Performance Analysis of Wireless Local Area Networks (WLANs)

by

Eng. Tamer Mohamed Samir Khattab

A thesis submitted to the
Faculty of Engineering at Cairo University
in partial fulfillment of the
requirement for the degree of
MASTER OF SCIENCE
in
ELECTRONICS AND COMMUNICATIONS ENGINEERING

Under the supervision of

Prof. Dr. Mahmoud T. El-Hadidi

Professor of Computer Networks
Electronics and Communication Dept.
Faculty of Engineering
Cairo University

Dr. Hebat-Allah M. Mourad

Assistant Professor of Communication
Electronics and Communication Dept.
Faculty of Engineering
Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2000

Performance Analysis of Wireless Local Area Networks (WLANs)

by

Eng. Tamer Mohamed Samir Khattab

A thesis submitted to the
Faculty of Engineering at Cairo University
in partial fulfillment of the
requirement for the degree of
MASTER OF SCIENCE
in
ELECTRONICS AND COMMUNICATIONS ENGINEERING

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2000

Performance Analysis of Wireless Local Area Networks (WLANs)

by

Eng. Tamer Mohamed Samir Khattab

A thesis submitted to the
Faculty of Engineering at Cairo University
in partial fulfillment of the
requirement for the degree of
MASTER OF SCIENCE
in
ELECTRONICS AND COMMUNICATIONS ENGINEERING

Approved by the
Examining Committee

Prof. Dr. Mahmoud T. El-Hadidi	**(Thesis Main Advisor)**
Prof. Dr. Mohamed G. Darwish	**(Member)**
Prof. Dr. Mohamed Z. Abdel-Mageed	**(Member)**

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2000

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACK	Acknowledgement
AP	Access Point
BSA	Basic Service Area
BSS	Basic Service Set
BSSID	Basic Service Set ID
CCA	Clear Channel Assessment
CDF	Commulative Distribution Function
CDMA	Code Division Multiple Access
CEPT	Conférence Européenne des Postes et des Télécommunication
CF	Contention Free
CFP	Contention Free Period
CP	Contention Period
CRC	Cyclic Redundancy Check
CSMA/CA	Carrier Sense Multiple Access with Collision Detection
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CTS	Clear To Send
CW	Contention Window
DA	Destination Address
DBIR	Direct Beam Infrared
DBPSK	Differential Binary Phase Shift Keying
DCF	Distributed Coordination Function
DECT	Digital European Cordless Telecommunications
DFE	Decision Feedback Equalizer
DFIR	Diffused Infrared
DIFS	Distributed Coordination Function Inter Frame Space
DQPSK	Differential Quadrature Phase Shift Keying
DS	Distribution System
DSSS	Direct Sequence Spread Spectrum

EPA	Equilibrium Point Analysis
ESS	Extended Service Set
ETS	European Telecommunication Standard
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FDM	Frequency Division Multiplexing
FFT	Fast Fourier Transform
FHSS	Frequency Hoping Spread Spectrum
FSK	Frequency Shift Keying
GFSK	Gaussian Frequency Shift Keying
GMSK	Gaussian Minimum Shift Keying
HIPERLAN	High Performance Local Area Network
IBSS	Independent Basic Service Set
IEEE	Institute of Electrical and Electronic Engineers
IFS	Inter Frame Space
IR	Infrared
ISBSS	Infrastructure Basic Service Set
ISI	Inter Symbol Interference
ISM	Industrial Scientific Medical
ISO	International Standards Organization
LAN	Local Area Network
LBR	Low Bit Rate
LBT	Listen Before Talk
LD	Laser Diode sources
LED	Light Emitting Diode sources
LFSR	Linear Feedback Shift Register
LLC	Logical Link Control
LW	Light Wave
MAC	Medium Access Control

MPDU	Medium Access Control Protocol Data Unit
MSK	Minimum Shift Keying
NACK	Negative Acknowledgement
NAV	Network Allocation Vector
NB	Narrow-band
NPMA	Non Preemptive Multiple Access
OOK	On Off Keying
OSI	Open System Interconnect
PC	Point Coordinator
PCF	Point Coordination Function
PDF	Probability Density Function
PHY	Physical Layer
PIFS	Point Coordination Function Inter Frame Space
PLCP	Physical Layer Convergence Protocol
PLCP_PDU	Physical Layer Convergence Protocol – Protocol Data Unit
PLL	Phase Locked Loop
PN	Pseudo-random Noise
PPM	Pulse Position Modulation
PRMA	Packet Reservation Multiple Access
PSD	Power Spectral Density
PSK	Phase Shift Keying
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RA	Receiver Address
RF	Radio Frequency
RTS	Request To Send
SA	Source Address
SIFS	Short Inter Frame Space
SS	Spread Spectrum

SSS	Single Station Superposition
STA	Station
SWAP	Shared Wireless Access Protocol-Cordless Access
TA	Transmitter Address
TDM	Time Division Multiplexing
U-PCS	Unlicensed Personal Communication Systems
UV	Ultraviolet signals
VL	Visible Light signals
WEP	Wired Equivalent Privacy
WLAN	Wireless Local Area Network

ACKNOWLEDGMENT

I would like to express my sincere gratitude to Prof. M. T. El- Hadidi for being the guiding force behind the research necessary for the completion of this dissertation. Without his guidance, encouragement, support, ideas, and scientific enlightenment this work would not have been possible.

My great appreciation goes also to Dr. H. M. Mourad for her continuous support, and guidance for the work done in this thesis.

I would like to acknowledge the support of my colleagues at both the National Telecommunication Institute (NTI) and IBM Egypt for their help and support.

I am indebted to my father, my mother, my father-in-law, my mother-in-law, my wife, my son, and my brother and sister for all the encouragement and moral support they have provided me throughout my academic career. No words can express my gratitude to them. I therefore dedicated my work to them.

ABSTRACT

Attention is increasingly being directed to computer networks with mobility, such as wireless local area networks (WLANs). This is being motivated by the growing number of Notebook personal computers, as well as the need for fast installation and re-installation of local area networks in several applications (e.g. construction firms, retail activities, hospitals,… etc.) .

This thesis addresses the subject of WLAN and presents an overview of its current technology and the standards that have been proposed for its implementation (namely, IEEE 802.11 and HYPERLAN of ETSI). The thesis also analyzes the performance of a WLAN, which is based on the CSMA/CA access control method, using a closed queuing model approach. The solution utilizes a technique called Single Station Superposition (SSS) technique, which provides a set of four interrelated equations for the throughput, delay, probability of collision, and probability of having a busy medium. Numerical results are obtained using the Minerr() function of the MathCad package. The use of this technique allowed for analyzing CSMA/CA protocol including the effect of the important back-off algorithm. This feature was neglected in the analytical work for CSMA/CA found in the literature review. It was only considered in simulation work. This is because it complicates the analytical model and makes it mathematically intractable using classical techniques. Moreover, even in the simulation work on the subject no studies were made on the effect of changing back-off parameters on the system performance. It was found from our results that back-off parameters, such as maximum number of retrials on a collided packet, affect the system performance and may result in degradation of performance if not chosen correctly.

Another set of performance results is obtained for the case of a hybrid wireless/wired LAN network. For this part a protocol was proposed to interconnect a wireless network to a wired network. The protocol integrates a modified version of the packet reservation technique for the WLAN, and the CSMA/CD for the wired LAN. Both networks are interconnected via a

wireless access point. Expressions for the average throughput of the hybrid network and the average packet delay between two arbitrary stations are derived and solved numerically. Although, this type of hybrid networks is becoming essential, as it allows for resource sharing and data transfer between the new wireless networks and the already established wired networks, no analysis or simulation work was found in the literature review for it. The analysis we presented in this thesis serves two main purposes. One purpose is to demonstrate that this type of network interconnection can be analyzed using the mathematical models of the two network parts (wireless and wired parts). The other purpose is to show that the proposed protocol gives an acceptable performance that can be optimized through tuning of some network parameters according to the traffic characteristics.

Chapter 1

Overview of Wireless Local Area Network (WLAN)

Chapter 1
Overview of Wireless Local Area Network
(WLAN)

1.1.   The Need for Wireless Local Area Networks

In the last two decades the technology and the market of computing was moving from big centralized resources main frame computer with terminals into smaller more distributed resources personal computers. As a result of this movement resources are becoming isolated and not efficiently utilized. On the other hand, communication is becoming more and more essential for all business, scientific, and other tasks.

The need for resource sharing and communication capabilities was behind the evolution of computer networks. One important and commonly used type of computer networks is the Local Area Network (LAN). LANs are characterized by limited range, private usage, and high speeds.

In the last five years the mobility in computing and communications became more and more essential, especially for business usage. Mobility implies smaller sizes and more power considerations. It also, does not fit with fixed wired connections. The only suitable communication for mobile applications is wireless communication.

The implication of mobility on computers appeared in the evolution of Notebook personal computers. On the other hand, the effect of mobility on computer networks appeared in the evolution of Wireless Computer Networks of which Wireless Local Area Networks are one kind.

1.2.   Categories of WLANs

Wireless communication technology is becoming more and more advanced and used in varieties of applications in the last three decades. The main characteristic of this technology is that it uses air or free space for transmitting information form a source to a destination. The information is transferred in the form of an electromagnetic wave.

According to the frequency of this electromagnetic wave its characteristics (such as attenuation, ability to penetrate obstacles, refraction, …etc.) are defined.

Since WLANs uses wireless communication it is categorized according to the wireless electromagnetic wave used. This type of categorization is called Wireless Media Categorization.

There are mainly two main categories according to this feature; these are

1. Radio Frequency WLANs
2. Infra Red WLANs

The next chapter will describe both in details.

As WLAN is a Local Area Network, the categorizations that apply to other LANs apply to it. This means that WLAN is also categorized according to the network topology (such as Star, Ring, and Bus). It is also categorized according to the Medium Access Control (MAC) Protocol used (such as Contention Based MAC and Contention Free MAC).

One more characteristic that apply only to WLAN is the so called network architecture. Network architecture depends on the life time of the network as well as the path the signal takes from a source to a destination.

A description of all these categories as well as a comparison between them will be given in the next chapter.

1.3.   WLAN Standards

As any other commercial products, WLANs are produced by different vendors. To make sure that customer satisfaction is met, standards are needed.

Standards assure that certain services are provided with a certain level of quality as well as compatibility between different vendor products.

WLANs has two main Standard organizations that produced two sets of standards for WLANs. These organizations and their standards are:

- *Institute of Electrical and Electronic Engineers (IEEE),* which produces the 802.11 standards
- *European Telecommunications Standards Institute (ETSI),* which produces the High Performance LAN (HIPERLAN) standards.

Both standards coexist and will be described in details in Chapter 3.

It is worth noting to mention that standards only apply to physical and data link layers of ISO-OSI model. Other layers are left for application and usage to define.

1.4. Medium Access Control Protocols for WLANs

As mentioned before WLAN is like any other LAN categorized by the MAC protocol used for sharing the medium. One of the most used protocols is the so called Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). This protocol is adopted in many products in the market and is the one chosen for the IEEE 802.11 standards.

Due to its importance, this protocol is studied in details in Chapter 4. Although the protocol algorithm is not complicated, a mathematical model for the protocol is difficult to establish. This is why all the work done on the protocol till now either ignored an important feature of the protocol or used simulation to evaluate the protocol.

In Chapter 4 a recently proposed analysis technique is applied to the protocol to get an analytical model for it. The numerical results of the analysis are obtained and found to be close to those obtained via simulation. Analytical results are more needed than simulations, as they allow for more understanding of the behavior of the protocol.

1.5. Connecting Wireless LAN to Wired LAN

As wired LANs are already existing since they are older and because wired LANs allow for more resources because they generally have higher rates and are not limited in terms of size and power consumption as wireless LANs, there has been an increasing demand to connect wireless LANs to wired LANs. This type of connection produces a rather complicated situation to analyze mathematically. This is why this area was not yet investigated in terms of performance evaluation.

In Chapter 5 a proposed protocol which is a mixture of a wired protocol and wireless protocol is described with all the modifications done to the access technique. An approximate analysis technique that reuses the analysis done for each protocol independently is performed in order to get performance parameters for this type of networks.

1.6.   Thesis Objective

While simulation techniques can give an expectation of how communication protocols will behave and their expected performance to a good accuracy, analysis of communication protocols always gives better understanding of how the protocol works and performs under different conditions. It also helps a great much in enhancing the protocol and tuning its parameters for better performance.

In this thesis we aim at analyzing the performance of Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol. This protocol is one of the major MAC protocols used for wireless LANs. It is also adopted in the IEEE 802.11 wireless LAN standard. The performance of this protocol has been studied in some literatures like [CHHAYA96] and [WEINMILLER97], but in the first it was studied avoiding the random exponential back-off part of the system and only simulation results were given, which includes this important feature in the analysis. In the second reference, it was only a study through simulation and no analysis has been made. The reason why most of the literatures found in the subject avoid the random exponential back-off feature is that it produces a system model that is mathematically intractable [CHHAYA96]. We have used a technique that simplifies the mathematical model yet achieving good result accuracy to analyze the protocol including the effect of the back-off feature.

It was also noticed that the needs for interconnecting wireless LANs to an existing wired back-bone LAN is increasingly becoming essential in many areas. Hence, we suggested a protocol that integrates a wireless LAN protocol and a wired LAN protocol together to form a hybrid wireless/wired LAN. Performance of this protocol was studied to achieve two goals:

- Analyzing the proposed protocol to verify its applicability and performance enhancements possibility.
- Giving a first bitch on how interconnected wireless/wired LANs can be modeled to achieve performance analysis of such LAN types.

1.7.   Thesis Outline

In order to achieve the objectives mentioned in the previous section the thesis will go as follows:

In Chapter 2 we will start by giving the reader an overview of the technologies involved in wireless LANs. In this overview we give detailed descriptions of the available physical media used for wireless LAN communications. We then move into the protocols used for wireless LANs. In each part of these we tried to explain the challenges facing WLANs technology and give a comparison between the different alternative solutions.

Knowing all the different possible solutions used for WLANs the reader might wonder about interoperability between products from different WLAN vendors. This is why in Chapter 3 we move into giving an overview of the ongoing standardization effort for WLANs. In this chapter we give a detailed description of the two main wireless LAN standards; namely the IEEE 802.11 and the ETSI HIPERLAN standards.

In Chapter 4 analysis of the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol is obtained using Markov queuing model of the protocol. The mathematical representation of the average system throughput and delay are obtained from the Markov model using a technique called Single Station Superposition. Equations for throughput and delay are solved numerically to get performance analysis results.

In Chapter 5 interconnection between wireless LANs and wired LANs is studied. In this study we propose a protocol for this interconnection. The protocol proposed integrates two different known protocols. One of them is used for wireless LANs and called Packet Reservation Multiple Access (PRMA), the other is used for wired LANs and called Carrier Sense Multiple Access with Collision Detection (CSMA/CD). We then analyze the hybrid protocol to know if it gives an acceptable performance and if it can be enhanced. The analysis uses the already established models for the used protocols and adds to them the necessary modifications to accommodate the hybrid situation. The analysis is also given for the purpose of taking a first step in analyzing this type of hybrid networks, as there were no literature found in the subject.

In Chapter 6 we conclude our work by stating what we have achieved in the thesis and propose future areas of research related to the subject of the thesis.

Chapter 2

Wireless LAN Technology Overview

Chapter 2
Wireless LAN Technology Overview

2.1.   <u>Introduction</u>

Wireless LANs - like any other Local Area Network - use a variety of communication technologies to achieve data connection between network users. These technologies are mainly characterized by the variations in the first two layers of the ISO-OSI model, namely the *Physical* and *Data Link* layers.

The Physical layer technologies can be divided into groups according to one of the following criteria:

? Network architecture.
? Network topology
? Wireless Medium (electromagnetic wave) type.

On the other hand, the Data Link layer technologies are categorized according to only one criterion, which is the Medium Access Control (MAC) technique used in the network.

In this chapter we will give an overview of the alternative technologies used in both layers. We will start by the physical layer, which focuses on the network architecture, the implementation topology and the electromagnetic wave used (transmission medium).

The description of the network architecture part will consider both infrastructure architectures and add-hoc architectures from the point of view of connectivity lifetime. On the other hand, it will describe both point-to-point and diffused broadcast signal transmissions.

In a section that will follow different implementation topologies such as bus topology, star topology and ring topology will be described. The description will focus on the suitability of this topology to wireless transmission in general, and to different electromagnetic waves and different network architectures.

To complete the physical layer description, the different transmission electromagnetic wave types are described. These types are mainly Radio Frequency (RF) transmissions and Light Wave (LW) transmissions.   Radio frequency part includes both narrow-band (NB) transmission and spread spectrum (SS) transmission (both direct sequence DSSS and

frequency hopping FHSS). On the other hand, light wave transmissions are limited only to infrared (IR) transmissions.

Having described the different physical layer alternatives and characteristics, we move into the data link layer. In this layer we focus on the different medium access control (MAC) techniques used in wireless LANs. We will first start by describing the two main categories of MAC; namely contention-based and contention-free MAC. Then we will give a description of the different MAC techniques used within each category.

Finally, at the end of each section we will give some tables that compare between the different technologies and techniques described in this section. These tables as well as a section for conclusions are a briefing of the main results we obtained from the literature overview given in the chapter.

2.2. Physical Layer Technologies

The physical layer is categorized according to the architecture, the topology and the type of the electromagnetic wave used. The categories of these three criteria will be described in the following subsections as well as the advantages and disadvantages of each category.

In general, wireless communication uses electromagnetic waves in free space or air as a transmission medium to send information from its source to its destination. The types of network architectures used in wireless LANs will be described first. After that the different wireless topologies will be explained, and finally the types and characteristics of the electromagnetic waves used will be given. This ordering of description is selected because, the architecture of the network is mainly selected to suite the needed application which is the main user requirement. According to the suitable architecture and the working environment some limitations are put on the network topology. After selecting the suitable topology, the wireless LAN designer must select the suitable wireless transmission medium that can work for the selected topology and environment with best performance.

2.2.1. Wireless LAN architectures

Wireless LANs use different network architectures to achieve connectivity between a number of nodes in a certain topology. The term architecture here refers to two main criteria. These criteria are [STALLING97A]:

7

? The network connectivity lifetime and planning.

? The way the signal is transmitted from one node to another.

Network connectivity could be with a previously planned topology and permanent connectivity. This means that the network is designed and then established to last for a long time with the same configuration as in the case of office or factory permanent wireless nodes connection and the case of connecting wireless nodes to a backbone wired LAN. On the other hand, network connectivity could be with no pre-specified topology and with a temporary connection, as in the case of networks established in meeting rooms between a group of portable computers to help people transfer some files during the meeting. Another example is the military tactical situations where networks are established in temporary strategic places for short periods of time. These two types of wireless LANs are divided into two architectural categories, which are respectively, called [STALLING97A]:

? Infrastructure WLANs (Fixed wire replacement or extension).

? Add-hoc WLANs.

With respect to the way the signal is transmitted from one node to another, there are two main types of wireless connections used for WLANs; namely *Point-to-Point* and *Diffused Broadcast*. These types are categorized according to the path the signal takes from the transmitter to the receiver. Description of these two types will follow with an emphasis on the suitability of each of them to the wireless architecture described before.

