols is needed to handle access to this new network medium. At the same time, however, the new IEEE wireless network MAC and physical layer (802.11) protocols must be compatible with the other members of the 802.xx family. This requires the consideration of such details as addressing, bit error rates (BERs), varying bit rates, and collision handling, especially when one interface experiences a collision and the other is free. These are some of the key reasons why the existing 802.xx family of protocols falls short of producing the desired results in a wireless environment .

The design of a suitable MAC protocol is further complicated by the desire to share the network with other services. This need to incorporate isochronous services such as voice and video in with data transmission, and ensure the desired quality of service to each user, is problematic at best.

A closer look at the IEEE current standards for wireless networks will give insight into how some of these problems, unique to a wireless environment, can be handled.

3.2 IEEE 802.11

The current IEEE 802.11 MAC standards for wireless LANs outlines some interesting aspects of the MAC sublayer interface. It deals primarily with the architectural services that are provided by this layer. The standard also deals with such issues as virtual carrier detection, power down or sleep-state operation, interfaces to existing wired LANs utilizing one of the 802.xx family of protocols, and different types of mobility when accessing a wireless computer network.

The IEEE 802.11 standart defines two different types of stations. This distinction is made primarily by the amount of movement associated with each type of attachment. The first type of station is referred to as a portable station. This defines a station which does not change its point of attachment or location during the process of sending or receiving information, but only changes this location in between transmissions. It is also possible for a portable station to be actually stationary and never change its point of attachment. The second type of station defined is called a mobile station. As the name implies, this station moves between locations as well as points of attachment while transmitting or receiving. The standard has to be able to handle both types of attachment and handle both if they coexist in the same network.

3.2.1 Basic services

The IEEE 802.11 standard describes a group of basic services that all compliant systems must provide. The standart currently defines seven of these services, which break into two main categories. These categories are: station services and distributed system services (DSS). Flexibility has been worked into the current standard by leaving how the services are provided less stringent. For instance, when implementing a distributed system, the standard does not require a layer-specific implementation. Instead, the IEEE 802.11 standard allows for centralized or distributed systems and link or network layer implementation.

There are four services the IEEE 802.11 standard requires a station (STA) to handle.

The first, authentication, means that the station must be able to prove it is who it says it is. This also means, a station must be able to identify itself as well as verify the identity of other stations connected to the network. This need for authentication leads into the next service a station must provide: privacy. Because the IEEE 802.11 standards standart requires a uniform interface that all compliant networks will be able to use to access the network, the MAC layer must provide the needed privacy and defense against eavesdroppers. The purpose of this service is primarily to offer the same level of assumed security that is inherent in wired networks. For a wired network that is isolated from other networks, the act of merely restricting access to the wired resource adds a considerable level of privacy. In most wired systems, an eavesdropper can be detected. This is not the case with wireless networks. With a wireless interface, no method exists for detecting when someone else is listening. The privacy service, therefore, allows for the use of encryption services to offer the same level of privacy offered by wired 802.xx alternatives. Encryption allows access to the network to be restricted to only those members who have the proper deciphering key. The actual method of encryption used is not restricted by the standard. The encryption service can then be made network specific in that each network can decide what level of encryption is needed to provide the desired level of privacy. The remaining two services offered by the station deal with the actual establishment of the network. These two services are association and disassociation. As expected, the service of association defines how a station becomes attached to the network and disassociation defines how it leaves.

To understand the second set of services covered by this document it is necessary to understand some of the basic network infrastructure outlined in the 802.11 standart for the wireless networks. Since the wireless medium is a multiple access medium, there is a need for some form of coordination among stations. The IEEE 802.11 standard provides mechanisms for coordinating stations and calls a group of coordinated stations a basic service set (BSS). Several BSSs can co-exist in the same physical location allowing the handling of heavy traffic areas. The BSSs can be further combined via an access point (AP) into a distributed system (DS) of these coordinated services groups, each of which is independent. This grouping is called an extended service set (ESS) and the seamless integration of the BSSs in the ESS is handled by distributed system services (DSS).

The second set of services defined in the 802.11 standard are DSS. There are three currently defined DSS. The first of these is distribution. As the name implies, the distribution service deals with the routing of packets between BSSs in one ESS. In other words, this is the service that allows the DS to determine which BSS a station belongs to and route the MAC data units (MDU) to the appropriate BSS. The next service provided is integration. This service allows different DSs to communicate as well as controls communication between 802.11 and the remaining family of 802.xx compliant networks. Finally, the last service that is provided is re-association. Although at first this appears as if it should be a station service and grouped with association and dis-association, it is however a DSS. This can be seen when considering the movement of a mobile station that must have a re-association pending prior to disassociating with its current BSS. This joint association is needed to maintain seamless connectivity during movement between BSSs. As such, the management of this then joint association (the current association and the reassociation being established) is best handled at the DS level of the network since it will require the proper routing of MDUs.

3.2.2 Virtual carrier sensing

The IEEE 802.11 MAC and physical layer specification standard is based on a carrier sensing multiple access with collision avoidance (CSMA/CA) protocol with random back-off time for when the channel is sensed busy. Although carrier sensing is primarily a physical layer mechanism where power levels and code violations are used for a determination, the IEEE 802.11 standards standard uses instead what is termed as virtual carrier sensing in the MAC sublayer specifications. The motivation behind this change in layer for the implementation of carrier sensing comes mainly from the problems, described in chapter 2, with the physical layer implementation of this service. This virtual carrier sensing is performed utilizing information contained in short control packet exchanges between communicating parties prior to actual data frame transmission. This exchange of control packets is the same exchange used in the multiple access with collision avoidance (MACA) protocol outlined in section 3.3.1.1 of this thesis and is described in detail in that section.

3.2.3 Power management

One of the more interesting issues addressed in the MAC section of this standard is the issue of power management. Since power management is an important concern in the implementation of a wireless network, how this standard specification deals with this issue is of interest. Power management considers the question of how to monitor the network and the transmission of data while a computer is in a sleep mode. To understand how the standard deals with the question of power management it isnecessary to understand some of the basic infrastructure of a 802.11 compliant wireless network.

In its most basic form, the IEEE 802.11 standard requires four network components. These are the stations (STA) which are the actual computers attached to the network, BSS which are the set of coordinated stations, DS which is a set of BSSs, and an access point (AP) which is any station that provides access to a DS. The basic service units that must be provided by the STAs and DS have been already discussed. The main unit of functionality, however is the AP. This is a station that interconnects distribution systems and is the only way a member of one BSS can communicate with a member of another BSS. It is basically a station that functions as base station controlling the BSS synchronization via beaconing and similar services. Therefore, in order to understand the workings of the network its is important to discuss the operation and service of the AP.

The access point (AP) is defined as “any entity that has STA functionality and provides access to the DS. ” When a station joins a BSS it must become associated with an AP. To assist in this process, the AP transmits beacons at a predetermined interval. These beacons are used by the STA to locate an AP and form an association, synchronize with the BSS, and also to handle power management traffic.