2.2.1.1. Point-to-point wireless connection

Point-to-point connections are mainly dependent on making the signal of the transmitter to be directed and intended to a single receiver (Fig 2-1). This implies that the signal of the transmitter is required to be highly directive, and the transmitter-receiver pair to be well aligned towards each other. This criteria results in reducing the needs for high signal power from the transmitter [FERNANDES94].
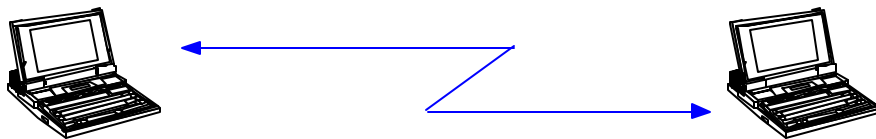
Fig 2-1 Point-to-point wireless connection

8

It can be seen from the above description that this connection has the following characteristics:

? Point-to-point connection requires a highly directed signal, which make it more suitable to IR transmissions than to RF transmissions. This is because it is very difficult to generate a highly directive RF signals, while it is the nature of light waves to be directive.

? This type of connection is more suitable to infrastructure networks as it requires alignment of the transmitter and receiver, which is not suitable for temporary ad-hoc networks.

? As transmitter-receiver alignment is required for this type of connection, it is obvious that mobility is very limited and usually fixed terminals are more suitable.

? Because this type of connection is used for fixed node networks, and as the signal level at the receiver is very high because of the highly directive transmission signal, this type of connection have lower bit error rates. Hence, the transmission over this type of connection is quite reliable.

? As this connection is directive and uses line-of-sight transmission, it is less susceptible to interference from other sources. It is, however, more susceptible to shadowing effect, which is the phenomena caused by moving objects, such as people inside an office, which intercepts the line of sight path between the transmitter and the receiver instantaneously causing a loss of connection for a short duration.

? Regarding security issues, point-to-point connection has the advantage of being interrupted if intercepted by a hostile receiver. This interruption could be detected by the communicating nodes, and hence any eavesdropping could be detected easily.

? Since point-to-point connections are more reliable and signal levels at the receiver are higher, they have larger communication range, which totally depends on the transmitter signal level. Ranges up to few kilometers are achievable [STALLING97B]. It must be noticed that the coverage range in case of point-to-point connection is measured in lengths on the line between the transmitting node and the receiving node, and not in areas as in the case of diffused connections. This coverage length is given the notation d, where d is the transmitter radio range.

2.2.1.2. Diffused broadcast wireless connection

As the name implies, this connection is by nature a broadcast connection. In this type of connection the transmitting node signal is propagated in a way that makes it detectable by all the other nodes in the network, whether they are the intended receivers or not (Fig 2-2). Nodes listening to the broadcast signal can detect if they are intended receivers or not via the destination address imbedded in the transmitted frame. The transmitted signal in this case

should not be directive at all. On the contrary they should be diffused with wide beam range. This means that no alignment is needed between the transmitter and the receiver. It also means that higher transmitted signal power is needed, as the density of the signal power at the receiver is very low [STALLING97A].
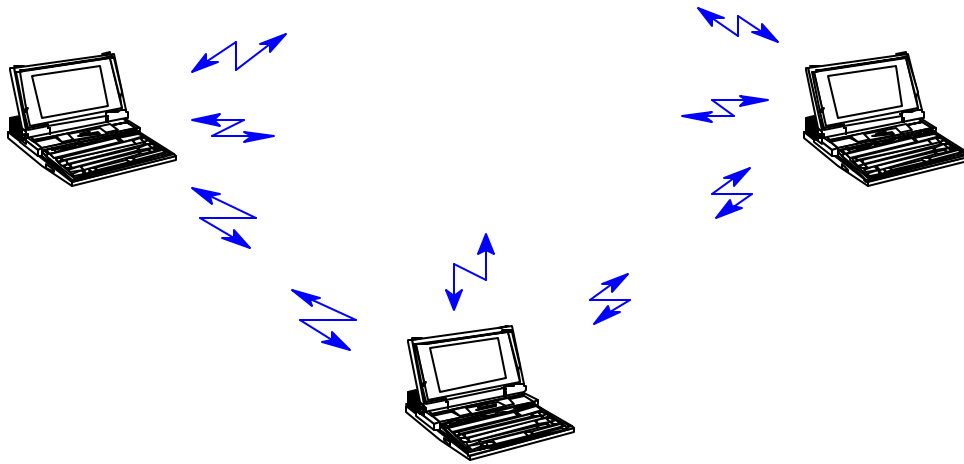
Fig 2-2 Diffused broadcast wireless connection

A careful look at the above description of diffused broadcast connection type, will result in extracting the following characteristics for this type of wireless connection:

? Diffused broadcast connection is basically dependent on non-directive signal communication. This means that it is more suitable for RF signals because they are naturally non-directive signals. But IR signals could be used with the help of central diffusion devices such as ceiling passive reflector or active satellite or by using the beam diffusion from room walls (Fig 2-3).

? Diffused broadcast connections can be used for both infrastructure and ad-hoc networks because they need no pre-specified network architecture.

? This type of connection allows nodes to be mobile with the limitation that they do not get out of the radio contact range. Mobility is more allowable with RF signals because they need no diffusion device and they can penetrate walls. On the other hand, lower level of mobility is achieved when IR is used because they do not penetrate walls and their transmitters and receivers should be aligned towards the diffusive device, in case one is used, and they have limited coverage ranges.

? Unguided broadcast transmissions suffer from lower signal levels at the receiver, and also multi-path fading effects specially when used indoors. These facts makes diffused broadcast communications less reliable as they suffer from higher bit error rates.

? Because the signal can reach the receiver from different paths, diffused broadcast connections are not sensitive to obstacles and moving objects, especially when RF signals are used.

? By its nature diffused broadcast connections are publicly available signal types, which means any receiver that is not from the network can listen to the transmitted signal, specially with RF. This means that the security levels are very low and the ability to detect eavesdropping is very weak. For all these reasons, encryption must be used for network security, and again authentication must be used for session's security.

? As it can be seen from the above, diffused broadcast connections are less reliable and have higher bit error rates. They also have lower signal levels at the receiver side because signal diffusion reduces its concentration levels on a specified area. These facts lead to smaller coverage range for diffused connection. For WLANs the range is around 70 m [STALLING97B]. The range is still dependent on the transmitter power, but in this case it is measured in areas, not lengths. The coverage area of a certain network is measured by the maximum circular area, which allows for a connection between any two nodes in the network moving within this area.

2.2.2.   Wireless LAN topologies

Theoretically speaking, all types of network topologies used for wired LANs can be used for wireless LANs, but practically this is not true. The fact that wireless communication channels have different characteristics than wired channels, and the demand for mobility and add-hoc connectivity in WLANs are the reasons why this is not true [DAVIS95].
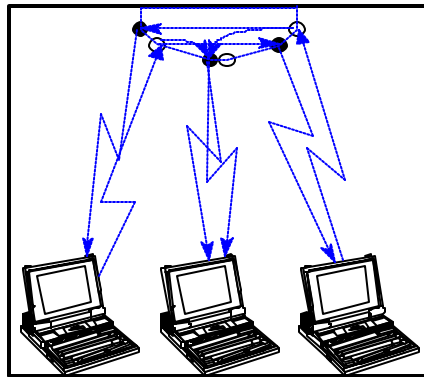
Active ceiling satellite   Passive ceiling reflector
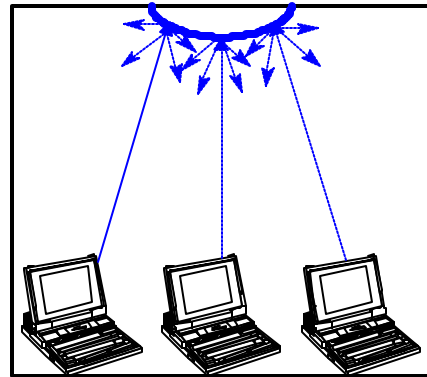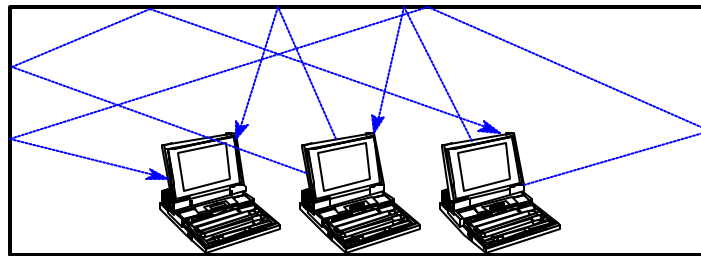
Reflection from walls

Fig 2-3 IR diffusion techniques

The main characteristics of wireless channels, regardless of the type of the electromagnetic wave used, that results in having restrictions when selecting the topology are stated in the following lines [BAUCHOT95].

? Wireless communication channels are more noisy and less reliable than wired channels.

? Wireless communication channels are naturally public broadcast channels.

? Wireless devices are usually battery operated devices, as they are portable, which requires a very strict power consideration in such devices.

? The wireless communication range is limited either due to regulations or laws of physics.

? Wireless nodes are required to have mobility and support ad-hoc networking.

Due to the above characteristics of wireless communication channels, and as the transmission medium is not a guided medium, we cannot use the same topologies used in wired LAN

physically. This is because all topologies of WLAN have the common feature of using a shared medium. Wired LAN topologies could be achieved on wireless LANs media logically rather than physically except for certain cases.

In the following a description of some wireless LAN topologies will be given with an emphasis on the suitable architecture and limitations of each.

2.2.2.1. Ring topology

Ring topology can be achieved in wireless LANs using point-to-point connections [STALLING97B]. This topology is used for infrastructure LANs as it requires fixed nodes and pre-specified network planning.

This topology is not commonly used in wireless LANs as it restricts mobility and in general produces lower reliability for the network. It is also not easy to insert or remove nodes from the ring without service interruption.

The main advantage of this topology is coverage area, which is the circle with circumference given by $d \times n$, where d is the radio coverage distance of any node and n is the number of nodes (Fig 2-4).
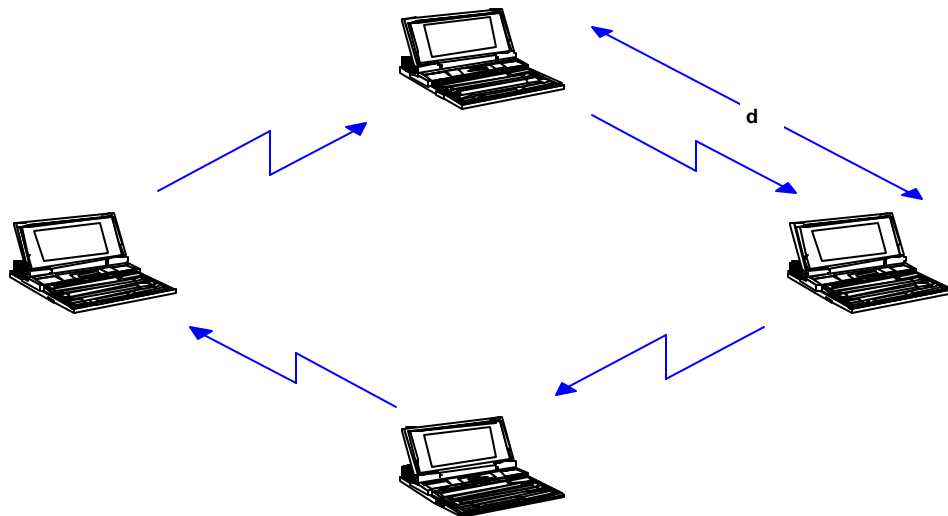
Fig 2-4 Ring topology using point to point connection

2.2.2.2. Bus Topology

Bus topology is achieved in wireless LANs using diffused broadcast connection [DAVIS95]. The bus is the free space or the air itself, and ad-hoc or infrastructure networks could be achieved using this topology. The coverage is the circular area with diameter d, where d is the transmitter radio range (Fig 2-5).

Bus topology is one of the most commonly used topologies in wireless LANs because of its flexibility in inserting new nodes without network interruption. It also allows for mobility and its reliability is high because of its non-single-point-of-failure nature.

2.2.2.3. Star topology

Star topology can be achieved using both point-to-point and diffused broadcast connection. The only limitation in architecture for star topology is that there must be a predetermined hub. Hence, ad-hoc network is not suitable. Consequently, infrastructure networks are more suitable. Here in the case of star topology the coverage area is the circular area with radius d (Fig 2-6). This means that star topology has larger coverage area than bus topology and they are also more suitable for interconnecting wireless LAN to wired LAN because in this case a central point must exist for transferring wireless traffic to wired traffic and vice versa.

Star topology on the other hand is less reliable than bus topology because of its single-point-of-failure nature.
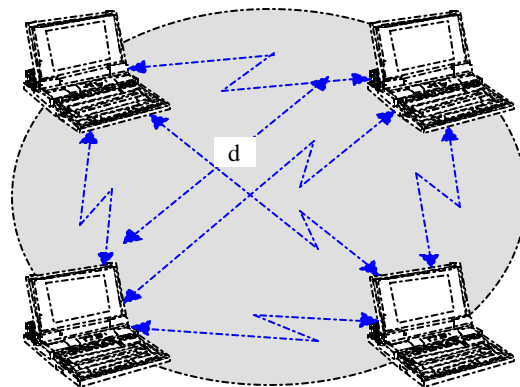
Fig 2-5 Bus topology using diffuse broadcast connection
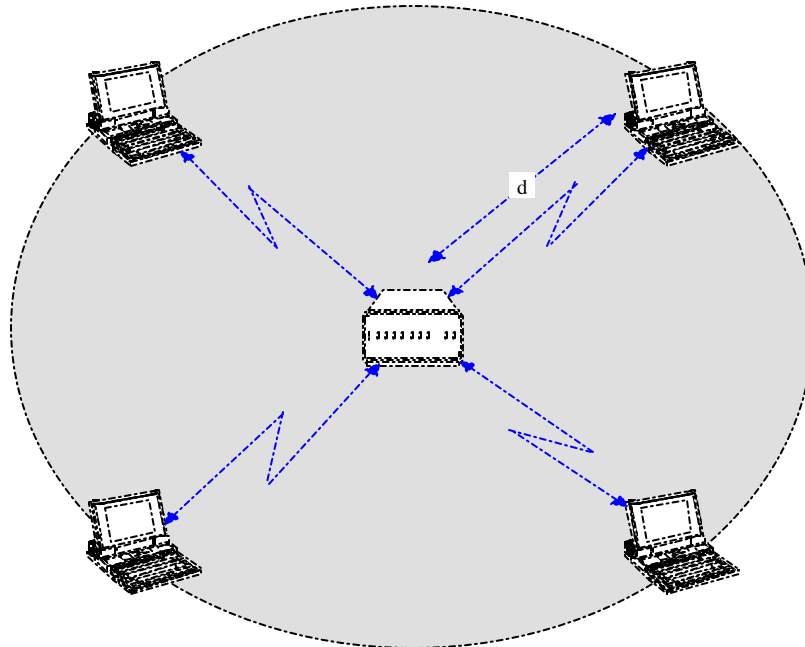
Fig 2-6 Star topology using diffuse broadcast connection

2.2.3. Wireless LAN transmission medium

There are two types of wireless local area network transmission media, which are classified according to the type of electromagnetic wave used i.e. the frequency of the carrier (Fig 2-7). The first type is Radio Frequency Wireless LANs (RF WLAN), while the other is Light Waves Wireless LANs (LW WLAN) which is only limited to Infrared WLAN (IR WLAN). In the following we are going to explain these two types of wireless transmission media in a somewhat detailed way to express the major differences between them, the advantages and disadvantages of each, and the suitable working environment for each.

This description will concentrate on the communication channel aspects such as frequency ranges, bandwidths, antennas used, data rates obtainable, and transmission schemes, as well as operating environment restrictions such as susceptibility to interference and obstacles, distances covered, and power levels achievable and allowed.
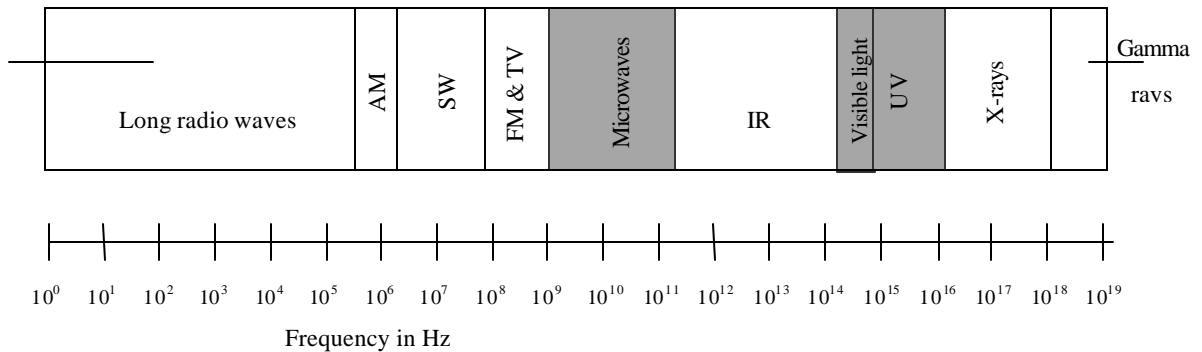
Fig 2-7 The electromagnetic spectrum

2.2.3.1.   Radio Frequency wireless LAN (RF WLAN)

Radio frequency wireless communications always face the problem of the very crowded frequency bands. Hence, the need to license the operating frequency is very much essential for reliable and stable operation. A great effort has been spent by the Wireless LAN vendors to achieve licensed frequency bands for their products [DAVIS95].

Radio frequency wireless communication systems are mainly divided into two categories. These categories are according to the bandwidth used and the frequency of the transmission. These categories are [STALLING97B]:

?   Narrow-band transmissions.
?   Spread spectrum wide-band transmissions.

Radio frequency electromagnetic waves have wide variations in their characteristics according to the frequency range used. These variations result in the following general frequency dependent characteristics: [BANTZ94]

1. The lower the frequency band the less complex and less expensive the devices used in transmitter and receiver circuitry. This is especially true for the ranges below 3 GHz, because in this range the silicon bipolar transistor can be used, and this electronics technology is the cheapest.

2. The lower the frequency the higher the penetration capabilities of the radio wave. This is especially true for frequencies lower than 5 GHz. On the other hand, although radio

frequency waves at 5 GHz bands are more susceptible to obstacles; they still can penetrate few walls. This means that they still have better penetration capabilities than light waves.

3. The higher the frequency used for radio frequency transmission the shorter the wavelength. This makes it possible to use small size antennas for higher frequencies.

4. The higher the frequency the lower the occupancy of the band. This is because lower frequencies are more capable of penetration and less power hungry, so they are always occupied first before going to higher frequencies. Also, higher frequency carriers can carry more information bandwidth than lower frequency ones. Due to these facts, operation on higher frequencies has the advantage of larger bandwidths available and assigned for them.

5. Higher frequencies are more directive and less capable of penetration through objects. This property makes them immune against interference from other devices operating on the same frequency band.

Narrow-band transmission systems are the traditional radio frequency transmission technique. The system transmits and receives data on a specific radio frequency band. The techniques used for narrow-band transmissions rely on keeping the transmitted signal bandwidth as low as possible. To avoid interference between different data sources, multiplexing techniques such as Frequency Division Multiplexing (FDM) and Time Division Multiplexing (TDM) are used. Both of these techniques rely on dividing the channel into sub-channels either in the frequency or the time domains. Even if TDM is used, the frequency on which the resulting multiplexed signal will be transmitted is still in need to be guaranteed for this signal so as not to suffer from interference with other sources.

From the above discussion, it is obvious that strict regulations must be applied in the frequency domain allocations to achieve reliable communication in case of narrow-band transmissions. These regulations are the main problem that faces any new user to the RF band. This is because the RF band is already overcrowded, and the traditional user of any sub-band is given the priority over any new user of this band, to achieve regulation stability and products maintainability. The lower the frequency of the required sub-channel, the more crowded the sub-band is. This is because, the lower frequency means cheaper equipment and more penetration ability.

One of the most famous radio frequency allocation management organizations is the Federal Communications Commission (FCC). This organization is responsible for regulating the RF spectrum between different users and devices in North America. There are similar

organizations in Europe such as Digital European Cordless Telecommunications (DECT), and Conférence Européenne des Postes et des Télécommunication (CEPT).

As mentioned before, narrow-band transmission needs to license the radio frequency band it is working on. The current regulations of frequency allocation assigns the following frequency bands for narrow-band Wireless LAN transmissions:

? Microwave band of 18-19 GHz, with low power (a fraction of a watt) [O'DONNELL94].

? Also worth noting is that the FCC regulations has opened the frequency band of 1890 MHz - 1930 MHz for license-free narrow-band wireless transmission [BANTZ94]. This band has been known as the Unlicensed Personal Communication Systems (U-PCS). In June 1994 the FCC reduced this band to a 20 MHz bandwidth from 1910 MHz to 1930 MHz, segmented into a 1910-1920 MHz sub-band for asynchronous applications such as WLANs and a 1920-1930 MHz sub-band for isochronous applications such as cordless phones. The 1910-1920 MHz sub-band requires the addition of a Listen before Talk etiquette that allows different devices to operate on it without interfering with each other. This band is also shared with point to point microwave links, but as such links are highly directive, they usually don't cause problems [LAMAIRE96]. This band should be used with very low power radiation levels, typically 100 mw [BANTZ94].

? Another regulation is the CEPT/DECT regulations, which assigns the 1880-1900 MHz [DAVIS95] band for Wireless LAN to be used also under low power conditions (25 mw) [PAHLAVAN95].

? Finally, the Industrial Scientific, and Medical (ISM) band which is licensed by both CEPT and FCC can be used for narrow-band transmission under the restriction of very low power levels (25 mw) [PAHLAVAN95]. This band contains three frequency ranges: 902-928 MHz, 2400-2483.5 MHz, and 5725-5850 MHz.

The above mentioned bands, and the general characteristics of narrow-band RF transmission result in the following characteristics for narrow-band WLANs:

? Narrow-band transmissions uses high frequencies which implies in addition to the previously mentioned frequency dependent general characteristics the following specific characteristics:

a) In the above mentioned bands the ranges around 1900 MHz and ISM band are capable of penetrating few walls, but in the case of the 18-19 GHz band this ability is very much reduced and line of sight is required.

b) As higher frequencies require line of sight communications, they require more directive antennas. This means that for the case of 18-19 GHz band directive antennas are required

and hence point-to-point connection between nodes is the only obtainable one. On the other hand, in the case of the band around 1900 MHz and ISM band the antennas could be omni-directional and hence broadcast connection between nodes could be achieved.

c) Devices operating on the band around 1900 MHz and ISM band are more susceptible to interference from other devices than those operating at the 18-19 GHz band. This is because electromagnetic waves at 1900 MHz band are more capable of penetration through obstacles and use omni-directional antennas.

d) Higher carrier frequencies are allocated higher bandwidths. This can be seen from above, as the band of 18-19 GHz has a bandwidth of 1000 MHz, while the other bands; namely the 1910-1920 MHz and the 1880-1900 MHz have bandwidths of 10 MHz and 20 MHz respectively. Also, the ISM three frequency bands 902-928 MHz, 2400-2483.5 MHz, and 5725-5850 MHz have bandwidths of 26 MHz, 83.5 MHz, and 125 MHz respectively, which is increasing as the band starting frequency increases.

e) Higher frequency electromagnetic waves suffer from higher attenuation in the free air transmission. This fact makes them more power hungry, and as they are directive their power could be increased to achieve higher signal to noise ratios at the receiver but the regulation limits this power to a fraction of a watt only [O'DONNELL94] to reduce interference to the lowest possible values. On the other hand, as lower frequencies are capable of penetration through obstacles, their powers are limited to very low values (as low as 25 mw) to avoid interference [PAHLAVAN95].

f) Although higher frequencies are susceptible to obstacles, their coverage ranges can be increased. Because of the directive antennas used for them, they are more suitable to fixed outdoors-wireless links as linking two LAN parts between buildings. This is the main use of the 18-19 GHz band, which are capable of reaching distances up to 20 km [O'DONNELL94] in outdoors connections. On the other side lower frequency bands are limited to a power level of 25 mw with non-directive antennas. This reduces their coverage to a range of 15-45 m [PAHLAVAN95].

? Data rates achievable using narrow-band transmission techniques are dependent on the frequency ranges, the bandwidth allowed, and the modulation techniques used. As the frequency ranges are either in the UHF band or in the microwave band, then they allow for very high data rates, but the allowable bandwidths are very limited because of the RF crowded frequency. This limited bandwidth puts a very strict limitation on the maximum achievable data rates. On the other hand, one of the most limiting factors for achievable data rates in RF transmissions is the electronics technology limitation. This is because the higher the data rate is, the more sophisticated the circuitry of the transmitter and receiver is. A final thing to be said about this issue is that the current standards and devices available

for narrow-band transmission Wireless LANs can achieve data rates ranging from 5 Mb/s up to 20 Mb/s [PAHLAVAN95] and [LAMAIRE96]. The future is aiming at achieving data rate ranges reaching up to 500-1000 Mb/s especially in the 18-19 GHz band without compression, and in the lower bands using *M-arry* binary modulation techniques to reduce the data rates to a suitable range for the allowable bandwidth.

? Transmission schemes used for narrow-band transmissions are mostly QPSK or FSK binary modulation schemes. These schemes are applied either using one carrier at the center of the bandwidth used and with the addition of equalization to reduce Inter Symbol Interference (ISI), or using multi-sub-carrier systems to reduce ISI associated with Fast Fourier Transform (FFT) as combining algorithm [STALLING97A].

? Equalization is needed to remove Inter Symbol Interference due to the effects of multi-path dispersion.

? Narrow-band transmission contains no implicit encryption techniques, and because radio transmission in general uses an open medium (free air), the security of such transmission techniques is very low, and they are very much susceptible to eavesdropping and spying. To improve the security of narrow-band signal transmissions over the channel, an encryption algorithm and an associated decryption algorithm must be applied at the transmitter and the receiver, respectively. On the other hand, for the level of session's security, an authentication technique must be applied.

This concludes the brief description of the properties and characteristics of narrow-band RF wireless LANs. A brief description of the second RF transmission technique will be given in the following lines.

Spread spectrum transmission systems were used initially by the military in World War II because of its high reliability, imbedded security, and anti-jamming capabilities. The Spread Spectrum (SS) communication techniques used in WLANs are mainly divided according to the signal-spreading scheme into two categories [PROAKIS95], [VITERBI95]:

? Direct Sequence Spread Spectrum (DSSS)

? Frequency Hopping Spread Spectrum (FHSS)

In both categories the main idea is to spread the Power Spectral Density (PSD) curve of the transmitted signal much more beyond its bandwidth over the entire allowable channel bandwidth. This causes the power level of the signal to be reduced with the same ratio of the increase in the bandwidth, as the total area under the PSD is constant for the same signal, and equals the power of the transmitted signal (Fig 2-8).

In both techniques the spreading of the signal is performed using a pseudo-random code called Pseudo-random Noise PN. This randomness in the spreading code is the reason of the implicit security in Spread Spectrum techniques. Also the fact that the spread signal has a low level of power (in most cases comparable to the channel noise power), makes it difficult to detect the presence of these signal, as they will be hidden inside the background channel noise.

Power

Spread Spectrum signal PSD

Normal signal PSD

Frequency
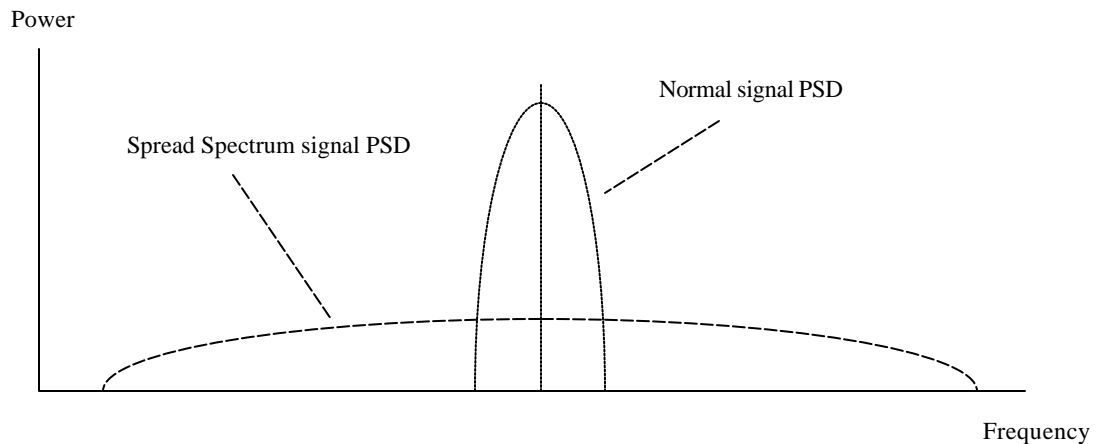
Fig 2-8 The effect of spreading the signal

Direct Sequence Spread Spectrum uses a PN code, which has a bit rate very much higher than the bit rate of the transmitted signal. This PN code is binary multiplied (XORed) with the transmitted signal before modulation. This causes the resulting signal to have a bandwidth approximately equal to that of the PN code. Hence the transmitted signal is effectively spread over the bandwidth of the PN code. Taking the bandwidth of the PN code to be equal to the bandwidth of the allowable channel, the transmitted signal will be spread over the bandwidth of the allowable channel (Fig 2-9). On the other hand, at the receiver the spreading process is reversed by XORing the received spread signal, after demodulation, with the same PN code [PROAKIS95].
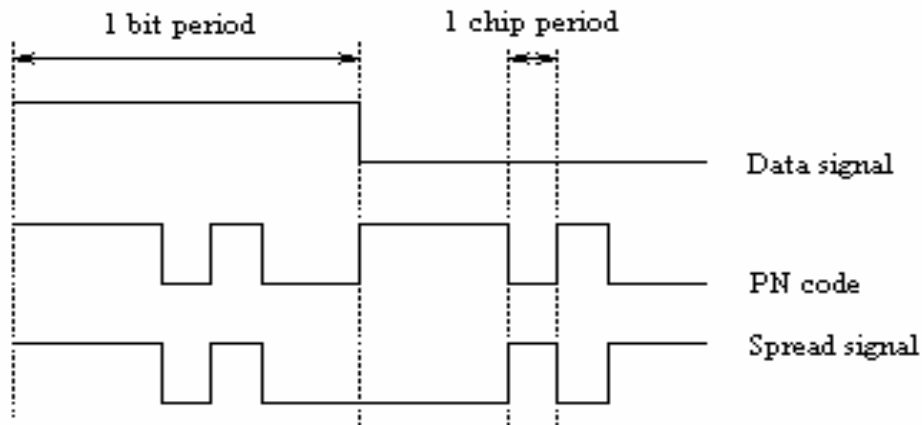
Fig 2-9 Direct Sequence Spread Spectrum

Frequency Hopping Spread Spectrum on the other hand, uses the PN code to change the frequency of the carrier between different frequencies covering the range of the allowable bandwidth, or most of it. This means that the PN code controls the frequency of the modulator causing it to hop pseudo-randomly between different frequencies in the allowable bandwidth. Hence, the spreading factor here is dependent on the selection of these frequencies (Fig 2-10).

As in the case of DSSS, here also the process is reversed at the receiver. Thus, the receiver changes its center frequency to follow exactly the same pattern of the transmitter's carrier frequency hops [PROAKIS95].

As it can be seen form the previous description, Spread Spectrum techniques in general contribute low power spectral density levels of transmitted (spread) signal to the channel, and they have high immunity against interference from narrow-band transmission sources. Accordingly, no strict frequency management regulations are necessary for them as they produce very low level of interference to normal narrow-band transmissions occupying the same band. Hence, the FCC and CEPT has allowed Wireless LANs using Spread Spectrum transmissions to share the ISM band with other primary narrow-band transmission devices operating in this band with power levels as high as up to 1 watt. This higher power value allows for higher coverage range of Spread Spectrum WLANs.
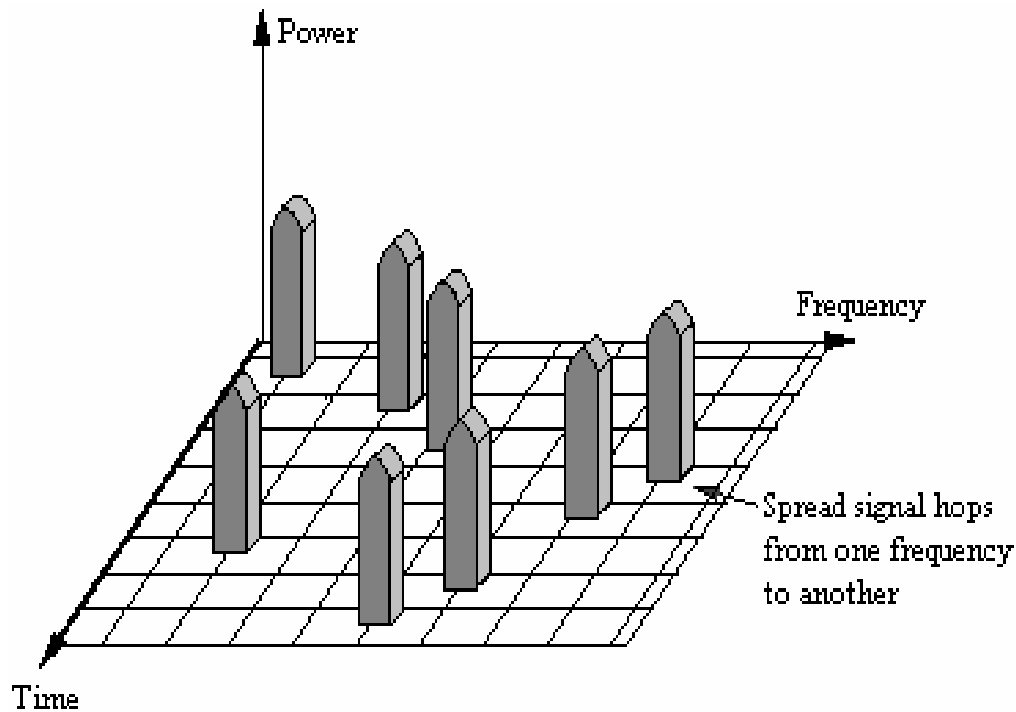
Fig 2-10 Frequency Hopping Spread Spectrum

The ISM band, as mentioned previously, occupies three different frequency bands. These bands are already assigned to narrow-band priority users. The bands as well as their priority users are as follows [PAHLAVAN95], [BANTZ94] and [DAVIS95]:

? 902-928 MHz, used by *baby monitors*, *amateur radio*, and *cordless phone sets*.

? 2400-2483.5 MHz, used by *microwave ovens*.

? 5725-5850 MHz, used by *radar and navigation devices*.

These bands have certain regulation when used for DSSS and FHSS. These regulations are made to make Spread Spectrum signals in this band immune against interference from the narrow-band priority users of these bands. They also ensure that the Spread Spectrum signal will not introduce to the channel more than the allowable interference noise so that priority users can operate reliably and safely. These regulations are as follows:

? For DSSS the bandwidth expansion ration must not be less than a factor of 10. This means that the bandwidth of the transmitted signal must not exceed 1/10 of the allowable channel bandwidth [BANTZ94].

? For FHSS in the 902-928 MHz band, the channel is divided into sub-channels (hopping bands) of 0.5 MHz bandwidth. This gives us 52 hopping frequency bands, of which 50 bands should be used by the hopping pattern, and the other 2 are left for interference avoidance. On the other hand, both the 2400-2483.5 MHz and 5725-5850 MHz bands are divided into sub-channels of 1 MHz bandwidth. This gives us 83 and 125 hopping frequency bands respectively, of which 75 bands should be used, in each case, by the hopping pattern. The other 8 bands, and 50 bands, respectively, are left to be used for interference avoidance [BANTZ94].

The above mentioned bands and the regulations for their usage, as well as the general characteristics of Spread Spectrum communication techniques result in the following characteristics for wireless spread spectrum RF LANs:

? Spread spectrum technique uses lower frequencies than narrow-band techniques which implies in addition to the general frequency dependent characteristics the following specific characteristics:

1. Although the RF signals in ISM bands are capable of penetration through obstacles to a somewhat high extent, especially in lower bands, the nature of spread spectrum techniques makes RF communication at these bands very highly resistant to interference. This resistance against interference is achieved due to the PN coding used in the spreading and de-spreading processes. This coding effect is the basic of the so called Code Division Multiple Access CDMA. Also Frequency Hopping Spread Spectrum techniques when used with a coding scheme called Repetition Code can have multi-path resolving capabilities, which gives them resistance against Fading effects, thus eliminating the need for equalization.

2. Spread Spectrum techniques already reduces the power of the signal at a certain band, by spreading it over a much more wider band. Hence, more power levels can be associated to signals using this technique without being afraid of affecting other narrow-band or different PN code spread spectrum signals occupying the same band. The FCC regulations allow for a power level up to 1 watt for spread spectrum operation in the ISM band [PAHLAVAN95]. On the other hand, in Europe the Digital European Cordless Telecommunications (DECT) standards limits the power to 250 mw, while the European Telecommunication Standard (ETS 300 328) limits it to 100 mw [RUNE95].

3. Because more power levels are allowed for Spread Spectrum communications, and also because of the relatively higher penetration ability of RF signals in the ISM band frequency ranges, Wireless LANs operating using these techniques have a larger coverage ranges. Of course, these ranges are very much dynamic according the operating environment, such as

whether the communication takes place in an indoor or an outdoor environment, and the number of walls or other obstacles the RF signal has to go through from the transmitter to the receiver. On the average, for DSSS the coverage ranges are from 30 m to 270 m, and for FHSS the ranges are from 30 m to 100 m [PAHLAVAN95].

? Besides the limitations on data rates that comes from the bandwidth limits of the ISM bands, there is another more strict limitation coming from the Spread Spectrum regulations for using the ISM band. In fact, these are the actual factors that determine the transmission rates achievable in Spread Spectrum WLANs. For DSSS, the regulations require that the spreading factor to be equal to 10 or more. Combining this with the bandwidths available, we find that the maximum data rates achievable with DSSS are 2 Mb/s, 8 Mb/s, and 10 Mb/s for the 902-928 MHz, 2400-2483.5 MHz, and 5725-5850 MHz bands, respectively. To increase the data rates multi-amplitude and multi-phase techniques are used. Also, some vendors use multiple codes to transmit the data over one channel between a certain transmitter and receiver pair with higher rates. Practically, data rates ranges of 2-20 Mb/s are achievable [PAHLAVAN95]. On the other hand, for FHSS, the regulation limits the sub-channels bandwidth to 0.5 MHz in the 902-928 MHz band, and to 1 MHz in the 2400-2483.5 MHz, and 5725-5850 MHz bands. This means that the data rates achievable are 500 kb/s for the 902-928 MHz, and 1 Mb/s for the bands 2400-2483.5 MHz, and 5725-5850 MHz. Here also, multi-amplitude and multi-frequency techniques could be used to increase data rates. In fact, data rates in the ranges of 1-3 Mb/s are achieved with FHSS [PAHLAVAN95].