To describe the method used for power management, several other definitions are necessary. First, there are three power management modes defined in the 802.11. These are based on the condition the station is in. They are transmit, meaning the station is currently transmitting; receive, meaning the station is currently receiving; and doze, meaning the station is not able to transmit or receive. Based on these station states, the AP then defines four packet handling modes. The first of these is continuous-active-mode (CAM). In CAM a station can freely receive frames. This means that the AP is not required to buffer any of the frames. The next mode is temporary-active-mode (TAM) which is the same as CAM but on a temporary basis. The next two modes are the real mechanisms that handle the power control of the stations.

The first of the main power management modes is power-save-polling (PSP). When a station is in this mode it is asleep, and the AP buffers all packets destined for the station. The station finds out about the buffered packets by waking up just long enough and just far enough to activate the receiver and receive one of the AP's beacons. In these beacons is a traffic indication map (TIM) indicating which stations have packets waiting for them at the AP. It is not necessary for a station to wake for every AP beacon since the AP will continue to buffer the packets and list their presence in the TIM until the station wakes. The station, however, must wake at a preset interval which is an integer value of the beacon interval long enough to check the TIM. If a station notices that it has packets waiting at the AP, it wakes completely up entering CAM or TAM and sends a short poll message to the AP letting it know that it is clear to receive. The stations state in the AP table is then changed to awake. Once the AP is notified that the station is awake, it will begin to send any buffered frames it has to the station. The AP waits for each sent frame to be acknowledged before sending the next. All sent and acknowledged frames are removed from the AP's buffer.

The next of the possible power saving modes is power-save-non-polling (PSNP) mode. When a station is in this mode, it has powered down enough to save energy but has left its receiver on. In PSNP, the AP does not buffer packets for the station but rather sends them through using the regular packet exchange mechanism.

For the above operations to work properly it is imperative that the station keep the AP informed as to what mode its is currently in. These procedures require the AP to remain awake at all times. Since an AP could also be a portable computer bydefinition, it is important that the computers that are part of a BSS take turns as APs. Due to this restraint, all members of an ad-hoc network partake in a distributed beacon and synchronization algorithm defined in section 3.3.1.1 of this document.

3.3 Network structures

When interconnecting wireless networks and ensuring fair and controlled media access, there are two main structures that permeate the available literature. Further details, illustrating how these network structures are engineered as well as the difference between them, are presented in the following sections.

3.3.1 Ad-hoc networks

The current IEEE 802.11 standard defines an ad-hoc network as:

An ad-hoc network is a network created for a specific purpose, typically in a spontaneous manner. The principle characteristic of an ad-hoc network is that the act of creating and dissolving the network is sufficiently straightforward and convenient so as to be achievable by non-technical users of network facilities (i.e. no specialized 'technical skills' are required with little and/or no investment of time or additional resources required beyond the stations which are to participate in the (ad-hoc) network.)

The most characteristic feature of an ad-hoc network is its peer-to-peer nature. Figure 3.1 shows a connection structure for a peer-to-peer ad-hoc network. Note in this figure that the connection path between any two stations is shown by a dotted line and not all stations are shown as being connected to all of the others. Each computer attached to such a network communicates directly with its nearest neighbor. This peer-to-peer interaction is unlike the cellular network structure where all computers communicate only with a base station. Since the IEEE 802.11 standards standard requires communication from station to AP, in an ad-hoc environment all stations can be considered APs. The exact procedure for implementing this is discussed in section 3.3.1.1 of this document. The use of an ad-hoc type of structure allows several unique and perceived beneficial advantages some of which are discussed below.

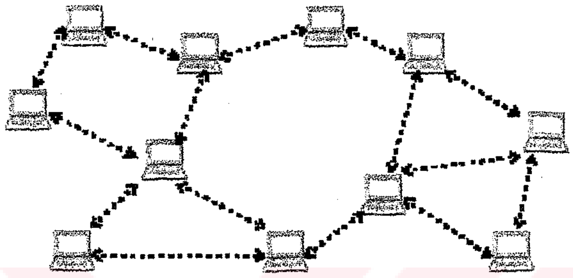


Figure 3.1: Connectivity of an ad-hoc network

One of the first advantages observable with an ad-hoc structure is the ease of set-up. This refers to the lack of pre-planning required in establishing such a network. Primarily, no cables have to be laid, no pre-arranged base stations have to be constructed, and if the frequency band used is licensed, then no pre-clearance of frequencies used within the area of establishment has to be conducted.

Of course, ad-hoc networking has its disadvantages as well. One of the main ones results because the stations are communicating with each other and the topology is dynamic. This means that no one station can assume accessibility to all of the others . For instance, a station may be able to send and receive information from two other stations that do not have direct access to each other [Stallings W (2000)]. This may require a third station to route the information sent back and forth between these two stations while not swamping the channel. A system designed using a routing algorithm to handle connectivity in a ad-hoc network is discussed in section 3.3.1.1 of this document.

One method of trying to ensure connectivity in an ad-hoc network is to implement a broadcasting or flooding strategy. This method was used in the well known and well studied ALOHA network. As numerous research has shown, the maximum channel utilization of a pure ALOHA system was only 18% . Although this throughput and subsequent maximum channel utilization was doubled with slotted ALOHA, it is clear that both of these methods provide too low a channel utilization to be feasible. [Tanenbaum, A.S. (1998)]

Another approach to using broadcasting is to sense the channel before transmitting or carrier sense multiple access (CSMA). Besides exposing the network to the well known hidden and exposed station problems described in chapter 1 , this method is problematic. The main problem comes from the fact that the decision to transmit is based solely on information collected at the transmitter and not at the receiver . If the stations were to share state information concerning the channel, the sensing may be improved . However, the very dynamic nature of a wireless network makes storing and transmitting state information infeasible . The methods outlined in the IEEE 802.11 standards standard for virtual carrier sensing effectively handle the hidden and exposed station problems. This method of virtual carrier sensing is discussed in section 3.3.1.1 of this document

Another method of establishing an ad-hoc network is for the stations to setup a temporary infrastructure. One such infrastructure is the spokesman election algorithm (SEA) . Here the stations elect one of the stations to act as a base station and transmissions are to and from base station only [Chen, K. (1998)]. The problems with this method are obvious. Since no one station can guarantee connectivity to all of the other stations, the use of one of them as a base station will cause pockets where the signal fades in and out, disconnecting and reconnecting these stations to the network. Also, since the most obvious use of ad-hoc networks is to interconnect notebook or portable computers, the use of one of these units as a base station and having it handle and control all traffic would put an unnecessary strain on its limited power supply.

It is important to remember that any method used to interconnect these units must provide a way to efficiently use their limited power supplies and allow the units to enter a sleep state while not being accessed. This means that the stations need a way to check packet headers while still asleep and only wake if the packet does indeed belong to it. It may also be required that these units to be used similarly to a router and as such they must re-transmit packets. Unless all of the power concerns of portable computers are fully addressed, an ad-hoc networking strategy cannot be considered a viable solution. The IEEE 802.11 standard uses some unique approaches for handling power saving in an ad-hoc network. First, this standard treats all members of the ad-hoc network as APs requiring them to implement the beaconing and buffering functions of the AP. In this way, the ad-hoc network only needs to maintain synchronization to arrange for power savings. The implementation and establishment procedures of an ad-hoc network via the IEEE 802.11 standard is discussed in the next section.

3.3.1.1 Models of ad-hoc networks

Below is an explanation and discussion of some proposed models for adhoc networks. This discussion starts by looking at the most popular method to date and progresses through some of the less well know proposals.

One of the most popular methods of implementing an ad-hoc wireless network is the multiple access with collision avoidance (MACA) protocol. This MACA protocol is the same method, with minor variations, used in the IEEE 802.11 for implementing virtual carrier sensing. The main ingredients of this system are its two types of short, fixed-length control packets. Utilizing primarily two control packets, the channel is effectively controlled. These control packets also help alleviate the hidden and exposed station problems.

The first of these control packets is the Request To Send (RTS). This packet is sent by the station wishing to transmit to the station it wishes to transmit to. For instance, if station **A** wants to transmit to station **B**, then it sends a RTS packet to station **B** listing the length of the proposed data transmission. Station **A** then sets a timer and waits for station **B**'s response. If station **B** is busy receiving a transmission from another station, it will not respond to the RTS. This will cause station **A**'s timer to expire, and station **A** will reschedule its transmission attempt for a later time. Note that, because the packet is small and of fixed length, it will have minimal interfering effect on the transmission station **B** is currently receiving, thus reducing the amount of collided packets and reducing the amount of packet retransmissions needed.

If in the above scenario station **B** was not busy, it would respond to station **A**'s RTS with the second of the short control packets or the Clear To Send (CTS). In the CTS, station **B** specifies the amount of data station **A** is clear to send. This could be less than the original amount requested by station **A** and as such allows station **B** to control the amount of access time station **A** is allowed. Upon receipt of the CTS, station **A** will start its data transmission. Upon completion of the data transmission to station **B**, station **A** will await an acknowledgment (ACK) from station **B**. If station **A** receives a NACK instead of an ACK, it will reschedule the transmission and try again later. Here is one area where the IEEE 802.11 differ from the pure MACA protocol. In the IEEE 802.11 standard, if two station are using this exchange, then every frame sent will be acknowledged separately. If station **A** does not receive an ACK from station **B**, it assumes station **B** did not receive the frame and reschedules transmission. There is no NACK provided in the IEEE 802.11. Instead the absence of an ACK serves as an implicit NACK. A state transition diagram for this exchange can be found in figure 3.2 below.

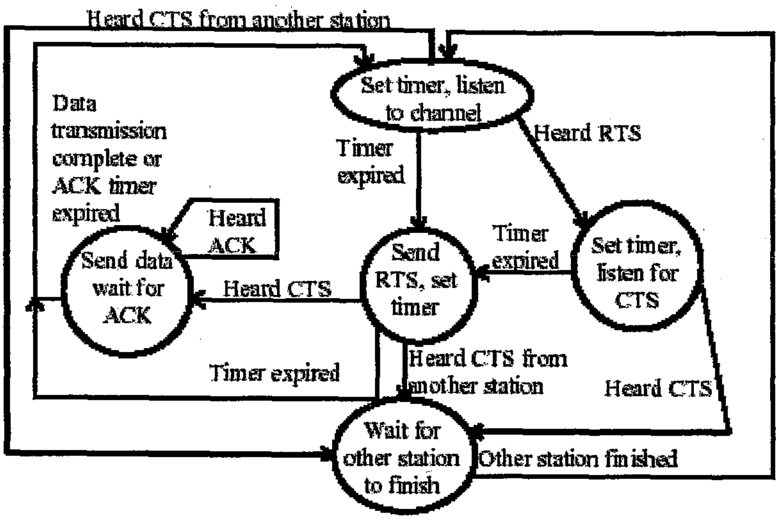

Heard CTS from another station
Set timer, listen to channel
Data transmission complete or ACK timer expired
Timer expired
Heard RTS
Heard ACK
Send data wait for ACK
Heard CTS
Send RTS, set timer
Timer expired
Set timer, listen for CTS
Timer expired
Heard CTS from another station
Heard CTS
Wait for other station to finish
Other station finished

Figure 3.2: State transition diagram for virtual carrier sensing

There are several clear advantages to this transmission exchange. First, referring to figure 3.2, note that any station hearing a RTS for another station refrains from using the channel until the time for a return CTS has expired. In this way, if no CTS is heard, then the station hearing the RTS does not delay sending and the channel wastage associated with the exposed station problem described in chapter 1 is avoided. Now if a station hears a CTS it will refrain from using the channel until sufficient time has expired for the DATA-ACK exchange length specified in the CTS. Also note that a station does not have to have heard an RTS prior to hearing the CTS to refrain from sending. By listening-in to the control packet exchanges, a station can determine the presence of a hidden station and refrain from sending thus overcoming the hidden station problem.

The IEEE 802.11 standard makes a small modification at this point. Since the highest probability of a collision is right after the ending of a transmission, the 802.11 requires a random back-off period before the station can attempt another RTS broadcast. The standard defines a net allocation vector (NAV) which contains the amount of time for transmissions discovered by overhearing the RTS or CTS or both. The additional back-off time is added to this NAV to determine when the station can attempt another broadcast .

The back-off timer is a key design issue and several improvement have been suggested. The original implementation of MACA used a binary exponential back-off (BEB) which doubled with every unsuccessful RTS-CTS exchange and reduces to zero with a successful RTS-CTS exchange. The problem with this method is that a station with a lower BEB timer could in fact monopolize the channel making the access unfair. An improvement to the system is to include the current value a station has for its back-off timer in the header of the control packets. Therefore, whenever a successful RTS-CTS exchange occurs, all stations who hear the exchange will have the same back-off timer value and all will have a fair chance at channel access [Bharghavan, V. & Demers, A. (2001)].

There are other modifications to the simplified exchange outlined above. For instance, if a station **A** does not hear an ACK in response to the data it sent to station **B**, it will reschedule the transmission. When station **A** sends the new RTS to station **B** to re-send this data, station **B** will respond with an ACK instead of a CTS. This helps eliminate channel wastage due to redundant data. Another enhancement is if station **A** had sent an RTS to station **B** but did not hear a CTS in response, station **A** would reschedule the transmission attempt for a later time. This later time would also include a random back-off period. When station **B** becomes clear, instead of waiting for station **A** to attempt another transmission, station **B** can send a request for a request to send (RRTS) to station **A.** Station **A** will respond with a RTS and the transaction will follow the usual path outlined in figure 3.2.

As mentioned earlier, in the IEEE 802.11, all stations attached to an ad-hoc network are considered APs. If this assumption is made, then the other communication conditions can continue to operate as per the 802.11 specifications. The only changes are now that all stations send beacons which are used for a distributed synchronization process. The basic procedure is that all stations have a beacon timer. A station will wake just prior to the expiration of this beacon period and determine a random time to wait before transmitting a beacon. The station listen to the channel until the additional back-off timer has expired. If the station hears a beacon from another station in this interval, it cancels its scheduled beaconing, and if no frames are waiting for it, returns to sleep. If the station does not hear a beacon from another station prior to the random back-off expiration, it will send a beacon. Also, a station will notify other stations of waiting frames during this beaconing period.

One issue that has not been addressed yet is how to establish an ad-hoc network. The IEEE 802.11 uses a simple method for network start-up. Since all stations on an ad-hoc network are APs, they are required to send beacons at regular intervals. A station wanting to join an ad-hoc network simply listens for a beacon until a preset time has expired. If no beacon is heard, it assumes that no other stations are attached yet and starts the beaconing process. The rest of the operation follows the beaconing method outlined above.