? For DSSS, the used transmission scheme is QPSK, which allows increasing the bit rates up to 20 Mb/s. Also, as PSK has higher signal to noise ratio, the coverage range of DSSS is higher than that of FHSS. When FHSS is considered, the transmission scheme of choice is GFSK (Gaussian Frequency Shift Keying). Again as this scheme is used with a multi-frequency modulation, more data rates are achievable. It is worth noting here, that the FSK has lower signal to noise ratio. Also, as regulations prohibit the use of synchronized FHSS, the only used scheme is non-coherent (non-synchronized) FHSS. The non-coherent FHSS suffers from what is called recombination losses at the receiver. Both these facts result in a coverage range for FHSS communications that is less than that of DSSS [STALLING97A].

? Due to the PN coding used for spreading the signal in Spread Spectrum techniques, they have a built in security scheme. This achieves the signal security against eavesdropping over the transmission channel. The other level of security is the session security, which is achieved through authentication techniques [STALLING97A].

This concludes the discussion of RF transmission medium for WLANs. The next medium to be described is Infrared transmissions.

2.2.3.2. Infrared wireless LAN (IR WLAN)

Unlike RF transmissions, infrared transmissions in wireless communications do not need any licensing. There are no regulations for using infrared transmissions on free space or air, except the health regulations applied to the levels of the transmitted power [FERNANDES94].

One of the major properties of IR transmission is that it is by nature a highly directive transmission scheme. This means that they can be used for point to point communications, but not for broadcast systems. To use IR transmission for broadcast systems, some extra devices - such as a reflector or a repeater - should be used.

Light wave signals can be in one of three types. These types are categorized according to the wavelength of the transmitted carrier as follows: [STALLING97A]

? Infrared signals (IR).
? Visible Light signals (VL).
? Ultraviolet signals (UV).

The signal adopted in WLANs is the infrared signal because it is safer for people than ultraviolet, and it is more immune to ambient noise than visible light. Hence, in the following sections a detailed description will be given for only this scheme because of its applicability to WLANs, while the other two schemes are not used in wireless communication applications. In fact visible light is sometimes used for communication but usually in optical fiber techniques, while ultraviolet is used for measurements and sterilization application rather than communication applications.

Infrared transmission systems essentially exhibit the properties of any light wave transmission. Thus, they cannot penetrate through walls and other obstacles; they are of limited transmission ranges; and the geometrical optics techniques such as lenses and other refractors can be used to direct them.

Infrared transmissions are used widely in a variety of applications such as optical fiber transmissions and remote controls. Most of IR applications use one of two widely used IR wavelengths; namely the 800 nm and the 1300 nm wavelengths [STALLING97A]. IR

carriers used for WLANs occupy approximately the wavelength ranges 800-1300 nm [STALLING97A] and [PAHLAVAN95].

Infrared sources are mainly divided into two categories:

? Laser Diode sources LDs.

? Light Emitting Diode sources LEDs.

LDs provide higher speed because they produce a coherent source of light, that is, a very narrow band of frequencies (1-5 nm spectral-width[1]), but with higher power density concentration. This high power density concentration could cause severe eye damages if directed to the retina. Hence, unless high data rates are needed LDs are only used for optical fiber communications. On the other hand, LEDs produce lower power density concentration, and wider spectrum (25-100 mm spectral-width) [STALLING97A].

For Wireless LANs, IR carriers generated from LEDs are used. The wavelengths used are in the ranges:

? 800 nm - 900 nm [PAHLAVAN95].

? 850 nm - 950 nm [STALLING97B].

These wavelengths correspond to frequencies:

? 375 GHz - 333.333 GHz for the 800 nm - 900 nm IR.

? 352.941 GHz - 315.789 GHz for the 850 nm - 950 nm IR.

---

[1] Spectral-width here means that the carrier is not a single frequency carrier, as in the case of RF modulation, but it is rather a set of carriers (group of frequencies) that are modulated with the transmitted signal.

```
Transmitter: Encoder — Modulator — Laser or Light emitting diode
Optical Channel
Receiver: Optical Filter — Photo-diode detector — Demodulator — Decoder
Source data ... Received data
```
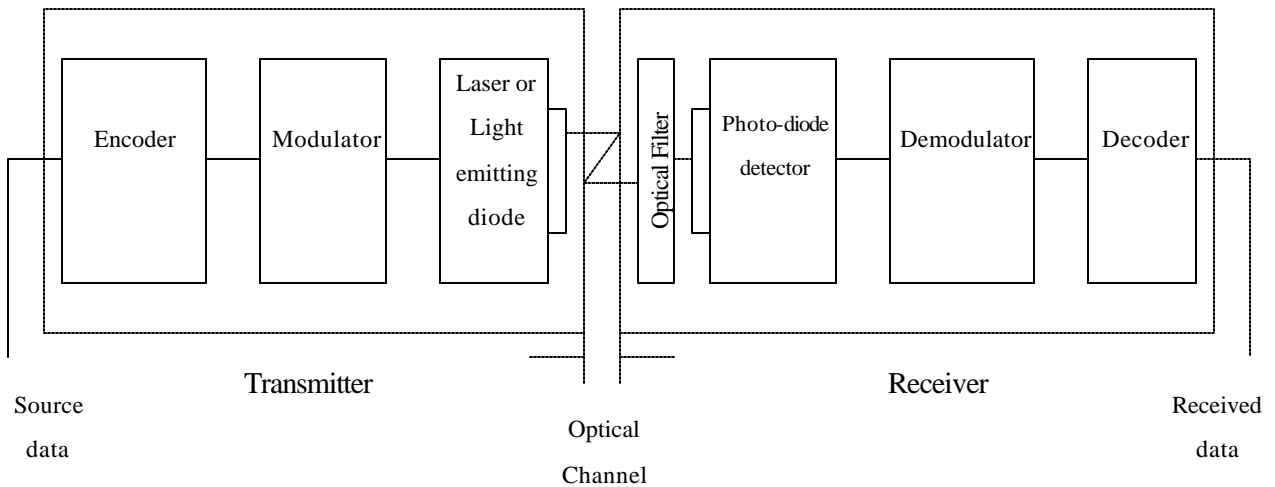
Fig 2-11 The optical transmitter and receiver

The wireless communication systems using IR transmission medium have the following characteristics: [STALLING97A]

- Infrared devices can work at high frequencies, yet with a cheap price. On the other hand, infrared detectors can only detect the amplitude of the received signal, and the sources are not capable of generating a single frequency carrier. Finally, the devices used for infrared transmission have the characteristic of separating the transmitters called *emitters*, and the receivers called *detectors* from each other. This means that for a single station there's a separate transmission device and a separate receiving device. In general these devices are in two categories. The first of these categories uses *Laser Diodes* (LD) for emitter and *Photo Diodes* for detector. The second uses *Light Emitting Diodes* (LED) for emitters and *Photo Diodes* also for detector (Fig 2-11). The LED sources are cheaper but with more carrier spectral-width and lower power levels. The LD sources are more expensive but with less carrier spectral-width and higher power level.

- Infrared transmission, as well as any other light wave transmission can not penetrate through opaque objects, such as walls, doors, and human bodies. They can only penetrate transparent materials as glass and water. Moreover, infrared waves can be reflected with shiny surfaces, and refracted with optical refracting materials.

- By nature, IR transmission, as well as any other light wave transmission technique, is a directive, and line of sight transmission technique. This means that the suitable topology for it is the point-to-point communication, but it still can be used for broadcast communication

28

with the help of a reflector that causes a diffusion to the light beam scattering it in all directions.

- Although IR communication uses directive methods in general, the power levels of these signals are generally of low values. These low values are because of health constraints to avoid damages of the eye and skin. It must be noted that when we speak about IR transmission the power level is represented with the intensity, which is measured in power per unit area. Usually, the limitations are on LD sources as they have high power intensities. The safe limit for LD sources is approximately 100 w/m² [FERNANDES94]. The same limit can be applied to LED sources in case they produce high power intensities.

- Because IR frequency range is not occupied by many devices, and it does not need any licensing, and also because they are much more higher in frequency than RF waves, more bandwidths are available for operation on this type of transmission. This can be seen by noting that the first band, which is the 800-900 nm band has a bandwidth of approximately 41.66 GHz, and the second band which is 850-950 nm has a bandwidth of approximately 37.16 GHz. It can be easily seen that these bandwidths are much larger than those available for RF transmission.

- IR transmissions are not affected by any kind of electrical interference. They are also not susceptible to interference form other IR WLANs in the neighbor because they do not penetrate walls. On the other hand, the ambient light contains a sufficient amount of IR signals to interfere with IR WLAN. The sources of IR waves in ambient light are mainly the sun light and fluorescent lamps. The interference effect of these sources could be very much reduced by the use of electronic band pass filters, which allow only the frequency of the required signal to pass. IR transmission also has the advantage of allowing the use of *Optical Filters[2]* to prevent undesirable waves from reaching the detector. As LED sources generate a wide-spectral carrier, their receivers should have a wide-band filter to allow for the transmitted signal to be detected. This wide-band receiver gathers more noise from the channel than a narrow-band one. On the other hand, LD sources allow for the use of narrow-band filters at the receiver. Hence, LD sources have better performance in terms of interference compared to LED sources.

- As IR transmission suffers from low power of radiation, high attenuation in the ambient atmosphere, and scattering effects due to humidity and dust in atmosphere, they have a low coverage range. This range is very much limited in case of LED sources, while it is longer in case of LD sources. It is also important to notice that point-to-point connection with IR

---

[2] Optical filters are materials that are transparent for a certain range of light wavelengths.

transmissions in outdoor and using a focused beam can reach coverage ranges up to few kilometers [STALLING97B]. On the contrary, indoor environment coverage ranges are longer in case of broadcast diffused connection than the case of point-to-point connection. This is because of the more power levels allowable for diffused connection, the ability to use active repeaters instead of just a reflector for diffused connections, and the ability to avoid emitter-detector alignment difficulties in case of point-to-point connections [STALLING97A]. Typical values for diffuse mode are 17-67 m of coverage range using LED sources in general except in the case of high data rates. On the other hand, point-to-point mode has a coverage range of only up to 27 m using LD sources or LED sources, but LD sources are preferred because of their fine beam width.

- As it can be seen from above, IR transmission channels have large bandwidths, which allows for higher data rates, but the limitation on data rates comes from the electronics technology, and also from the fact that Light Waves has a very wide range of noise sources and scattering materials. Also LED sources are much more susceptible to noise interference than LD sources and diffused techniques are much more vulnerable to interference than point-to-point connections. Due to these facts, the actual achievable data rates for IR WLANs are much less than those allowable by the bandwidth. For LED sources the data rates are 1-4 Mb/s for diffused connection, and 10 Mb/s for direct connection [PAHLAVAN95]. For data rates higher than 10 Mb/s LD are used [STALLING97A].

- As IR emitters are not capable of generating single frequency carriers, the most suitable and used transmission scheme is the amplitude modulation. Moreover, as the data rates achievable are already high, and as the implementation of intensity IR detectors is more expensive than on-off detector, multi-level amplitude modulation is not needed and usually not used for IR transmissions. The scheme of choice is *On Off Keying OOK* which is a binary amplitude modulation using the base-band signal to modulate the carrier light beam. Another scheme that is used to reduce the power requirement for the emitter, especially for LD sources is the *Pulse Position Modulation PPM*. In both cases to avoid inter symbol interference the multi-sub-carrier modulation is used [STALLING97A].

- Because IR transmissions are generally of limited coverage ranges, and they are not capable of penetration through walls, less security restrictions are required for them. A user outside the intended network room does not receive any significant amount of the optical signal, which means that eavesdropping from outside the office is practically very difficult. Also, as IR communication requires line of sight, the users will easily detect any interception to the transmitted beam. So, the nature of the signal itself has an imbedded security on the level of the network security. The actual security issue is again the

authentication security, and for more secured data transmission encryption may be used, but usually it is not essential.

This concludes the discussion of the Wireless LAN Physical layer technologies. The results of this discussion can be summarized in tables Table 2-1, Table 2-2, and Table 2-3.

	Infra Red		Radio Frequency		
	Diffused IR	Direct IR	FHSS	DSSS	Narrow-band
Data rate (Mb/s)	1-4 LED sources used	10-20 >10 using LD sources	1-3	2-20	5-20
Mobility	Stationary / Mobile	Stationary	Stationary / Mobile		
Range (m)	17-67	27	30-100	30-270	15-45 in ISM 20 km at 18 GHz
Eavesdropping immunity	Highest	Highest	High	High	Low
Frequency MHz / Wavelength (λ) nm	λ=800-900 λ=850-950		ISM band (902-928 2400-2483.5 5725-5850)		ISM band & 18-19 GHz
Transmission technique	OOK PPM		GFSK	QPSK	QPSK FSK
Radiated power	<1 w/m^2 for LED <6.4 w/m^2 for LD	<1 w/m^2 for LED <3.2 w/m^2 for LD	< 1w		25mw for ISM band < 1w for 18 GHz band
Topology	Star, Bus	Ring	Star, Bus		
Interference avoidance	High	High	Highest	Higher	Low
Frequency licensing	Not required		Not required		Required unless ISM

Table 2-1 Properties of different physical layer technologies [PAHLAVAN95]

	Ring	Star	Bus
Wireless media	Direct IR	Diffused IR & RF	
Coverage (Tx range is d)	The distance between to successive nodes in the ring $< d$	Area with radius d	Area with diameter d
Reliability	Lowest (any node failure causes the network to fail)	Low (hub failure causes the network to fail)	Highest (nodes connectivity are independent)
Architecture	Infrastructure	Infrastructure	Infrastructure Ad-hoc
Use for interconnecting wireless and wired LAN	Not feasible	Most preferred	Can be used
Node mobility	Stationary	Mobile within coverage area Cellular technology could be used with multiple hubs to improve mobility	Mobile within coverage area

Table 2-2 Properties of different topologies when applied to WLANs

	DSSS	FHSS
Interference immunity	Low (using the processing gain which makes it not effective for high power interfering signals)	High (by changing the hopping pattern to avoid the contaminated frequency band)
Synchronization	High (within the chip duration)	Low for slow hopping and higher for fast hopping
Receiver complexity	More complex because of higher synchronization restrictions	Less complex specially for slow hopping
Jammers	Affected by pulsed jammers	Affected by follower jammers and could be avoided by fas t hopping
Diversity	Achievable using the PN code	Achievable by fast hopping or by repetition codes
Data rates (Mbps)	2-20	1-3

Table 2-3 Comparison between DSSS and FHSS

2.3. <u>Medium Access Control Technologies</u>

The MAC is the main function of the Data Link Layer in the 7-layers ISO-OSI model. Its function is to guarantee collision free transmission over the Physical Layer. This can be accomplished through one of several available techniques. We have grouped these techniques in two major categories as follows:

- *Channel Multiplexing*, which allows for channel sharing by dividing the physical channel into several sub-channels and starting several transmissions simultaneously over them.
- *Single Channel Access*, which allows for channel sharing by granting access to the channel to only one station at a time.

2.3.1. Channel multiplexing

Channel multiplexing has three main techniques:

- Frequency Division Multiple Access (FDMA).
- Deterministic Time Division Multiple Access (TDMA).
- Code Division Multiple Access (CDMA).

Since channel bandwidth is limited in the case of RF WLAN and single carrier can not be practically produced in the case of IR WLAN, the use of FDMA is not adopted in WLAN. On the other hand, because data transmissions takes place in bursts, the conventional fixed deterministic TDMA is not applicable in case of WLAN [BANTZ94]. This leaves us with CDMA. Unfortunately, CDMA suffers from the near-far phenomena, which is very much induced in case of indoor communication. It also reduces the bandwidth allowed for a user because it spreads the signal. Hence, its use as a multiplexing technique for WLAN does not produce adequate performance [PAHLAVAN95]. From all of this, it can be seen that channel multiplexing techniques are not the right choice for WLAN and simultaneous multiple transmissions are not used.

The used MAC techniques are mainly dependent on granting channel access for only one transmitting node at a time i.e. Single Channel Access, and sometimes to improve performance an extra multiplexing is added to the Single Channel Access method.

2.3.2. Single channel access

In these techniques the MAC is required to either resolve the contention for the channel by nodes wishing to transmit, or to avoid the occurrence of such contentions in the first place. Hence, there are two main categories for single channel access MAC: [BANTZ94]

- Contention-Free (Controlled or Centralized) Access.
- Contention-Based (Distributed) Access.

Each of these two categories includes many MAC access techniques. Not all of these techniques are suitable for WLAN. In the following a description of each category will be given and a brief discussion of some of the MAC techniques under each, which are applicable to WLAN, will be given.

2.3.2.1. Contention-free access

This category of MAC techniques depends on avoiding any contention between the nodes of the network. This is achieved through a central network controller that manages the granting of the medium access, so that it is given to only one node till it finishes its current frame. Consequently, this technique is also called Controlled or Centralized Access. This central network controller could be one of the network nodes with this function added to it or it could be a dedicated device that only performs this function.

The function of the central control point (network controller) could be performed using several algorithms. Some of these algorithms are: [TOBAGI76]

Priority assignment in which the central control point assigns priorities to all other nodes. Accordingly, when more than one node wishes to access the medium at the same time, it is granted to the highest priority one and the other nodes should try to get access later. This technique has the following properties:

- It minimizes the transmission delay because either the transmission request is served immediately or it is dropped and the transmission should be started all over again.
- It reduces the throughput of the network, which is the number of successfully transmitted information packets per unit time, because transmission request packets for non-served stations are retransmitted again wasting some of the channel capacity for these control packets containing no information.

- By its nature, the scheme is not a fair scheme i.e. some nodes are preferred over others in service. To achieve higher degree of fairness, the priority assignment function could be reorganized after each transmission.

Polling in which the central access point polls the network nodes in a sequential manner to see if any node is wishing to transmit. When it finds a node wishing to transmit, it gives it access to the medium and prohibits access from other nodes till the end of the current transmission. This technique has the following characteristics:

- The transmission delay is fixed and depends on the number of the stations in the network and the order of the transmitting station in the polling sequence.
- The throughput of the network is low because the probability of wasting a long time in polling non-transmitting stations before reaching the station wishing to transmit is high. During this time the channel is idle and no data is transmitted over it, which reduces the throughput.
- The scheme has an embedded priority as stations at the start of the polling sequence have priority over those at the end of the sequence. This priority effect could be reduced if the sequence is reordered each time a transmission is served.

Reservation access in which stations wishing to transmit, send their reservations to the central control point. These reservations are queued in the central control point, then the controller serves these requests according to a certain service discipline, like first come first serve or highest priority first. This scheme has the following properties:

- The transmission delay is unbounded and could reach high values. It depends on the reservation queue service discipline, because after a station sends its reservation it has to wait till it is told by the controller that it has been granted the medium.
- The throughput is very much enhanced because reservations that are not served in the current time slot are queued in the controller till a slot is assigned for them. Hence, no overhead is induced, resulting from re-transmitting the reservation packets over and over again.
- The scheme by its nature is a fair access scheme and the fairness degree is determined by the reservation queue service discipline.

The main point in all these techniques is to try to divide the channel bandwidth equally between nodes and use it as efficiently as possible. From the previous description of this MAC category and from network protocols studies, it is found that a Contention Free MAC has the following properties:

- Contention-free techniques have better performance in highly loaded networks. Its performance is less efficient in the case of light traffic loads.
- Contention-free techniques rely on a central controlling node. This makes them more suitable for use with star network topologies as the central controlling node will be the hub.
- As contention-free techniques are centralized techniques, they can make use of this centralization to retune the MAC parameters dynamically to accommodate the current traffic load and achieve higher performance. This gives the techniques the ability to be adaptive.
- As contention-free access techniques are centralized techniques they are more suitable to infrastructure WLAN. They cannot be used for ad-hoc networking, because they need to have a pre-specified access controller.
- Contention-free access techniques like any centralized scheme have lower reliability, because of network single point of failure.
- Network setup transient state usually takes less time in these schemes because they are performed only by the central controlling point.
- Contention-free access schemes are usually slotted access schemes. This time slotting helps in simplifying the algorithm used to avoid contention on the channel.