Another method of handling ad-hoc access is with optimum channel utilization multi-access (OCUM) protocol described in . The key architectural component of this protocol is the use of two channels. All stations have access to a message channel and a busy-tone channel, called the M-Channel and the B-Channel, respectively. The M-channel is broken into slots and utilized in a hybrid slotted ALOHA fashion. The basic principle is that when a station is receiving data, it transmits a busy tone on the B-channel. A station wishing to transmit must first monitor the B-channel to see if the M-channel is free. The slotted M-Channel has a preamble section where stations wanting to transmit contend for slots. When a station is ready to send, it listens to the B-Channel. If a busy tone is present, it refrains from sending. If it does not hear a busy tone, it then transmits a preamble on the MChannel. The station it wants to send to will broadcast a busy tone on the B-Channel if the preamble is successfully received. In this way, the B-Channel acts as a clear to send for the transmitting station.

The main problem with this implementation is that, although the busy tone eliminates the hidden station problem, it can in fact exaggerate the exposed station problem. Interestingly enough, the authors report a maximum throughput of 78% in their simulations or more than double the throughput of a slotted ALOHA protocol [Bharghavan, V. & Demers, A. (2001)]. Comparing this with the popular wired IEEE 802.3 (Ethernet) protocol, which has a maximum throughput of 85% , this performance is surprisingly good. Although on the surface it would appear as if the exposed station problem is not as severe a problem as the hidden station, it is also possible that the authors did not consider the exposed station problem in their modeling. This over sight would have caused the authors to assume the channel was busy at times when it could have been utilized for non-interfering traffic. Before any real conclusions can be drawn concerning how these results relate to the exposed station problem, further simulations of the model would need to be conducted.

Another interesting approach to ad-hoc wireless networking is that proposed by Charles Perkins . He has proposed treating each mobile unit as a router and using a distance routing protocol in the MAC layer to achieve and maintain connectivity. He points out that the standard routing algorithms need to be modified. This is because one of the main problems with standard routing algorithms is that they tend to exhibit their worst performance when the links they are routing over display a dynamic behavior. As discussed previously with wireless networking, one of the main characteristics of the links between stations is their very dynamic character. This has led to some revisions of the basic distance routing protocols to make them applicable to wireless networks. The proposal from Perkins follows.

Like all routing protocols, the main objective is to take any packet received and re-transmit it to the next hop on route to its destination via the shortest path. This protocol is no different. Each station maintains its own view of the network topology. It associates with each accessible station a link cost. This cost is a metric of the number of stations the packet has to transverse before reaching its destination. Each station shares its link cost information with all other stations in its area via a technique such as flooding. Each station updates its network view whenever it receives the broadcast of another station concerning link states. By forwarding packets in this manner, sharing information concerning links periodically, and barring the formation of loops, the packets can be assured of reaching their destinations via the shortest paths.

A problem associated with the distribution of information concerning path lengths or link values in this manner is that it could cause a form of oscillation. In other words, if the link cost associated with a link keeps changing up and down by a metric of one then the shortest path to that link could change with every link state broadcast from neighboring stations. Also it is possible that one of the stations may have old, out of date information in its table and as such could cause the other stations nearby to change to the faulty data. To overcome these problems, Perkins has proposed that each entry in the link state table of each station should contain a sequence number and a settling time for each link to a reachable station. The sequence numbers are generated by the destination stations and the settling time is the length of time the entry must remain in the table before the station shares this link information with its nearest neighbors.

One of the problems with this and other routing algorithms is the formation of loops. Both short and long lived loops are possible. Also, the distribution of the link state information must not swamp the channel. Therefore limitations must be placed on when this information is worth the expense of transmission. One case where transmission of the information should occur immediately is if a break in a link is discovered. Failure to share link breakage information as soon as it is discovered can cause lost packets. When a break is discovered, the link is assigned a value of infinity and this information is shared immediately with all stations in hearing range. Any station possessing an alternate route to the station with the broken link, responds immediately with this information. This helps to maintain the connectivity in the network, especially under the dynamic link condition of a wireless medium. More details concerning this method of implementing a ad-hoc network can be obtain by referring to .

Other methods of implementing ad-hoc networks exist but most authors avoid this structure in favor of a cellular network architecture.

3.3.2 Cellular

Cellular technology has been used widely by the telephone industry for several years. The main structure is to break and area into a continuous group of cells. The frequency band is then broken down so that cells adjacent to each other do not share the same region of the spectrum. Cells that do use the same frequency are separated by enough distance as to not interfere with each other. An example of this is shown in figure 3.3. In this figure, only three frequencies are used to cover the area. As can be seen in figure 3.3, the spectrum can be reused and efficiently divided to cover large geographical areas with only limited frequency bands. This same idea can be applied to an indoor environment. Taking into account structural features, cells can be assigned dependent on the building layout.

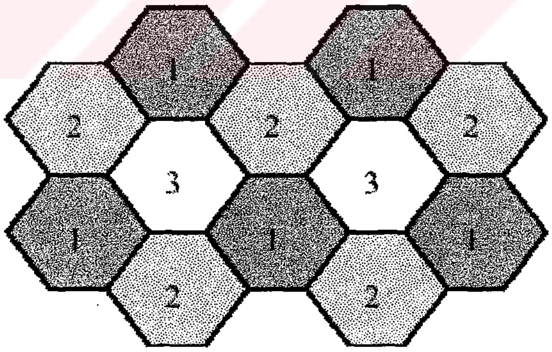


Figure 3.3: Cellular network frequency reuse layout

A cellular network is usually structured with each cell containing a base station networked to a higher level network. All stations in the cell communicate with the base station, there is no peer-to-peer traffic. A typical layout for a cellular network can be found in figure 3.4. Note that in this figure, the wireless links are shown by dotted lines and the wired link between base stations is shown with a solid line. Also note that all stations in the network are in contact with at least one base station. One important characteristic of this arrangement is that the uplink traffic or base station access via a mobile station is relatively small compared to the down link traffic from the base station to the set of mobile stations . The main reason for this discrepancy in traffic patterns is the traffic most envisioned as being generated via a mobile station is file transfers or similar large data transfers from a higher level network . With this type of traffic, only a brief message needs to be sent via the uplink to generate a large response via the downlink. Noticing this discrepancy between the volume of uplink versus downlink traffic can aid in channel assignment and collision avoidance methods. Some of these methods are implemented in the systems discussed in section 3.3.2.1 below.

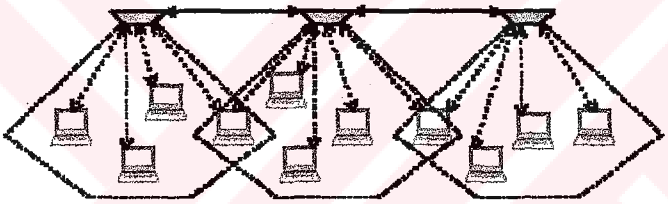


Figure 3.4: Cellular network station layout

One important consideration when designing a cellular network structure is the ability to maintain connection to the network when moving into another cell. Although this is more of a concern with cellular telephones than with indoor cellular computer networks, some of the smaller pico-cellular proposed structure may indeed cause a mobile computer user to change cells during operation. This problem concerns how far to overlap the cells at the boundaries and how to handle the hand-offs between cells [Chen, K. (1998)]. The cells must overlap at their boundaries to ensure connectivity, but there must be some method of handing control over to the base station in an adjoining cell when the station is sufficiently out of range. Many methods have been suggested for solving these concerns and many are addressed in the discussions of cellar networks which follows in section 3.3.2.1.