2.3.2.2. Contention-based access

When it comes to contention-based MAC, it must be noticed that there is no single controller for the network. On the contrary, all network stations perform the MAC function in the same manner. This means that the function of granting access to the network is distributed over all the network nodes. Hence, all the network nodes have to find some way to resolve any conflict that may happen when two or more nodes wish to access the medium at the same time. This is why this technique is sometimes called Distributed Access technique.

Contention-based access techniques are mainly divided into two main categories. This categorization is based on whether a node wishing to access the medium checks first for its busy state or not. The two main categories are: [KLEINROCK75], [BERTSEKAS87]

Carrier sensing MAC in which any network node wishing to transmit performs a *listen before talk* function first. In this listen before talk function, the station listens to the medium to see if there are any ongoing transmissions on the medium i.e. to check if the medium is busy or not. If the medium is detected to be busy i.e. there is a transmission on it, the listening station defers its transmission till the medium is free. When the medium becomes free, a station

wishing to send can start its transmission immediately or after a certain time, provided that the medium is still free after this time. Whether the station starts its transmission immediately after the medium is sensed to be free or not, is dependent on the specific Carrier Sensing MAC used. One of the most famous carrier sensing MAC techniques used for WLAN is the so called Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), which is a variation of the famous Ethernet MAC called Carrier Sense Multiple Access with Collision Detection (CSMA/CD). The carrier sensing techniques have the following properties:

- Packet delay could be slightly increased because of the listen process performed before each transmission, but as this listening process reduces collision, it generally reduces the packet delay.
- Throughput of the system is very much enhanced because of the reduction of the number of corrupted packets due to collisions.
- The scheme suffers from the problem of hidden terminal and has the advantage of the possibility of capture when used with WLAN, because of the wireless channel highly dynamic characteristics.

Non-carrier sensing MAC in which a station wishing to transmit does not listen to the medium to check its status. On the contrary, the station starts its transmission immediately when it has a packet to send or after a certain time depending on the specific scheme used. This of course could result in high collision rates because simultaneous transmissions from different stations could happen resulting in corrupting the data signal on the medium. To reduce the collision rate time slotting is associated with this scheme. One of the most famous WLAN MAC techniques, which belongs to this category is the Slotted Aloha transmission scheme. The Non-carrier Sensing schemes have the following properties:

- Although the overhead delay of listen before talk process is removed, packet delay could still reach high values because of the higher collision rates.
- Throughput of the system is very much reduced because of the higher collision rate induced in the scheme. This throughput is very much enhanced with the use of time slotting.
- The system does not suffer from hidden terminal problems, as there are no carrier sensing performed. It also still has the advantage of the possibility of capture because of the wireless media nature.

The main characteristic of these techniques is to try to resolve contention on the shared medium, which happens when two or more stations wish to transmit at the same time.

From the foregoing discussion and from network protocol analysis it is found that contention-based MAC protocols have the following characteristics:

- Contention-based techniques have better performance in light loaded networks. Its performance is less efficient in the case of high traffic loads.
- Contention-based techniques rely on a distributed control technique. This makes them more suitable for shared medium network topologies such as bus topology.
- Although contention-based techniques are distributed techniques, some adaptive techniques could be induced in them by assigning one of the nodes an extra function to collect traffic parameters and calculate MAC parameters from them. This is still very difficult and less efficient in terms of enhancing performance than the case of contention-free access.
- Contention-based techniques are suitable for both infrastructure and ad-hoc WLAN, as there are no predefined network topology needed for them. They are, however, not very much suitable for interconnecting Wireless LAN to a Wired LAN as this requires an access point.
- Contention-based techniques are distributed techniques. Hence, there is no single point of failure. Accordingly, they have higher reliability.
- Network setup transient state takes a somewhat longer time here because they need to be performed by all the network nodes.
- Contention-based access schemes could be either time-slotted or non-slotted schemes.

This concludes the discussion of the MAC techniques used for WLAN. In the following a table that compares between the two main categories is given to summarize the discussion that was given in this section.

	Contention-free	Contention-based
Performance	Better in high loads	Better in low loads
Topology	Star	Bus
Adaptability	More adaptive	Less adaptive
Network type	Infra-structure	Infrastructure Ad-hoc
Reliability	Low	High
Setup transient time	Low	High
Time slotting	Required	Could be used or not

Table 2-4 Comparison between contention-free and contention-based MAC

Finally, it should be noticed that one of the adopted techniques in WLAN products uses a hybrid mixture of contention free and contention based techniques. This technique is called Packet Reservation Multiple Access (PRMA). This technique is described in details in Appendix B. Another protocol that belongs to the contention-based protocols but includes also CDMA is the Inhibit Sense Multiple Access (ISMA) [ROOSMALEN94].

2.4. Conclusions

In this chapter we have given a review of the technologies used in WLANs physical and MAC layers. In the physical layer we have studied the different network architectures, the different topologies, and the different transmission medium characteristics. Afterwards, we studied the different MAC categories that can be used for wireless LANs.

During this review of literature we found that:

WLANs architectures are either permanent infrastructure or temporary ad-hoc architectures.

WLANs can be configured in any of the three known main topologies; namely bus, ring, and star topologies. The only restriction is that ad-hoc networks cannot be implemented using ring or star topologies as both needs pre-specified controllers.

The electromagnetic waves used for wireless transmission medium is either RF waves or IR waves. It was found that RF has lower bit rates, higher ability to penetrate obstacles, and is

more suitable for broadcast ad-hoc configurations such as the bus topology. On the other hand, IR is more directive, which makes it more suitable for infrastructure point-to-point applications such as infrastructure ring topology.

MAC protocols used for wireless LANs are either contention-based protocols, which has the advantage of ease of reconfigurations, or contention-free protocols which has better performance specially under high loads.

From the MAC protocols characteristics it was found that the contention-based protocols are more suitable for ad-hoc networks. On the other hand, contention-free protocols are more suitable for infrastructure networks and interconnecting wireless LANs to wired LANs.

Lots of varieties exist in both physical layer and medium access control techniques. This implies that standardization is needed to guarantee interoperability between different vendor products.

Chapter 3

Wireless LAN Standards

Chapter 3
Wireless LAN Standards

3.1.   Introduction

In the following, a detailed description of the two standards currently worked on for wireless LAN will be presented.   These standards are the IEEE 802.11 standard, and the ETSI HIPERLAN standard.   A brief description of the frequency sharing etiquette adopted by WINFORUM[3] will then conclude the chapter.

3.2.   Standards for Wireless LAN

Up till now, all standards for wireless LANs use unlicensed bands [PAHLAVAN95] in USA. On the other hand, in Europe the CEPT has ratified the 5.15 GHz - 5.30 GHz for use by Wireless LAN operating according to HIPERLAN standards.   In general, there are two approaches to regulate the wireless LAN bands:

1. The first approach is to develop standards that insure that products of different vendors can intercommunicate with each other using a set of inter-operable rules. This approach is used in both IEEE 802.11 and ETSI HIPERLAN.

2. The second approach is to define a set of rules or the so called *"Spectrum Etiquette"* that allow terminals produced by different vendors to have a fair share in the available channel frequency and time and coexist in the same band. The WINFORUM group adopts this approach.

The standards that are going to be mentioned in the following are supposed to affect only the first two layers of the OSI model; namely, the Physical and Data Link layers. In a more specific manner, the Data Link layer is subdivided into two sub-layers named *Media Access Control (MAC)* and *Logical Link Control (LLC)* in ascending hierarchical order (Fig 3-1). The Wireless LAN standards apply to the whole Physical layer and to the MAC sub-layer of the Data Link layer.

---

[3] WINFORUM is an alliance among the major computer and communication companies to obtain bands from FCC for the data-PCS and was initiated by Apple Computer in the US.

The Physical layer covers the transmission media type, interfacing to the media used, electrical voltage levels, and transmission rates. For wireless LANs these topics correspond to whether IR or RF is used, the frequencies used, and the transmission techniques adopted.

On the other hand, the Data Link layer deals with how the network is shared between nodes. More specifically, the MAC sub-layer is responsible for setting rules covering when each node can gain control of the network to send its messages.

Finally, the LLC layer is responsible for providing logical links (connection-oriented services) between nodes, as well as a unified interface to the upper (Network) layer. This interface is independent of the various frame formats used.

Although the standards only specify the characteristics of the Physical and MAC layer, they still have effects on upper layers of the OSI model. As an example for such effects, consider the case of an application running on a wired LAN that assumes a certain timeout for responses from the other side. If such an application is used with the same time-out with a wireless LAN, it will malfunction. This is because the delay in the wireless media is much more than it is in a wired media.

Another example of such effects, is the fact that wireless nodes generally tend to appear and disappear in the network more frequently than the case of wired nodes. This of course will have an effect on a routing application running on the network as it is supposed to keep track

	Application
	Presentation
	Session
	Transport
LLC	Network
MAC	Data Link
Physical	Physical
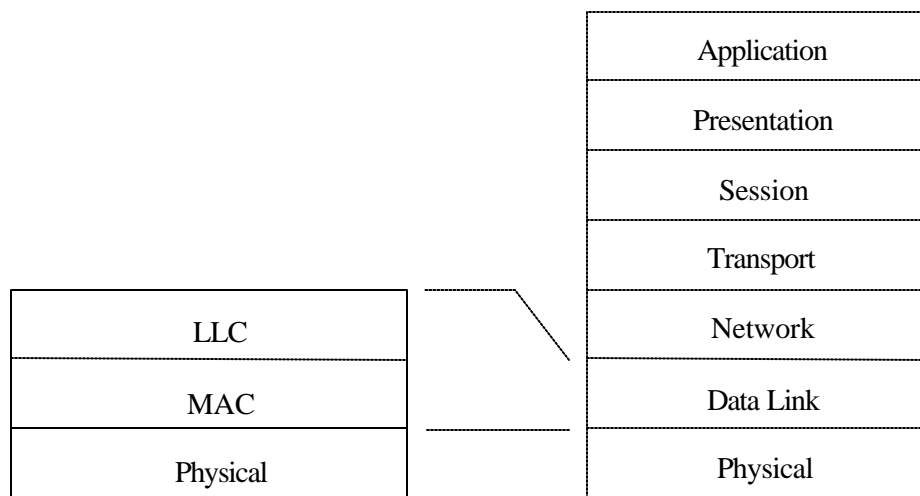
of the connected and disconnected nodes.

Fig 3-1 The ISO-OSI seven layers model

3.3.   IEEE 802.11 Wireless LAN Standards

The IEEE 802.11 committee has started working on the wireless LAN standard since 1990. It was supposed to be finalized by the end of 1995, but it is not finalized until now. Many of the MAC and Physical layer specifications are in a nearly final form by now. Although, the standards were supposed to be finished by the end of 1996 [LAMAIRE96], in a recent paper in the literature review the standards are still referred to as *draft standards* [CROW97].

The main scope of the 802.11 standard is to develop a Medium Access Control (MAC) and Physical layer (PHY) specification for wireless connectivity for fixed, portable, and moving stations within a local area [CROW97].

The purpose of the standards is twofold [CROW97]:

? To provide wireless connectivity to automatic machinery, equipment, or stations that require rapid deployment, which may be portable or handheld or which may be mounted on moving vehicles within a local area.

? To offer a standard for use by regulatory bodies to standardize access to one or more frequency bands for the purpose of local area communication.

To describe the standard we have to look at its three main parts; namely, the network architecture, the physical layer, and the medium access control layer.

3.3.1.   Architecture

The fundamental building block of the IEEE 802.11 architecture is what is called the Basic Service Set (BSS). It is defined as a group of stations that are under the direct control of a single coordination function whether this coordination function is a centralized (contention free) scheme, or a distributed (contention based) scheme. The geographical area covered by a BSS is called the Basic Service Area (BSA). Conceptually, all stations within the same BSS can communicate directly to each other if the channel is assumed to be ideal.

IEEE 802.11 standard supports both of the two WLANs connections described in the previous chapter. These connections are the ad-hoc connection and the infrastructure network connection. These connectio ns are supported in the IEEE 802.11 standards by adopting two types of network architectures. These two types are [ENNIS96]:

? Independent Basic Service Set (IBSS).

? Infrastructure Basic Service Set (ISBSS).

A description of these two types will be given in the following with an emphasis on the wireless connection type it supports.

3.3.1.1. Independent Basic Service Set (IBSS)

This is a deliberate grouping of stations into a single BSS for the purposes of inter-networked communications without the aid of an infrastructure network (Fig 3-2). This type of network is implemented using the ad-hoc wireless LAN connection type.
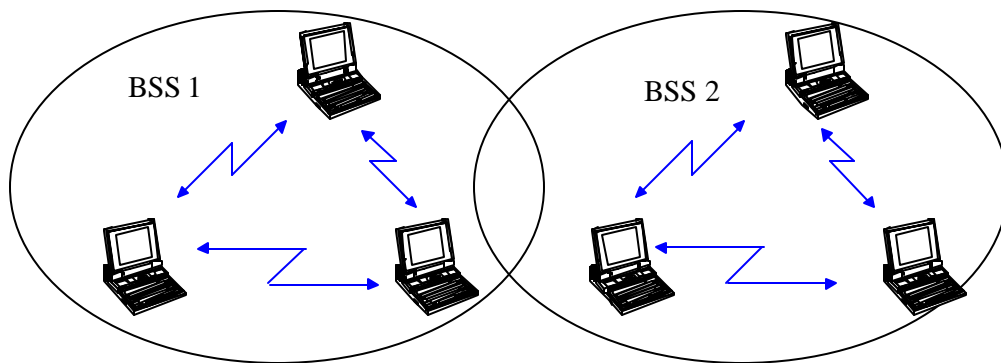
Fig 3-2 Two independent BSS with coverage area overlap

The BSSs are independent and cannot be interconnected together to form larger networks. On the other hand, if more than one IBSS contain a common area inside their coverage range, their operations in this area should not interfere. This means that neither of the two BSSs should be able to receive the communication signals of the other BSS, as they will be using two different wireless channels e.g. two different frequencies, or two different spreading codes.

This type of network has the following characteristics [ENNIS96]:

? Only one BSS can exist in the same network and multiple networks cannot be interconnected to cover larger area.

? It uses direct communication between nodes (peer-to-peer).

? It has a limited coverage area, which cannot exceed the maximum coverage area of one BSS.

46

- ? If multiple BSS's have a common part of the coverage area they should not interfere together (Fig 3-2).
- ? It does not support roaming and mobility is very limited.
- ? It does not support connecting wireless stations to fixed wired LAN resources
- ? Power management is not efficient, because no central point for traffic flow, that could monitor all the network activity, exists.
- ? Network access parameter refinement is not efficient, as network traffic control is not performed by a single station.
- ? Time bounded services are not supported with the required Quality of Service (QoS).

3.3.1.2.   Infrastructure Basic Service Set (ISBSS)

This is, as the name implies, mainly infrastructure network that is established to provide wireless users with specific services and range extension. This network is established using Access Points (AP), i.e. it consists of a group of Stations (STA) associated with an AP to form a BSS (Fig 3-3). The AP is the central controller for network traffic, i.e. it is equivalent to base station in cellular telephone networks.

AP supports range extension by providing the integration points necessary for network connectivity between multiple BSSs, thus forming an Extended Service Set (ESS). The ESS has the appearance of a one large BSS to the Logical Link Control of each station. The ESS consists of multiple BSSs that are connected together via their APs to a common Distribution System (DS). The function of the DS is not a part of the standards, as it could be any other IEEE standardized network or non-IEEE standardized network, but its required services are defined. While the DS could use the same Physical layer of the BSS, it is logically different because it is only used as a transport backbone to transfer packets between different BSSs in the ESS.

An ESS could provide a gateway access to a wired network. The device used for this connection is called a *Portal*. If this connection is to a wired 802.x network, the portal function is a bridge rather than a gateway.

The Portal is defined as a logical entity that specifies the integration point on the DS where the wireless IEEE 802.11 network integrates with a wired network.
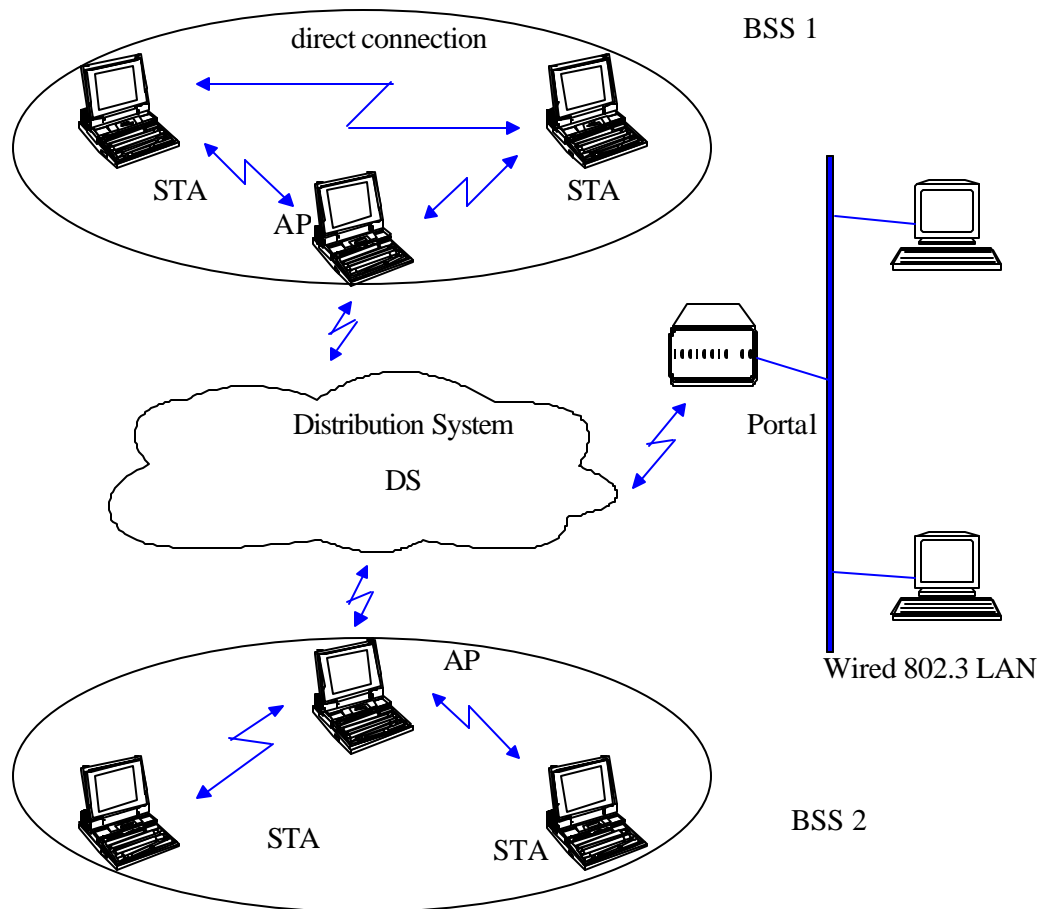
Fig 3-3 Infrastructure BSSs connected via a DS to form an ESS

Finally, the traffic between any two stations within the same BSS typically flows through the AP, but it could be direct from one station to the other.

This type of network has the following characteristics:

? More than one BSS can coexist in the same network forming an ESS and multiple networks can be interconnected to cover larger area.

? It uses centralized (Station - Access Point) communication in most cases (Fig 3-3 in BSS 2), but direct communication between nodes (peer-to-peer) within the same BSS could still be used (Fig 3-3 in BSS 1).