One of the main advantages of a cellular structure, versus the ad-hoc structure discussed previously, is that the coverage plan becomes definable . All stations either have contact with a base station or they are not connected to the network. There is no need to worry about the connection between mobile stations since all transmission go through the base station which acts as a controller and repeater.

Cellular networks also have disadvantages. One of the disadvantages of a cellular network when compared to an ad-hoc network is that cellular networks use base stations and these require careful planning and placement prior to operation . Also, although coverage is simplified, the need for cell-overlap introduces a problem referred to as self interference. This problem occurs when a station is in the overlap region between two cells and is registered with both base stations . Two stations in figure 3.4 exist in this state. In this scenario, it is possible for the station to receive transmissions from both cells simultaneously thus colliding at the station. Another problem is the up-down collision problem. If one channel is used to handle both uplink and downlink traffic, then it is possible for one station to send an uplink while another is trying to receive a downlink thus colliding with the transmission [Chen, K. (1998)].

In spite of some obvious problems, cellular networks have been used successfully in the telephone industry for years. This has lead to well established methods of dealing with most problems that can arise and maximizing reliable throughput.

3.3.2.1 Cellular system models

In this section we examine a few suggested implementation of an indoor wireless cellular networks. Some of these models implement broadband services and are suited for the integration of telephone, video, and computer data services on one network. All of the below models break the coverage area into cells, use base stations to control the channel in the cell, and connect multiple cells with a backbone network. Their differences come mainly from how they handle items, such as in-cell traffic and between-cell traffic.

One of the techniques for handling traffic within a cell is via a central queue. This is the method utilized in [Ioannou, Z. & Gurcan, M. & Tan, H. (1998)]. The basic operation of this protocol is to have each base station maintain a central queue for downlink traffic. When a station has information to send, it sends a reservation packet to the base station letting it know how many packets it needs to buffer in the queue. The base station will respond with an acknowledgment and can adjust the number of packets the base station wishes to buffer. Upon a successful buffering request and acknowledgment exchange, the station places its packets in the base station buffer. The base station will transmit the packets out of the buffer on a first in first out (FIFO) basis. To help alleviate collisions, the downlink channel is slotted into single data length slots. All packets leaving the base station leave on a slot boundary. The uplink channel, however, is not slotted and it is possible for collisions to occur on this channel.

Some of the interesting features of this particular architecture is that the cells are multichanneled. This means that each cell can be allocated more or less channels for longer or shorter service periods as the traffic demands dictate. This dynamic allocation of channels allows the network to accommodate varying traffic patterns in the indoor office environment. Also, if a cell is experiencing excessive uplink collisions, then an additional uplink channel can be added thus alleviating some of the problem. One of the negative side effects of this implementation is that it introduces higher packet delays. However, if this system is used in a data network where these higher packet delays can be tolerated, then the efficient channel allocation method offered by this protocol is a definite advantage.

Another cellular network implementation that utilizes a form of dynamic channel allocation is the one described in . This method of channel allocation is based on what is termed as an interference graph. The main concept is that the coverage area is broken into cells, each of which is controlled by a base station. The base stations themselves are connected via a fiber optic high speed backbone network. At regular intervals, the base stations interrogates the stations in their cells to determine the traffic demands. These traffic demands are then reported back to the controller of the base stations. The controller then plots an interference graph to determine the minimum number of time slots and minimum number of frequencies needed to handle the traffic in all of the cells. These time slots, which now guarantee maximum throughput without interference from adjacent cells, are distributed to all the cells which had reported traffic. The maximum amount of time before the next interrogation period is determined by the base station which had reported the most traffic. As far as the stations utilizing the cells are concerned, they buffer their packets into a central buffer using a separate uplink channel similar to the method described in . This implementation also maximizes downlink traffic and assumes that the traffic pattern of a cellular network will be heavily lopsided with the majority of traffic being downlink transmissions.

An interesting approach to indoor cellular networks is the pico-cellular network described in . This network is designed for the small coverage area associated with indoor networks, namely pico-cell. A pico-cell is defined as a cell that is only tens of meters in diameter. These small size cells are still large enough to serve an entire room. The pico-cellular network outlined in this article is really designed for traffic that would be hindered by discontinuity while moving between cells. This would not necessarily apply to data traffic but the model is interesting in how it handles movement by considering the building structure as part of the plan.

The basic architecture in consists of three main entities, namely server host (SH), which tracks the mobile units called mobile host (MH), and mobile support station (MSS), which are basically base stations that provide the access for the MH. The basic concept is to provide a mechanism where the next packet is always accessible to the MH, no matter which MSS's domain it is currently in. To handle this, the coverage area is broken into groups. These groups depend on not only where the next MSS is but also on the building layout. For instance, if a MH is in a room and a separate MSS handles the hall, while yet another MSS handle the room next door, then the MH need only belong to a group containing its current MSS and the MSS covering the hall right outside the door. If the MH were to venture into the next room, it would only be able to do so by venturing out into the hall first. This consideration of building layout can substantially reduce the size of the group a MH belongs to. It is the job of the SH to keep track of the MHs in its domain and forward packets in a multicast fashion to the each MH's group.

The easiest way to grasp how this protocol works is to look at how a MSS handles packet delivery. When a SH send a packet to a MH, it uses the MH's group ID which identifies the group of MSSs that receive the packet. When a MSS receives a packet for a MH, it first checks to see if the MH is in its current cell. If the MH is in the cell, the MSS simply delivers the packet. If the MH is not in the cell, the MSS buffers the packet and holds it until the MH enters its cell or the SH drops it from the group handling packets for that MH. When a MH enters a new cell it must register with the new MSS. Part of the registration procedure requires the MH to notify the MSS of the next packet number it is expecting. At that point, the MSS can discard any older packets freeing buffer space. With this method, the MH never notices any interruption in service.

The problems with the above plan when handling data communication are obvious. The slow or limited movement expected from a user accessing the network via a portable computer makes the buffering of packets unreasonable. However, some of this could be alleviated by adding an expiration time to the packets. This would allow the MSS to discard expired packets and reduce the needed buffer space. Another possible solution comes from the observation that data communications are not hampered as severely as audio communications if there is additional overhead when switching cells. Also, since the changing of cells is not as common with portable computers as it is with cellular phones, the additional buffering in adjacent cells may introduce unneeded overhead. As a result, the authors have proposed making the group size for data communications via a portable computer a group of one meaning just the current cell the portable computer is operating in. The nicest feature with this protocol is the flexibility it offers when handling several different kinds of services on the same network.