? It has a large coverage area, which can be extended by interconnecting more than one BSS to form a larger ESS through a DS.

48

- ? If multiple BSS should be connected together their APs should be able to connect to each other through the DS.
- ? It supports roaming between different BSSs in the same ESS.
- ? Wireless mobile stations can be connected through Portals to fixed wired LAN resources.
- ? Power management is highly efficient, because all network activities are monitored by the AP.
- ? Network traffic parameters refinement is efficient as network traffic control is performed by a single station, which is the AP.
- ? Time bounded services such as voice and video communication can be supported with an adequate QoS.
- ? It is very important to note that the IEEE 802.11 standards do not support multi-hop routing, which means that one node can be used as an intermediate link between another node and the base station. Consequently, in an ad-hoc network a station can only connect to another node directly, and in the Infrastructure topology a station can only send packets through the AP or directly to another station in the same BSS.

3.3.2. Physical Layer

Each kind of LAN requires unique specification of how the data is transmitted physically. This is done in the physical layer level. The IEEE 802.11 draft standard supports three types of physical layers [HAYES96] and [CROW97]:

- ? Frequency Hopping Spread Spectrum (FHSS) RF transmission using the 2.4 GHz ISM band.
- ? Direct Sequence Spread Spectrum (DSSS) RF transmission using the 2.4 GHz ISM band.
- ? Diffused Infrared (DFIR) transmission using an IR of wavelength in the range 850-950 nm.

In addition to having three types of physical layers (PHY), two different data rates are adopted. These data rates are as follows:

- ? The Basic Rate with bit rate equal to 1Mb/s, which is mandated to be supported by all physical layers.
- ? The Enhanced Rate with bit rate equal to 2Mb/s, which can be optionally supported by any of the physical layers.

It must be noticed that although the Medium Access Control (MAC) layer is the same for all of these physical layers, which means that the MAC Protocol Data Unit (MPDU) is the same for all, the Physical Layer Convergence Protocol frame format (PLCP) is different. This frame

contains the MPDU packet scrambled by whitening scrambling methods and its DC offset is removed by adding DC offset removal bits to it. Also, added to it the physical layer dependent controls such as preamble, and header parts. These frames will be explained in each physical layer independently as they vary from one to another.

Another important issue is the fact that the MPDU part of the PLCP frame, which is called the PLCP_PDU, can be sent using either basic access rate of 1 Mb/s or the enhanced access rate of 2 Mb/s. On the other hand, the other PLCP physical layer dependent parts are only transmitted using the basic access rate of 1 Mb/s.

The three physical layers mentioned above will now be explained in details, with emphasis on the data rates supported, the transmission techniques, the carrier frequency or light wavelength, and the modulation scheme.

3.3.2.1. Frequency Hopping Spread Spectrum (FHSS)

This is the first physical layer adopted by the IEEE 802.11 standard. The technique itself has been described in Chapter 2. The regulations for this scheme are as follows [CHAYAT96]:

? Frequency band is 2400-2483.5 MHz.
? Slow hopping is the adopted scheme.
? Transmitted power level is less than 1 watt.
? The band is divided into sub-channels each with 1 MHz bandwidth. This results in an 83 sub-channels. These 83 sub-channels allow for 83 different frequency hops (Fig 3-4).
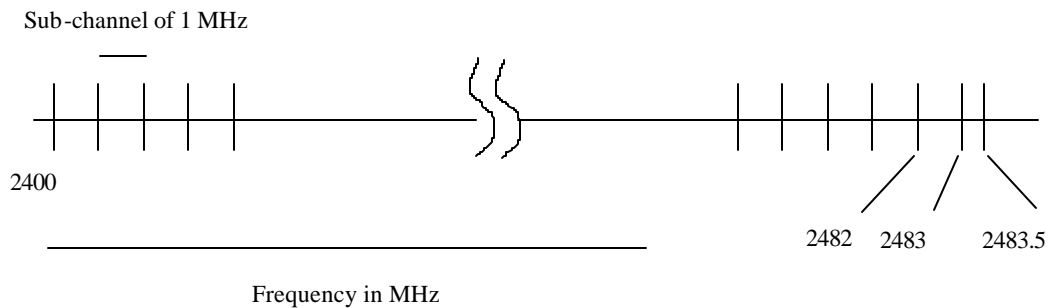
Fig 3-4 Organization of the 2400-2483.5 MHz band for FHSS

? Only 79 hopping center frequencies are used of the total 83 channels available in the band. The first channel, allowed to be used, is the one with the center frequency at 2402 MHz.

? The pseudo-random hopping sequences (patterns) are arranged into three different sets of hopping sequences. Each set contains 26 different hopping sequences each hopping sequence contains 79 frequency hops. This gives a total of 78 hopping sequences.

? A minimum hop distance of 6 channels is required by any hopping sequence to minimize adjacent channel interference and maximize multi-path resolving feature (Fig 3-5).

? Sequences from the same set are chosen so that they collide on only 3 channels on average and 5 channels at most over a hopping sequence cycle.

? A single BSS uses only one hopping sequence for all its transmissions; i.e. CDMA is not performed within the same BSS.

? If two or more different BSSs exists in the same radio contact area, each one must use a different hopping sequence from the same set.

? The frequencies $f_i(k)$ of the sequence number k, where k is from 1 to 78, is formed from the base sequence offsets, denoted by $b[i]$, where $b[i]$ is the center frequency number i in the basic (first) hop sequence minus the start frequency (2402 MHz), and i is the hop index inside one sequence in the range 0 to 78, using the equation: $f_i(k)=2402+[(b[i]+k)mod(79)]$.

? The minimum hopping rate of 2.5 hops/s is assigned to allow for a complete packet to be sent in one hop. Thus, if it interfered, the retransmission packet will be sent on another frequency. This increases the immunity against interference and reduces retransmissions.
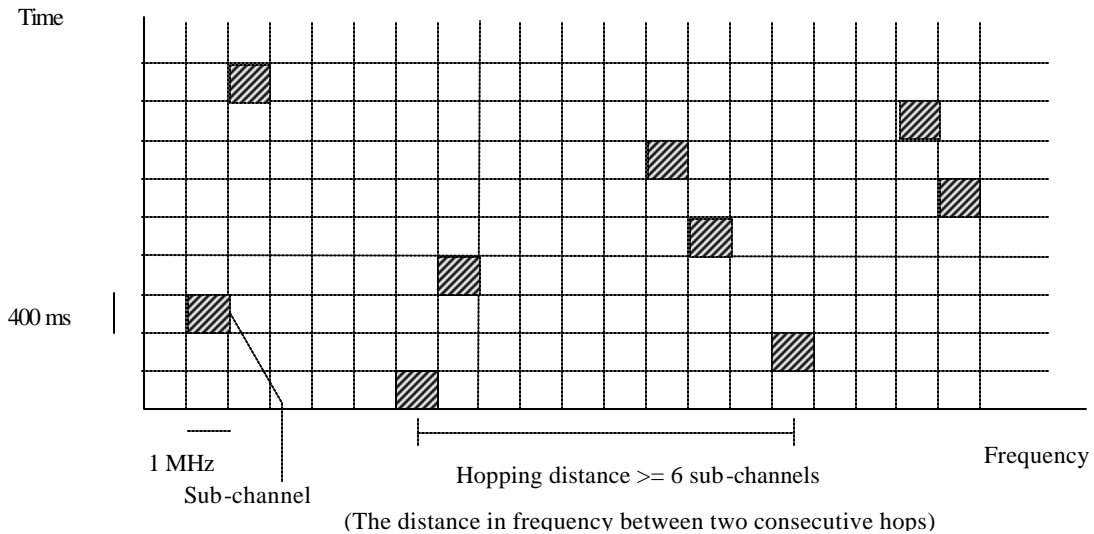
Fig 3-5 Hopping sequence requirements for 802.11

? For the basic access rate of 1 Mb/s the modulation scheme used is the 2-level GFSK (Gaussian Frequency Shift Keying).

? For the enhanced access rate of 2 Mb/s the modulation scheme used is the 4-level GFSK.

? Only the PLCP_PDU part of the frame is required to be scrambled so that the data becomes whitened. This scrambling is performed synchronously by XORing the data with periodic 127 bit Linear Feedback Shift Register (LFSR) sequence generated by the feedback polynomial $1+x^4+x^7$ (Fig 3-6) [CHAYAT96].

? The DC offset of the transmitted PLCP_PDU binary sequence should be removed to avoid its effect on Phase Locked Loop (PLL) based systems. This elimination is done by grouping the transmitted data in 32 symbols blocks and adding a symbol that holds the sign information of the block to it, thus getting a 33 symbols block. Next, the sign of the DC offset of every 33 block is checked against the sign of the previous block; if it is the same, the block is inverted, else it is left as is.

? Finally, the PLCP frame has the format shown in Fig 3-7.
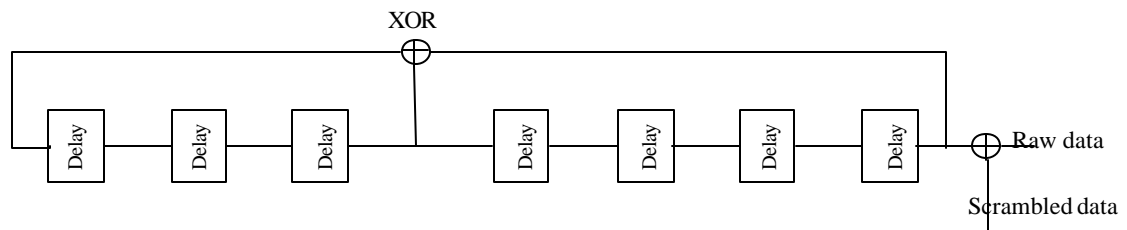
Fig 3-6 LFSR generator of scrambling bit sequence

PLCP preamble		PLCP header				PLCP_PDU
Sync pattern	SFD	PLW	4 bit	HEC		Payload data
80 bit	16 bit	12 bit		16 bit		(variable length)

PSF

Physical layer header and preamble always at 1 Mb/s   MAC data could be at

1 Mb/s or 2 Mb/s
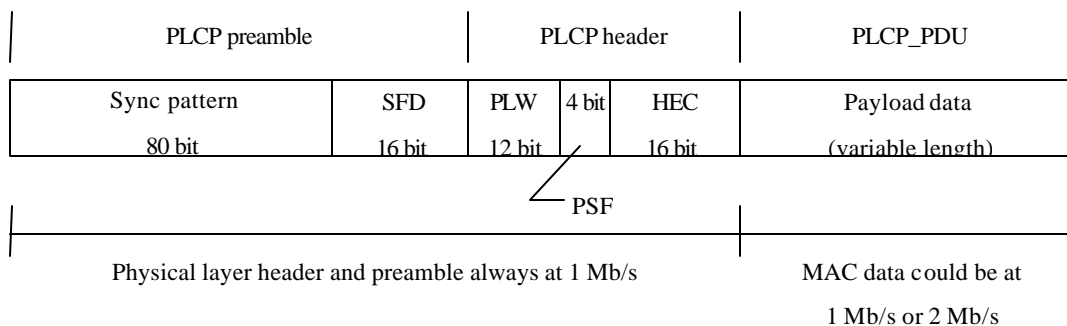
- *Sync pattern* is a synchronization pattern of 80 bits length.   It has the binary format '0101010101…0101' (with left bit transmitted first) and used for detecting signal presence, resolving antenna diversity, and acquiring symbol synchronization.
- *SFD* stands for Start Frame Delimiter, is a 16 bit block that has the binary format '0000 1100 1011 1101' (with left bit transmitted first). This field is used for frame level synchronization, as it marks the start of the frame.
- *PLW* is an acronym for PLCP_PDU Length Word, is a 16 bit block that contains the length of the PLCP_PDU part in octets.
- *PSF* stands for PLCP Signaling Field, consists of 4 bits, three of which are reserved for future use and the fourth least bit is used to signal the PLCP_PDU data rate. A binary '1' in this field indicates a bit rate of 2 Mb/s, while a binary '0' indicates a bit rate of 1 Mb/s.
- *HEC* is the Header Error Check, is a 16 bit CRC block that represents an error check sequence for the 32 bit PLCP header part only. The CRC is generated using the CCITT generator polynomial called *G(x)*, where $G(x)=1+x^5+x^{12}+x^{16}$.
- *PLCP_PDU*, is the packet received from the MAC layer (i.e. the MPDU), which includes the data required to be transmitted by the user (i.e. the MSDU), the addressing, and the MAC CRC.

Fig 3-7 PLCP frame format for FHSS in 802.11 standards

3.3.2.2.   Direct Sequence Spread Spectrum (DSSS)

This is the second Physical layer adopted by the IEEE 802.11 standard. Again the technique has been described before. The 802.11 regulations for this scheme are as follows [BOER96]:

- Frequency band is 2400 MHz - 2483.5 MHz.
- The band is divided into 11 sub-channels each will have a bandwidth of 11 MHz. These sub-channels are overlapping. The frequency spectrum diagram of these channels is shown in Fig 3-8.
- Transmitted power level is less than 1 watt.
- If more than one BSSs are going to overlap or coexist in the same radio coverage area, the center frequencies of their channels should be separated by at least 30 MHz. This leads to being able to have at most only two BSSs in the same radio coverage area.
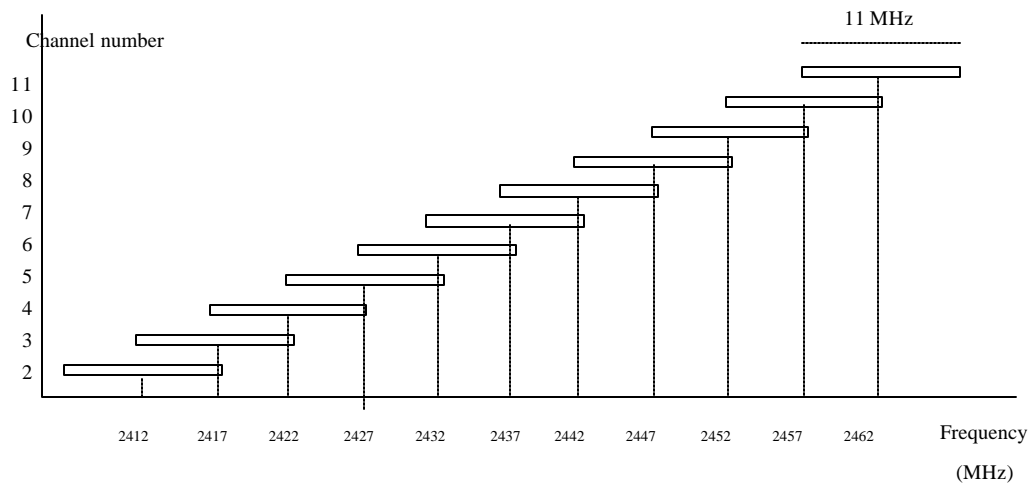
Fig 3-8 The 802.11 DSSS frequency sub-channels diagram

- The pseudo-random spreading sequences are defined to have 11 chips Barker sequence per symbol. This means that the processing gain is 11 i.e. 10.4 dB.
- The Barker sequence is an aperiodic random sequence that posses all the aspects of a pseudo random code. The 11-chip Barker sequence has the binary form '10110111000' and has an auto-correlation function that is constant and equal to -1 for any bit shift, except for the case of complete synchronization where it becomes 11 [STALLING97A].
- As one channel has a bandwidth of 11 MHz and the chip rate is 11 chips per symbol then the symbol rate per channel is 1 Mb/s if a binary scheme is used, i.e. no multilevel modulation is used.
- The spreading codes are chosen to be as orthogonal as possible, i.e. minimal correlation between spreading codes is achieved as much as possible.
- A single BSS uses only one spreading sequence for all its transmissions, i.e. CDMA is not performed within the same BSS.
- For the basic access rate of 1 Mb/s the modulation scheme used is the 2-level DBPSK (Differential Binary Phase Shift Keying).
- For the enhanced access rate of 2 Mb/s the modulation scheme used is the DQPSK (Differential Quadrature Phase Shift Keying).
- The entire PLCP frame is scrambled including the header and the preamble parts. The scrambling is performed with the same technique used for FHSS before. The scrambling is used to achieve two properties in the transmitted sequence. The first is the whitening of the

transmitted data bits, and the second is the DC blocking (DC offset removal) of the transmitted data packets.

- Finally the PLCP of DSSS has the frame format shown in Fig 3-9.

PLCP preamble		PLCP header				PLCP_PDU
Sync pattern	SFD	SG	SR	LN	HEC	Payload data
128 bit	16 bit	8 bit	8 bit	16 bit	16 bit	(variable length)

Physical layer header and preamble always at 1 Mb/s   MAC data could be at

1 Mb/s or 2 Mb/s
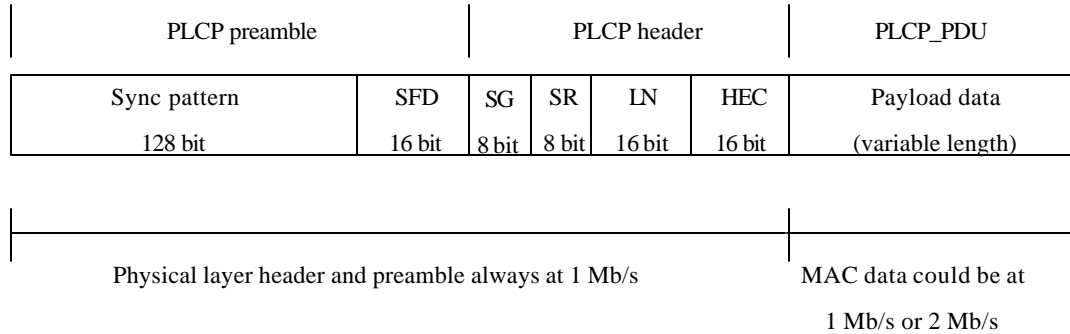
- *Sync pattern*, which is a synchronization pattern of 128 bits length. It has the binary format '111111…1'. It is scrambled with the data scrambling sequence, and used for detecting signal presence, adjusting gain settings, resolving antenna diversity, and finally for frequency offset compensation.
- *SFD*, which stands for Start Frame Delimiter, is a 16 bit scrambled block that has the binary format '0000 0110 1100 1111' (with left bit transmitted first). This field is used for frame level synchronization, as it marks the start of the frame.
- *SG*, which is an acronym for Signal Field, is an 8 bit scrambled block which contains a value indicating the transmission rate of the PLCP_PDU field. A binary value of '0110 0000' (with left bit transmitted first) indicates a 1 Mb/s with DBPSK and a binary value of '0010 1000' (with left bit transmitted first) indicates a 2 Mb/s with DQPSK. Other values of this field are reserved for future uses.
- *SR*, which stands for Service Field, consists of 8 bits. This field is reserved for future use except for the value of '0000 0000' which indicates an 802.11 compliant frame.
- LN, which stands for Length Field, is a 16 bit field that contains the length of the PLCP_PDU part of the frame in

Fig 3-9 PLCP frame format for DSSS in 802.11 standards

3.3.2.3.   Diffused Infrared (DFIR)

This is the third and last physical layer supported in IEEE 802.11 standards. The standard regulations for this technique are not completed yet, but the current specifications are as follows [HAYES96]:

- The used IR beam has wavelengths that are very near to the visible light band. The wavelengths used are in the range 850-950 nm.
- Diffused IR (DFIR) transmission techniques are used to avoid the need for line of sight clearance between the transmitter and the receiver.
- The signal transmitted power is less than 2 watts.

- The technique is used only indoors, and not used at all in outdoors.
- For the access rate of 1 Mb/s 16-PPM (Pulse Position Modulation) is used where 4 data bits are mapped into 16 coded bits.
- For the access rate of 2 Mb/s 4-PPM (Pulse Position Modulation) is used where 2 data bits are mapped into 4 coded bits.
- The PLCP frame has the general format shown in Fig 3-10.

PLCP preamble		PLCP header				PLCP_PDU
SYNC 80 bit	SFD 16 bit	DR 8 bit	DCLA 8 bit	LEN 16 bit	HEC 16 bit	Payload data (variable length)

Physical layer header and preamble always at 1 Mb/s   MAC data could be at

1 Mb/s or 2 Mb/s
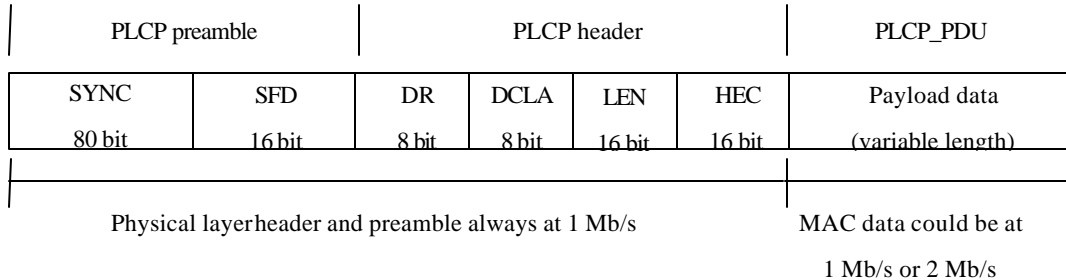
- *SYNC*, which is a the synchronization pattern.
- *SFD*, which stands for Start Frame Delimiter, is used for frame level synchronization, as it marks the start of the frame.
- *DR*, which is an acronym for Data Rate, contains a value indicating the transmission rate of the OLCP_PDU field.
- *DCLA*, which stands for DC Level Adjustment, is a field used to control the DC value of the data bits.
- *LEN*, which stands for Length, contains the length of the PLCP_PDU part of the frame in octets.

Fig 3-10 PLCP frame for DFIR in IEEE 802.11 standards

3.3.3.   Medium Access Control (MAC)

The IEEE 802.11 defines only one Medium Access Control (MAC) for all the previously mentioned Physical layers. This definition of one MAC for all the different Physical layers allows for an easy and cheap production of network controller chip sets as the same chip (network processor) can be used on different physical interfacing cards according to the Physical layer used.

Generally speaking a MAC layer is responsible for the following functions [DIEPSTRATEN96]:

- Channel allocation procedures.
- Protocol Data Unit (PDU) addressing.
- Frame formatting (data packetization).

The IEEE 802.11 MAC is required to have some more extra properties in order to cope with the characteristics of wireless channels and equipment, and to satisfy the different needs of the WLAN users:

- The MAC should support both the infrastructure (centralized) BSS and ad-hoc (independent) BSS.
- It should support asynchronous (burst) traffic, such as data packets, as well as time-bounded (time-critical) traffic, such as video and voice.
- It should support data fragmentation and reassembly.
- It should support a security scheme that compensates the wireless medium publicity.
- It should support power management capabilities.

3.3.3.1.   IEEE 802.11 modes and frame formats

The IEEE 802.11 MAC defines two types of medium access modes that can be used in an alternating manner. These types are: [STALLING97B]

? **Contention Period (CP)** which is required to be supported by all stations in the IEEE 802.11 WLAN.

? **Contention Free Period (CFP)** which is an option that can be supported by some stations in an IEEE 802.11 WLAN.

The IEEE 802.11 MAC defines three types of frames. Each type uses the fields of the *header* part of the standard MAC frame format, shown in Fig 3-11, in a different way. The three types of frames used by the MAC are:

? **The management frame** which is used for station association and dissociation with an AP, timing and synchronization, and authentication and de-authentication.

? **The control frame** which is used for handshaking in the CP, positive acknowledgment during the CP, and to end the CFP.

? **The data frame** which is used for the transmission of data in both the CP and the CFP, and can be combined with polling and acknowledgment during the CFP.

Bytes:

2	2	6	6	6	2	6	0-2312	4
Frame Control	Duration ID	Addr 1	Addr 2	Addr 3	Sequence Control	Addr 4	Frame Body	CRC

← 802.11 MAC Header →

Bits:

2	2	4	1	1	1	1	1	1	1	1
Protocol Version	Type	SubType	To DS	From DS	More Frag	Retry	Pwr Mgt	More Data	WEP	Rsvd

← Frame Control field →
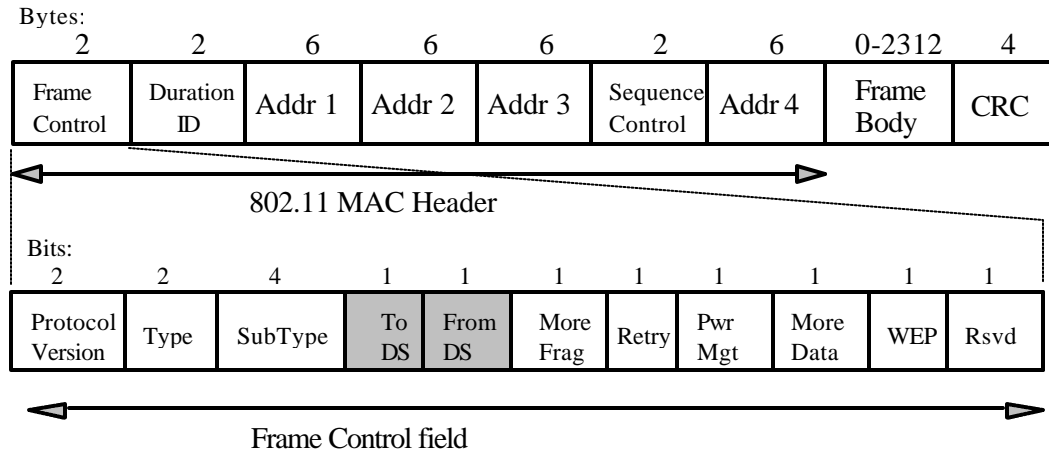
- *Frame Control* is detailed above.
- *Duration ID* is used to indicate the time in microseconds the channel will be allocated for successful transmission of an MPDU.
- *Addr1, Addr2, Addr3,* and *Addr4* are addressing fields that have different uses according to the "To DS" and "From DS" fields in the "Frame Control".
- *Sequence Control* is used to resolve duplicate packets received due to an error in the acknowledgment mechanism.
- *Frame Body* is the actual MSDU plus 7 bytes for encryption and decryption mechanism.
- *CRC* is a 32 bit CRC sequence used for error control of the whole frame.
- *Protocol Version* is reserved for use when other versions of the IEEE 802.11 protocol are implemented.
- *Type* is used to indicate the frame type. There are three types; namely, control, management, and data frames.
- *SubType* is used for further identification of the frame type e.g. Clear To Send, and Request To Send.
- *To DS* and *From Ds* are used together to specify the meanings of the addressing fields. The To DS indicates that the frame is transmitted to the Distribution System, and the From DS indicates that the frame is received from the Distribution System.
- *More Frag* is used to indicate that this is not the last fragment of a packet, i.e. not end of MSDU.
- *Retry* is used to identify that the frame is a retrial sent of an old corrupted frame.
- *Pwr Mgt* is used to specify whether power management is used or not.
- *More Data* is used to indicate that there are other data frames following this frame, i.e. not end of session.

Fig 3-11 IEEE 802.11 MAC standard frame format

To complete the description of the MAC frame format a table that indicates the meanings of the addressing fields for different values of the "To DS" and the "From DS" fields is given below (Table 3-1).

To DS	From DS	Addr 1	Addr 2	Addr 3	Addr 4
0	0	DA	SA	BSSID	-
0	1	DA	BSSID	SA	-
1	0	BSSID	SA	DA	-
1	1	RA	TA	DA	SA

Table 3-1 Different usage of the address fields

Where:

DA means Destination Address and indicates the final destination of the packet.

SA means Source Address and indicates the original source of the packet.

RA means Receiver Address and indicates the next receiver of the packet.

TA means Transmitter Address and indicates the last transmitter of the packet.

BSSID means Basic Service Set ID and is Physically the MAC address of the AP of the BSS. It is used for defining the BSS.

From studying Table 3-1 it can be seen that we have the following transmission cases according to the binary values of [To DS, From DS]:

? [0,0] the packet is transmitted in the same BSS.
? [0,1] the packet is received in a BSS from the DS system.
? [1,0] the packet is transmitted from a BSS to the DS system.
? [1,1] the data is transmitted through the DS from one AP to another.

From the usage of the addressing fields in the above cases it can be seen that the *Addr 1* field is always used for MAC filtering of packets since it always contains the intended receiver's address. On the other hand, the *Addr 2* field always contains the address of the last transmitter of the packet. The *Addr 3* field depends on the values of the To DS and From DS fields. Finally, the *Addr 4* field is only used in the case of transmitting the packet through the DS system to identify the address of the first source of the data.

In order to support both CFP and CP access schemes, the MAC defines two types of access schemes. The first is called Distributed Coordination Function (DCF), and the second is called Point Coordination Function (PCF). These two access schemes are going to be explained in details in the following.

3.3.3.2. Distributed Coordination Function (DCF) access scheme

This is a contention-based access scheme (some times called Best Effort access) [CROW97]. The scheme uses a modified version of Carrier Sense Multiple Access schemes called Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). The CSMA with Collision Detection (CSMA/CD) technique is not used, because it requires a listen while talking function to be performed by the transmitting node to detect collisions. This function will not be applicable in wireless media, because the signal of the transmitter near its antenna will be very much higher in power than that received from any other station. This is because of the high dynamic nature of wireless media regarding signal attenuation. As the name of the technique implies, the CSMA/CA depends on avoiding the collision as much as possible instead of detecting its occurrence.

The CSMA/CA scheme goes as follows (Fig 3-12): [CROW97]

1. A station wishing to transmit senses the medium to see if there are any other transmissions taking place.
2. If the medium is found busy the *Defer* process is performed.
3. If the medium is found free for a period larger than the inter-frame space duration, the station starts transmission.
4. The station waits for an acknowledgement from the destination to know if the transmission is successful or not.
5. If acknowledgement is not received after a pre-specified time-out, collision is assumed and i (where i is the retrials counter) is incremented then the *Defer* process is repeated.
6. If acknowledgement is received the station switches to listening mode.

Accordingly, the collision is detected, as well as any other transmission errors, via the acknowledgment scheme.

The *Defer* process is as follows:

1. The station defers its transmission.
2. The station presets a waiting counter to a random integer value in the range $[0,2^{2+i}]$ with a uniform random distribution.
3. The waiting counter starts counting down when the medium is free for more than the Inter Frame Space, and stops counting whenever the medium becomes busy.
4. When the counter reaches zero transmission starts.
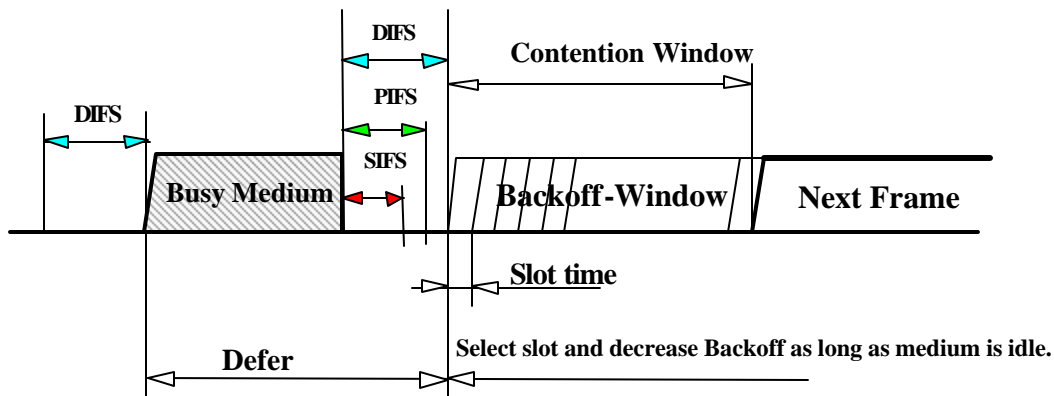
Fig 3-12 CSMA/CA channel events

What should be mentioned here is that the random back-off time is defined to be with uniform distribution with an upper limit called *Contention Window (CW)*. This CW doubles its value (with an upper bound) each time the packet transmission is unsuccessful, thus reducing the probability of collision in an exponential way. Also to allow for priority assignments when needed, an Inter Frame Space[4] (IFS) is used. This IFS can have one of three different values (Fig 3-12). The highest priority packet takes the shorter IFS called Short IFS (SIFS). The middle one is called Point Coordination Function IFS (PIFS), while the longest (i.e. the lower priority) is called Distributed Coordination Function IFS (DIFS). The DIFS is used for data transmission, while the SIFS is used for acknowledgment and other control functions such as the handshaking packets. Finally, the PIFS is used in the case of PCF access.

One more characteristic of the CSMA/CA technique used for WLAN is that the time is slotted, because it produces better performance in terms of throughput, as it results in reducing the probability of collision.

The main problem that faces this scheme is the so called *Hidden Terminal* problem. This problem in fact exists in all CSMA schemes when used for WLAN. The Hidden Terminal problem arises when two stations that are not in radio contact with each other try to transmit to a third station that is in radio contact with both of them. In this case, neither of the two

---

[4] Inter Frame Space intervals are mandatory periods of idle time between frames in the transmission medium

transmitting stations is capable of detecting the signal of the other one, which means that they might transmit simultaneously causing a collision at the receiving station end (Fig 3-13). This scenario means that the Carrier Sensing mechanism has failed in this particular case, which means that if this case is repeated many times in the network layout, performance may degrade to an unacceptable level.

To avoid the hidden terminal problem arising in CSMA in case of usage with WLAN, a handshaking technique in the form of a Request To Send (RTS) and a Clear To Send (CTS) packets is used. On the other hand, as this technique adds an overhead to the traffic, it is specified in the standards that it should not be used for short packet transmissions. Accordingly, an extra parameter must be specified for the CSMA/CA when handshaking is used. This parameter is called *RTS_Threshold*. This parameter represents the maximum length of MPDU that can be transmitted without using handshaking, i.e. if the MPDU length reaches this value or higher handshaking will be used.
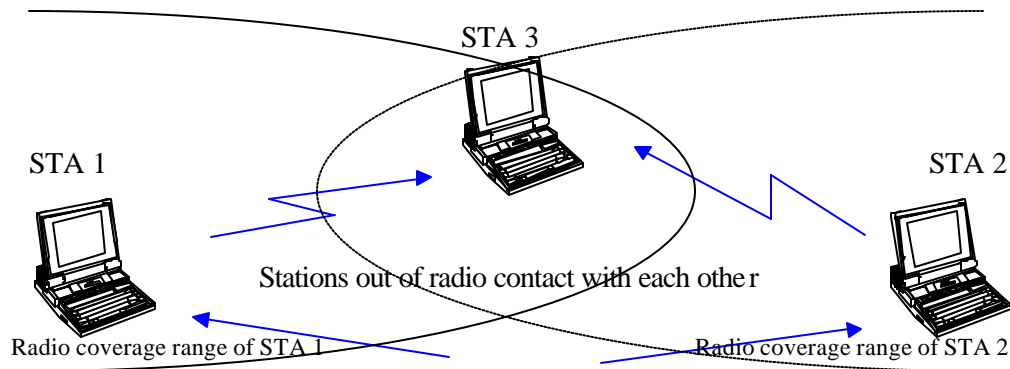
Fig 3-13 Hidden terminal problem on WLAN using CSMA

There is also another scheme that is used in association with handshaking to overcome the Hidden Terminal problem. This scheme is called *Virtual Carrier Sensing.* To allow for virtual carrier sensing, any station that is transmitting sends the duration of its MPDU in the handshaking and data packets. This duration information is used by all other listening stations to update their Network Allocation Vector (NAV). The NAV is an indicator of the amount of time that must elapse until the current transmission session is complete and the channel can be

sensed again for idle status. Virtual Carrier Sensing is performed by deferring any transmission whenever the NAV of the station wishing to transmit is nonzero. Accordingly, for a transmission to start from a station four conditions must exist. These conditions are respectively:

1. Back off random time is equal to zero.
2. Virtual Carrier Sensing is performed by making sure that NAV is zero.
3. Physical Carrier Sensing is performed which sometimes called Clear Channel Assessment (CCA).
4. Handshaking is completed with the destination informing the sender that it is ready to receive data.

Then there could be two modes of transmissions in the DCF access scheme. The first one is used for short packets and called CSMA/CA with Acknowledgment. This technique has been described before in this section and its channel event is shown with the acknowledgment (Ack) frame event, and NAV technique added to the scenario in Fig 3-14. The second is used for longer packets (equal to or greater than *RTS_Threshold*) and is called CSMA/CA with acknowledgement and handshaking (RTS/CTS).
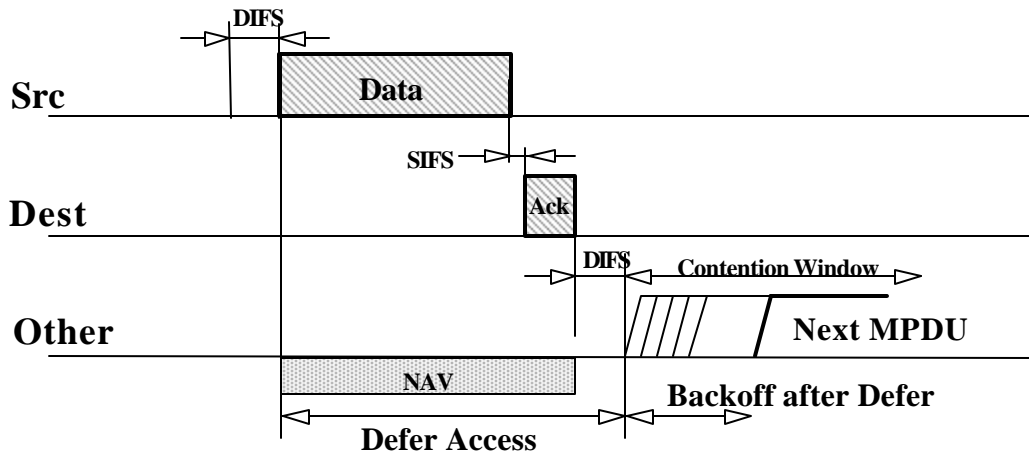
Fig 3-14 CSMA/CA with acknowledge

The CSMA/CA with RTS/CTS goes as follows (Fig 3-15):

1. The RTS control frame is first transmitted by the source station (after successfully contending for the channel) with a data or management frame queued for transmission to a specified destination station.