Other cellular network models exist and the list is virtually endless. As wireless communication has gained popularity, the methods of structuring cellular networks have also multiplied. The MAC protocol discussed in the IEEE 802.11 does lend itself nicely to a cellular structure. The BSS can be the member of a single cell with the AP filling in for the base stations. The IEEE 802.11 also provides a mechanism for connecting the wireless BSS to a wired 802.xx compliant network. The mechanism is called a Portal and it handles the movement of packets between the two interfaces. These mechanisms do not strictly define the layout or operation of a cellular network but only provide tool by which one could be established.

3.4 Code division multiple access (CDMA)

Another method proposed for handling a multiple access channel comes from the use of different orthogonal spread-spectrum codes for each of the different stations on the network. The basic principle is that the codes will not interfere with each other, since they are uncorrelated and will only appear as noise to any other signal on the channel. Viterbi contends that there are a large group of such codes. Therefore, the number of stations transmitting simultaneous is only limited by the number of signals the base station can receive simultaneously . He also shows that with the right receivers this group is large enough to accommodate practically any network . Others, however, have found that there is an actual limit to the number of codes that are truly orthogonal. This, however, is still not seen as a limitation since this group is still large enough to handle most traffic in a cell and with careful assignment of codes, they can still maintain orthogonality .

Besides the obvious advantage of multiple transmissions, another advantage of CDMA is that when each station is assigned its own code, further addressing can be eliminated. Because only one station would have each code, only that station would be able to receive data that was sent using the code. Although the IEEE 802.11 does not explicitly eliminate CDMA systems, it does not define such an implementation. The only use of multiple codes on the same channel is under the 2.4 GHz DSSS section of the standard and that only provides for the multiple codes to handle additional traffic .

3.5 Time division multiple access (TDMA)

The tried and true method of allocating a channel has always been time division multiple access (TDMA). The main principle of this method of access is to divide the channel into fixed time intervals and send information in these periods. The advantages are that the reservation of slots makes the handling of the channel simple . Also, since the channel is broken into fixed size slots and permission must be obtained before transmitting in one of the slots, collisions can be controlled if not eliminated. One of the disadvantages is that overhead is required to decide who needs and should have access to the slots. This is normally handled with some sort of contention period where the stations needing to transmit contend for permission. Several variations of TDMA exist in the literature going all the way back to the slotted ALOHA system. This thesis will not go into depth into any of these systems.

3.6 ATM

One method of data traffic control that has gained significant attention in recent years is asynchronous transfer mode (ATM). Although ATM specifications usually cover signaling methods as well as cell header and trailer information, no specific wireless interface has been defined yet. This has not limited the implementation of ATM over a wireless interface. In fact many have argued that ATM is the ideal transmission method for a wireless medium based on several factors some of which are discussed below. Because the ATM layer can be handled fully independent of the physical medium, but is used to control channel access and throughput , it is best discussed as part of the MAC sublayer.

As pointed out in , ATM networks have many features that make them attractive and these are recapped here. Some of the features that have caused an outpouring of support for ATM cell transport in a wireless network are its flexible bandwidth, accommodation of multiple service types, suitability for cellular structures, easy interface with higher speed backbone networks, and ability to provide end-to-end service. Its efficient multiplexing techniques also allow the smooth and efficient handling of burst data as well as isochronous, broadband services on one network. The virtual channel addressing method utilized by ATM cell transport is ideal for a cellular structure where each cell is isolated from the others. The dynamic addressing and locally unique VCIs (virtual channel identifiers) versus a globally unique addressing, allows for flexibility and control within each cell to be independent of the other cells. ATM is expected to continue to gain popularity as the need to integrate various types of service on one wireless network increases in the future. As a result, a deeper look into the principles and architecture of this protocol follows.

3.6.1 Basic principles and architecture

The architectural discussion which follows is taken primarily from . The main functions of the ATM protocol deal with data flow. The ATM layer is responsible for the multiplexing and de-multiplexing of cells onto the physical medium and making the connection appear as a constant stream of cells. To accomplish this, each connection is given upon establishment a virtual channel identifier (VCI) which is a unique mapping between the sending station and the receiving station. This mapping, however, is not globally unique, only locally unique, and as such the VCI only needs 16 bits compared to the 48-bit addressing suggested in the IEEE 802.11 standard .

Another service provided by the ATM layer that is attractive, especially when integrating several different types of services on one network, is the ability to adjust the quality of service (QoS) including bit rate and error rate to fulfill the needs of each service. This allows several heterogeneous services to coexist on the same network and not deteriorate the service level of other higher quality services.

Of course, like all network layers, this layer attaches a header before release to a lower layer and removes the header prior to release to a higher layer. The 5 octet ATM header consists of the following fields and each serves a specific control or routing purpose. These fields, as they fit in an ATM header, are shown in figure 3.5. The first field, which is 4 bits in length and optional, is the generic flow control (GFC) field. Its purpose is to regulate the uncontrolled and controlled transmission as well as control access from a number of terminals at the user network interface (UNI) . The next field, which is 8 bits in length, is the virtual path identifier (VPI). This field is basically used to establish a path from one station to another and in a wireless network, could be used to establish a path between cells. Note that the GFC field is optional and in its absence, the VPI field extends to 12 bits thus increasing the number of virtual paths that can be defined. The next field is the VCI. It is 16 bits in length and is a locally unique value given to each active connection within a network. The combination of the VPI and VCI are used by the ATM layer to determine the recipient. One unique characteristic of this approach is that the VCI is assigned at connection establishment and released when the connection is closed. This dynamic addressing assignment strategy allows for adjustments in network size and structure during operation. The next field, which is only 3 bits in length, is the payload type identifier (PTI). The purpose of this field is to help identify the type of information carried in the 48 octet payload section of the cell. The next bit is the cell loss priority (CLP) bit. Its purpose is to allow the dropping of cells in connections where such loss is not critical. Basically, this bit is used to identify cells that may be dropped when the congestion of the network mandates such measures. The final field, which is 8 bits in length, is the header error control (HEC) and is used not only for header error detection and correction but is also a synchronization of the cell at the receiver. Some have argued that on a wireless network the error correcting capabilities built in to these networks make this field redundant . It is also argued that the synchronization function can also be achieved using radio system framing and as such this field could be eliminated across a wireless network without a loss in performance . However, the IEEE 802.11 protocol uses a HEC in the header of the PLCP frames and has adopted the 32 bit CRC used in the ATM adaptation layer 5 (AAL_5) for error control of MAC packets .

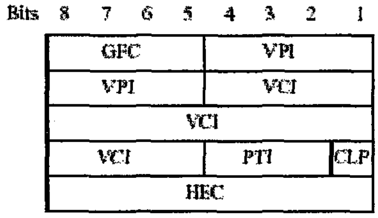


Figure 3.5: ATM header

Some adjustments do need to be made to the ATM signaling protocol for adaptation into a wireless format. For instance, the dynamic VCI and VPI addressing of ATM needs to provide for movement of users between cells in a cellular network . The smooth transition between cells requires some form of hand-off where VCI and VPI information can be exchanged without a noticeable delay by the user . Also, because of the noise variance of a wireless network, the QoS may need to be adjusted during the circuit lifetime. This is especially true if the user is moving between cells where one cell may have a larger traffic flow than the other one. This type of negotiation again needs to be conducted without losing the circuit . These factors must be addressed if ATM is to be implemented on a wireless interface.

3.6.2 Advantages and disadvantages

One of the most important advantages of an ATM network is its ability to support simultaneous virtual connections . Some have even pointed out that both ATM and CDMA share many similar characteristics making them ideal partners in a wireless environment . The largest debate, however has raged over the small cell size of the ATM cells. Several have argued that IP packets will be fragmented too much if they are transported over an ATM network and such fragmentation will reduce throughput of the network. This issue has been the focus of numerous recent studies and one such study's results are covered in the next section.

3.6.3 IP over ATM

Since most networks now speak IP, the real test of ATM is its performance at handling IP traffic. In one of the more interesting studies recently published, the authors started by first analyzing the typical IP traffic on a typical office network . This study then analyzed how this traffic would be affected if it were sent over an ATM network. Finally an actual network was examined to check the theoretical results. Because of the thoroughness of this study it will be discussed in further detail. The first part of this study that proved interesting was the IP traffic pattern of a typical office network. The authors studied this traffic flow and found that 70% of the traffic was what they called tiny IP packet, an additional 14% was medium IP packets, and the final 16% was large IP packets . These packets divided into Equivalent ATM Cells (EAC) as the tiny IP packets took 2 ATM cells, the medium IP packets took 3 ATM cells and the large IP packets averaging 13 ATM cells. This meant for a typical traffic pattern, one hundred IP packets would translate into 390 ATM cells. However, when RFC 1144 concerning header compression was applied, the tiny IP packet could then fit into one ATM cell. This meant that, by utilizing header compression, the typical 100 IP packets fit in 290 ATM cells. Since these tiny IP packets account for 70% of the typical office network IP traffic, the IP-ATM interchange would not be a hindrance. The next question answered in this article was how effective was the ATM cell structure at handling the IP traffic load. In other words, did the use of ATM cell for transporting IP packets leave too much wasted space inside the ATM cells? These researchers found that in all three identified categories of IP traffic, the loads were best served when payloads were between 40-96 bytes. Therefore the 48 byte ATM payload was not as bad as had been predicted by many in the IP network community.

CHAPTER FOUR
SECURITY

Given the increased productivity and growing popularity of wireless communications in general, and wireless data communications in particular, this cahpter outlines the security implications of WLANs.

4.1 Basics of WLAN Security

IEEE 802.11 contains several security features, such as open system and shared key authentication modes, the Service Set Identifier (SSID), and Wired Equivalent Privacy (WEP). Each of these features provides varying degrees of security and each is covered in this section. Also covered is information on how RF antennas can be used to limit, and in some instances shape, the propagation of the WM.

4.1.1 Limiting RF Propagation

Before any other security measures are implemented, it is important to consider the implications of RF propagation by APs in a wireless network. Chosen wisely, the proper transmitter/antenna combination can be an effective security tool that will help limit access to the wireless network to only the intended coverage area. Chosen poorly, they can extend a network beyond the intended area into a parking lot or farther.

Primarily, antennas can be characterized by two features.directionality and gain. Omni- directional antennas have a 360-deg coverage area, while directional antennas limit coverage to better-defined areas. Antenna gain is typically measured in dBi[2] and is defined as the increase in power that an antenna adds to an RF signal.

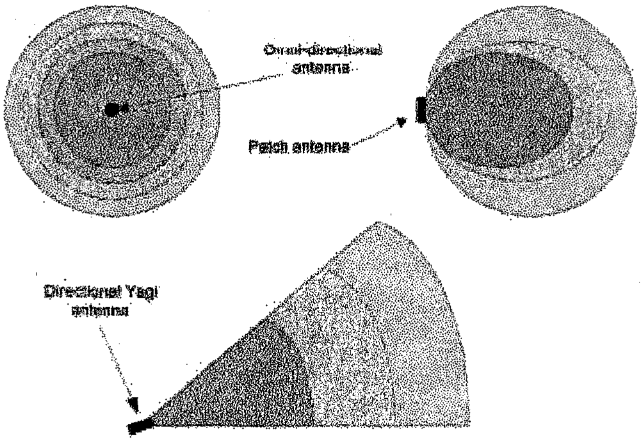


Figure 4.1. RF propagation patterns of common antennas.

Because current 802.11 products make use of the unlicensed Industrial, Scientific, and Medical (ISM) 2.4-GHz band, they are subject to the rules promulgated by the FCC in 1994 for spread spectrum use. These rules specify that any antenna sold with a product must be tested and approved by an FCC laboratory. To keep end users from using incorrect or illegal antennas with 802.11 products, the FCC also requires that any APs capable of using removable antennas must use nonstandard connectors.

In the U.S., the FCC defines the maximum Effective Isotropic Radiated Power (EIRP) of a transmitter/antenna combination as 36 dBm, where EIRP = transmitter power + antenna gain - cable loss. Essentially, this means that as transmitter power increases, antenna gain must decrease to remain below the 36 dBm legal maximum. For example, a 100-mW transmitter equates to 20 dBm. This transmitter combined with a 16-dBi antenna produces a total of 36 dBm, the legal limit. To increase antenna gain, we would legally be required to reduce transmitter power. In practice, most transmitter/antenna combinations sold today are well below the FCC maximum of 36 dBm.

The implications of all this are that transmitter power/antenna gain combinations are strictly regulated and limit the area that can be legally covered by any single AP. When designing WLANs, it is important to perform a thorough site survey and consider the RF propagation patterns of the antennas in use and the effective power of the transmitter/antenna combination. Also, because the ISM band is essentially open for use by anybody without licensing, it is important to consider the possibility of denial of service (DOS) from otherwise benign sources such as 2.4-GHz cordless phones. Finally, consider that a potential attacker may not be playing within the FCC rules. A resourceful attacker may be using high-power transmitters, high-gain antennas, and/or more sensitive receivers. Each of these can increase the effective range of wireless networks.

4.1.2 Service Set Identifier (SSID)

IEEE Std. 802.11b defines another mechanism by which to limit access: the SSID. The SSID is a network name that identifies the area covered by one or more APs. In a commonly used mode, the AP periodically broadcasts its SSID in a beacon. A wireless station wishing to associate with AP can listen for these broadcasts and can choose an AP to associate with based upon its SSID.

In another mode of operation, the SSID can be used as a security measure by configuring the AP to not broadcast its SSID. In this mode, the wireless station wishing to associate with the AP must already have its SSID configured to be the same as that of the AP. If the SSIDs are different, management frames sent to the AP from the wireless station will be rejected because they contain the incorrect SSID and association will not take place.

Unfortunately, because management frames on 802.11 WLANs are always sent in the clear, this mode of operation does not provide ádequate security. An attacker can easily listen on the WM for management frames and discover the SSID of the AP. Many organizations rely upon the SSID for security without considering its limitations. This is at least partly responsible for the ease with which some WLANs are compromised.

4.1.3 Authentication Modes

Before an end station can associate with an AP and gain access to the WLAN, it must perform authentication. Two types of client authentication are defined in 802.11: open system and shared key.

4.1.3.1 Open System Authentication

Open system authentication is a very basic form of authentication that consists of a simple authentication request containing the station ID and an authentication response containing success or failure. On success, both stations are considered to be mutually authenticated.

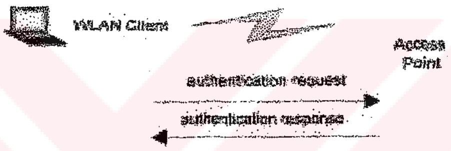


Figure 4.2. Open system authentication.

4.1.3.2 Shared Key Authentication

Shared key authentication is predicated on the fact that both stations taking part in the authentication process have the same .shared. key. It is assumed that this key has been transmitted to both stations through some secure channel other than the WM itself. In typical implementations, this might be set manually on the client station and the AP. The first and fourth frames of shared key authentication are similar to those found in open system authentication. The difference is that in the second and third frames, the authenticating station receives a challenge text packet (created using the WEP Pseudo Random Number Generator (PRNG)) from the AP, encrypts it using the shared key, and sends it back to the AP. If, after decryption, the challenge text matches, then one-way authentication is successful. To obtain mutual authentication, the process is repeated in the opposite direction. The fact that most attacks on 802.11b WLANs are based on capturing the encrypted form of a known response makes this form of authentication a very poor choice. It gives would-be hackers exactly the information needed to defeat

WEP encryption and is why shared key authentication is never recommended. It is better to use open authentication, which will allow authentication without the correct WEP key. Limited security is still maintained because the station will not be able to send or receive data correctly with an invalid WEP key (Paulo 2001, 12).

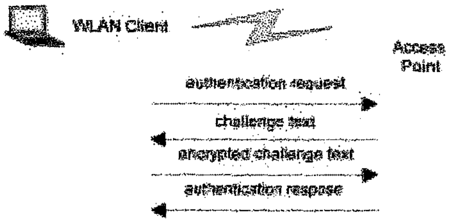


Figure 4.3. Shared key authentication.

4.1.4 WEP

As stated by the IEEE, WEP is designed to protect users of a WLAN from casual eavesdropping and was intended to have the following properties:

Reasonably strong encryption. It relies on the difficulty of recovering the secret key through a brute force attack. The difficulty grows with key length.
Self-synchronizing. No need to deal with lost packets. Each packet contains the information required to decrypt it.
Efficient. It can be reasonably implemented in software.
Exportable. Limiting the key length leads to a greater possibility of export beyond U.S. borders.

The WEP algorithm is essentially the RC4 cryptographic algorithm from RSA Data Security, Inc. It is considered a symmetric algorithm because it uses the same key for enciphering and deciphering the plaintext Protocol Data Unit (PDU). For each transmission, the plaintext is bitwise XORed with a pseudorandom keystream to produce cyphertext. The process is reversed for decryption.

The algorithm operates as follows:

*It is assumed that the secret key has been distributed to both the transmitting and receiving stations by some secure means.
*On the transmitting station, the 40-bit secret key is concatenated with a 24-bit Initialization Vector (IV) to produce a *seed* for input into the WEP PRNG.
*The seed is passed into the PRNG to produce a stream (*keystream*) of pseudo-random octets.
*The plaintext PDU is then XORed with the pseudo-random keystream to produce the cyphertext PDU.
*This cyphertext PDU is then concatenated with the IV and transmitted on the WM.
*The receiving station reads the IV and concatenates it with the secret key, producing the seed that it passes to the PRNG.
*The receiver.s PRNG should produce the identical keystream used by the transmitting station, so that when XORed with the cyphertext, the original plaintext PDU is produced.

It is worth mentioning that the plaintext PDU is also protected with a CRC to prevent random tampering with the cyphertext in transit. Unfortunately, the specification does not include any rules regarding use of the IV, except to say that the IV *may* be changed .as frequently as every MPDU.. The specification does, however, encourage implementers to consider the dangers of poor IV management. This is in some part responsible for the ease with which some WEP implementations are compromised.

CHAPTER FIVE
CONCLUSION

5.1 Conclusion

The intent of this thesis was to investigate some of the issues concerning the engineering of a wireless computer network in LANs especially in an indoor environment. During this investigation attention was paid to those areas that show a marked difference when implemented in a wireless versus a wired network.

One of the key areas investigated and discussed was the decision to move carrier detection and collision avoidance when implementing a wireless network to the MAC sublayer instead of the physical layer that is used in a wired network. In this investigation, the common wired network implementation was reviewed in chapter 2 followed by a discussion of why this technique would not work in a wireless environment. In chapter 3 a technique called virtual carrier sensing was explained as well as a discussion of how this technique overcomes the problems associated with the wired network approach. Also discussed in that chapter is the busy tone approach to the problem and oversights made in this solution.

This thesis also discusses some of the unique design issues involved in designing a wireless network. One area discussed was the need for interference suppression. This discussion was one of the key focuses of chapter 2. This chapter focused on signaling techniques and receiver and transmitter designs that help alleviate the interference experienced in LANs. Some of the other key issues discussed were the need for power management when using portable computers, and the main two models of wireless networks: ad-hoc and cellular. In the discussion of ad-hoc networks, some of the key issues were the establishment and maintenance of an ad-hoc network. Several proposed models of such networks were presented and discussed in chapter 3. In the discussion of cellular networks, the key issues were the cell boundaries and movement between cells, and several proposed models handling these problem were presented also in chapter 3. Other interesting points focused on in this thesis are spread spectrum signaling and the rules governing its use in the ISM bands covered by section 15.247 of the FCC regulations, and ATM protocols and their use in conjunction with IP protocols.

As a result of the security analysis we can say that the current IEEE 802.11 standard is known to lack any viable security mechanism, because the transport medium is shared potentially beyond the physical security controls of the organization and permits attackers easy and unconstrained access. So strong access control and authentication become essential in protecting the organization's information resources. Fortunately, attacks can easily be prevented through IEEE 802.11 management messages and additional steps ensuring the synchronization of the various state machines.

The purpose of this thesis was to identify the stops and to discover and evaluate techniques used to overcome these stops in implementing wireless LAN networking. Having read this thesis, the reader should be left with a sense of what is involved in the design and implementation of the physical layer and MAC sublayer of wireless LAN networks.

REFERENCES

Bantz D. and Bauchot F. Wireless LAN Design Alternatives *IEEE Network,* March/April 1998.

Bates R. J., Wireless Networked Communications. (McGraw-Hill: New York, 2000).

Bharghavan V., Demers A. MACAW: A MediaAccess Protocol for Wireless LAN's ACM SIGCOMM Computer Communication Review, October 2001.

Chen K. Medium Access Control of Wireless LANs for Mobile Computing IEEE Network, September/October 1998.

Dixon, R.C. Spread Spectrum Systems with Commercial Applications. (John Wiley & Sons, Inc. New York, 2000).

Ioannou Z., Gurcan M., and Tan H. Third Generation Wireless Information Networks (McGraw-Hill: Boston,MA, 1998).

Jensen M., Abidi A. Wireless Personal Communications: Trends and Challenges, (USA: Sybex Press 1999).

Stallings W. Data and Computer Communications, (Prentice-Hall, International Inc. New Jersey 2000).

Tanenbaum A. S., Computer Networks. (Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1998).

Viterbi A., Wireless Digital Communication: A View Based on Three Lessons Learned. IEEE Communications Magazine.1997.

Zeigler R. and Cioffi J. Wireless Personal Communications (IEEE Publishers: Boston, MA, 1999).