2. All stations in the BSS, hearing the RTS packet, read the duration field and set their NAV accordingly.
3. The destination station responds to the RTS packet with a CTS packet after an SIFS idle period has elapsed.
4. Stations hearing the CTS packet look at the duration field and again update their NAV.
5. Upon successful reception of the CTS packet, the source station is virtually assured that the medium is stable and reserved for its transmission. Hence it sends its data or management packet after an SIFS idle period is elapsed.
6. Again, stations hearing the data or management packet look at the duration field and once more update their NAV.
7. Upon a successful reception of the data or management frame by the destination station, it transmits an acknowledgement (Ack) frame after an SIFS idle period is elapsed.
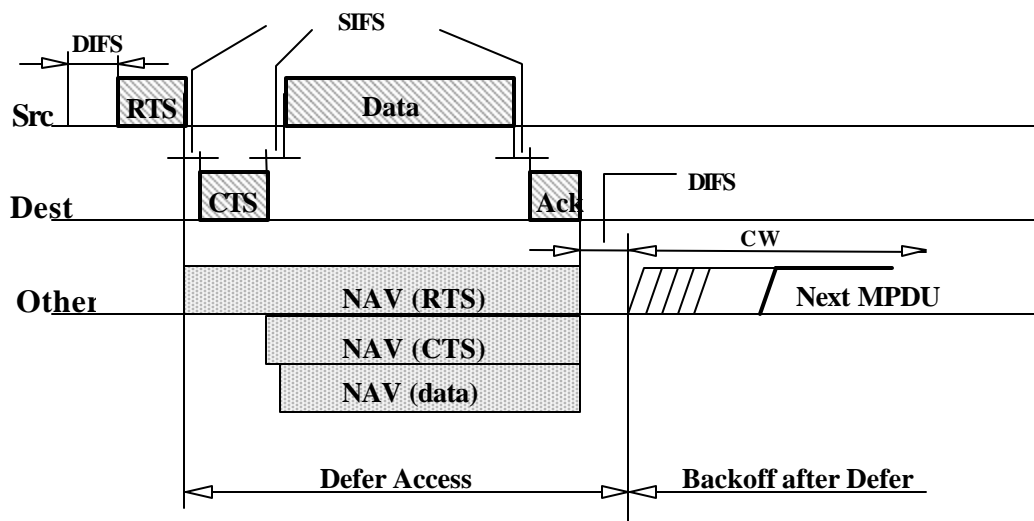
Fig 3-15 CSMA/CA with RTS/CTS and acknowledge

From the above description it can be seen that the handshaking technique, not only minimizes the possibility of collisions due to hidden terminal problem, but also reduces the capacity of the channel wasted when collision occurs. This is because most of the probable collision situations left will occur with a RTS or a CTS packets. The RTS packet is 20 bytes long and the CTS packet is 14 bytes long. These packets are very much small compared to the data packet, which might reach up to 2346 bytes. This means that the capacity of the channel wasted during collision is reduced.

One more thing that must be mentioned here is the ability of the MAC layer in the IEEE 802.11 to support *Data Fragmentation.* Data Fragmentation is used to break down large MPDUs to smaller ones, so as to increase transmission reliability. The fragmentation function is performed as follows (Fig 3-16):
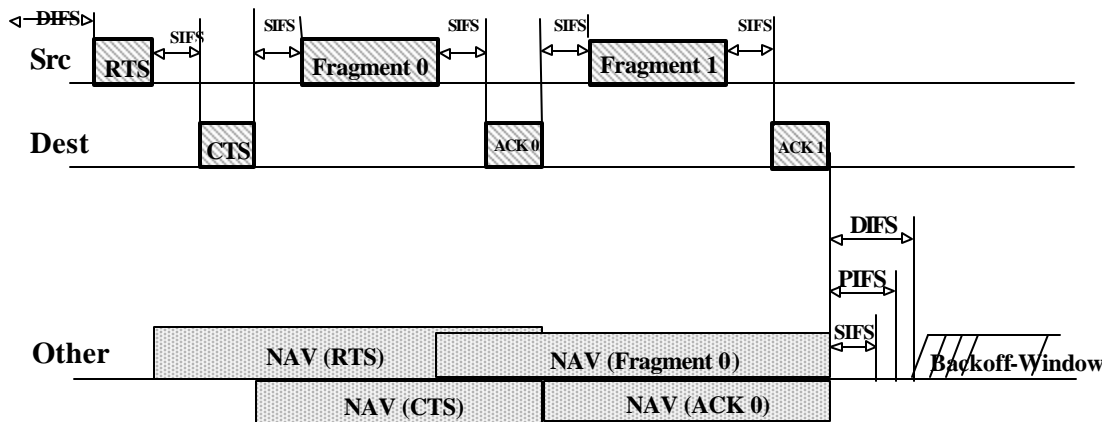
Fig 3-16 Fragmentation in CSMA/CA with RTS/CTS

1. When a Large MSDU is handed down from the Logical Link Control (LLC) layer to the MAC layer its contained in an MPDU and the length of the resulting MPDU is checked against a manageable parameter named *Fragmentation_Threshold*.

2. If the MPDU length is greater than the Fragmentation_Threshold, the MSDU associated with it is broken into multiple fragments.

3. Each fragment is included in an MPDU resulting in MPDUs of lengths equal to Fragmentation_Threshold, except for the last fragment, which results in an MPDU of length less than or equal to Fragmentation_Threshold.

4. All fragments are transmitted sequentially and the medium is not released till the end of the last fragment or if the source does not receive an acknowledgement for a fragment.

5. The destination acknowledges each fragment independently with an Ack frame.

6. The source station sends the fragment, then the destination upon receiving it sends an Ack after and SIFS idle time. The source upon the reception of the Ack for the previous fragment sends the next fragment after an SIFS idle period. The use of the SIFS here is the reason why the source station continues to hold the medium. This goes on till the end of the Ack of the last fragment.

7. If a fragment is not acknowledged, the source station stops the transmission and starts contending for the medium again using the DIFS idle period.

8. When the source station is granted the medium again, it starts its transmission with the last unacknowledged frame.

9. If RTS/CTS is used it is only sent before the first fragment, and contains duration for NAV update of this fragment transmission and its acknowledgement only. Other NAV updates are performed using the duration fields in the next fragments and their acknowledgments in the following manner: The duration field of the first fragment contains the time for the transmission of its acknowledgement plus the next frame and the acknowledgement of the next frame. The duration field of any next fragment contains the transmission time of its successor fragment plus the acknowledgement of that successor fragment. Finally, the duration field of the last fragment contains the transmission of its acknowledge only.

From the above discussion of the DCF access mechanism, it can be seen that this mechanism, as well as any other contention based MAC scheme, has the advantage of allowing fair access share to the medium for all stations in the network. On the other hand, it suffers from not supporting wireless access to backbone networks, e.g. wireless station access to a wired LAN server. It also suffers from not supporting Time Bounded Services (TBS) such as video and voice transmissions, because the delay over the network does not have an upper limit. To support these services the other scheme, called PCF is used.

3.3.3.3. Point Coordination Function (PCF) access scheme

This is a contention free access scheme that can optionally be used by some or all of the stations in an IEEE 802.11 WLAN [CROW97]. The scheme is a connection-oriented scheme. It depends on a centralized access technique that uses a station called the Point Coordinator (PC), which has priority control of the medium. The PC station has the ability to allow only one station to have priority access to the medium in a contention free scheme, by polling stations in a certain order allowing only the polled station to transmit, if it has packets to send. All other non-polled stations are not allowed to transmit and are required to remain silent. The function of the PC is performed by the AP within each BSS. Stations within a BSS capable of operating in a Contention Free Period (CFP) are called *CF-aware* stations. The method by which the polling tables are maintained and polling sequence is determined, is left to the individual vendors or producers of WLANs.

The PCF access method is required to coexist with the DCF with the following rules:

- The PCF logically sets on top of the DCF (Fig 3-17).
- The CFP repetition period (CFP_Rate) is used to determine the frequency with which the PCF occurs.
- Within a repetition interval, a portion of the time is allocated for the contention free traffic, and the remainder is provided for contention based traffic (Fig 3-18).
- The CFP and CP alternation is under the control of the PC.
- The CFP repetition interval is initiated by a Beacon Frame and ended by the Beacon Frame starting the next interval (Fig 3-19). The Beacon Frame is transmitted by the AP and besides starting the CFP, Beacon Frames helps in maintaining synchronization and timing in the BSS.
- The CFP starts just after the Beacon frame and ends by the transmission of CF-End frame, also generated by the AP.
- The duration of the CFP repetition interval is a manageable parameter that is an integer number of the Beacon Frame duration.
- The maximum size of the CFP is manageable and called CFP_Max_Duration.
- The minimum value for CFP_Max_Duration is the time required to transmit two maximum size MPDUs including overhead, the initial Beacon Frame, and a CF-End frame.
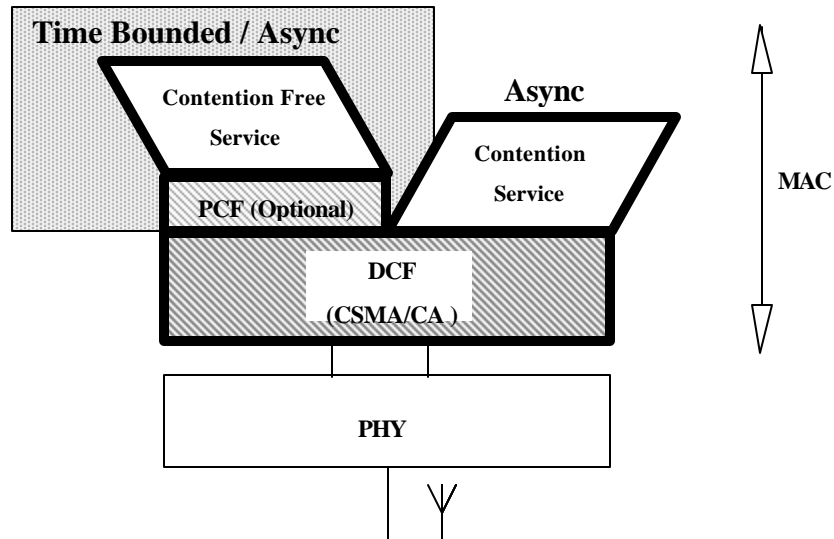
Fig 3-17 The use of optional PCF with DCF
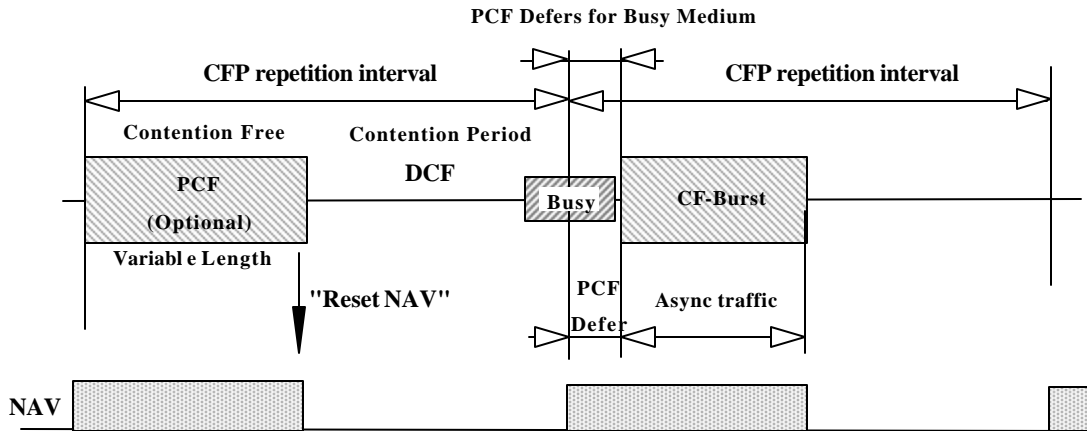
Fig 3-18 Alternation between CFP and CP
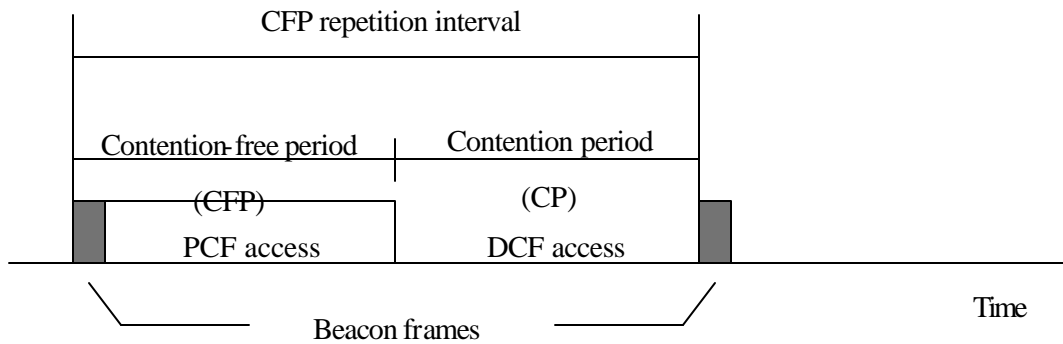
Fig 3-19 CFP repetition intervals and beacon frames

- The maximum value for CFP_Max_Duration is the CFP repetition interval minus the time required to successfully transmit a maximum size MPDU during the CP including the time for RTS, CTS, and Ack. Therefore, time must be allocated for at least one MPDU transmission using the CP.

- The AP can determine the relative lengths of the CFP and CP in each CFP repetition interval individually according to the network traffic characteristics.

- Both the DCF and the PCF periods defer to each other if the medium is not free. This means that the CFP repetition interval starts are not at constant duration.

68

- At the nominal beginning of each CFP repetition interval, all stations in the BSS update their NAV to the CFP_Max_Duration. This means that any contention traffic is prevented. This lasts till the last PCF transfer (CFP-End), which resets the NAV allowing contention based traffic to start.
- During the CFP, the only time a station is allowed to transmit is in response to a poll frame from the AP, or to transmit an Ack, an SIFS interval after receiving an MPDU from another station.

From the above description it can be seen that the CFP although not occurring periodically, it guarantees that its starting time delay will not be more than the time required to transmit one MPDU in the CP. This time is equal to the time needed to transmit RTS, CTS, maximum MPDU, and Ack plus cumulative IFSs. This means that this techniques can allow for time bounded services, as a maximum delay limit can be achieved for the transmission packets of a certain session if it is assigned a fixed slot location in the CFP.

The operation of the IEEE 802.11 MAC in the PCF mode can be one of three scenarios: Station to AP communication, station to station communication, and AP to a non CF-aware station.

1. Station to Access Point communication (Fig 3-20):
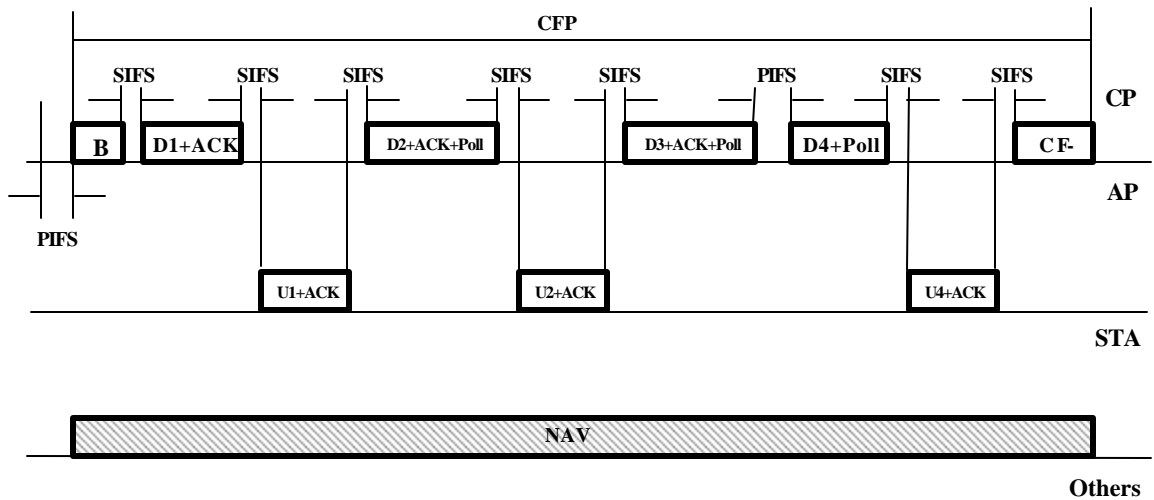
Fig 3-20 Station to AP scenario

69

At the nominal beginning of the CFP, the PC (Point Coordinator), which is actually the Access Point (AP), senses the medium. If the medium is free for a PIFS period, the PC transmits a *Beacon Frame* to initiate a CFP.

The PC starts CF transmission an SIFS after the *Beacon Frame* is transmitted by sending a *CF-Poll* (no data), *Data*, or *Data plus CF-Poll* frame. It must be noted that if the PC has no traffic buffered and the network is lightly loaded, the PC can terminate immediately the CFP by sending a *CF-End* frame.

When a CF-aware station receives a *CF-Poll* frame, it responds after an SIFS interval with either a *CF-ACK* or with a *Data plus CF-ACK* frame when it has data to send.

When the PC receives a *Data plus CF-ACK* from a station, it can send a *Data plus CF-ACK plus CF-Poll* to another station. The *CF-ACK* part of the frame is used to acknowledge the receipt of the previous *Data* frame.

If the PC sends a *CF-Poll* frame to a station and it has no data to send it replies with a *CF-ACK* frame only.

If the PC fails to receive a *CF-ACK* for a transmitted data frame, the PC waits a PIFS interval then continues transmitting to the next station in the polling list.

2. Station to station communication (Fig 3-21):
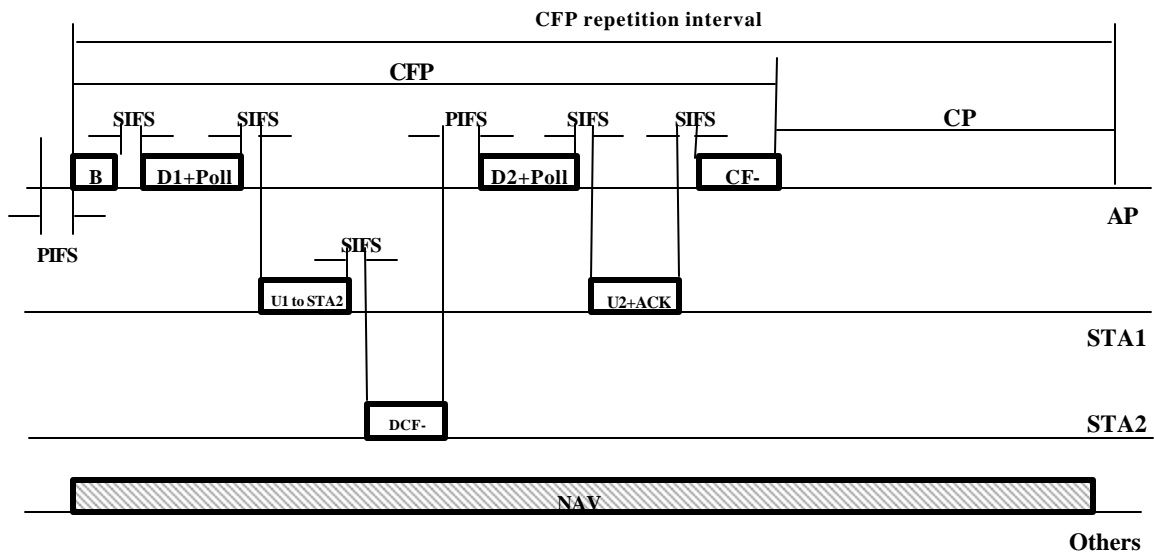
Fig 3-21 Station to station scenario

After receiving a *CF-Poll* frame from the PC, a station may choose to communicate directly to another station in the same BSS. In this case it will directly send a *Data* frame to the other station.

When the destination receives the *Data* frame, it responds with a *DCF-ACK*, after an SIFS idle interval, to the sending station. The PC waits a PIFS interval after the *DCF-ACK* before transmitting any further frames.

3. AP to non CF-aware station:
The PC may also choose to send a *Data* frame to a non CF-aware station. The receiving station waits an SIFS interval after successfully receiving the frame then sends a *DCF-ACK* to the PC.

Finally, the PCF is also capable of performing fragmentation in the same manner described before for the DCF. It must be noted that the receiving station, in both cases of DCF and PCF, is responsible for the re-assembly process.

3.3.4. Additional specifications in IEEE 802.11 standard

Power management is one of the responsibilities of the IEEE 802.11 MAC. It is indeed beyond the scope of this thesis, but details of it can be found in [DIEPSTRATEN96], and [LAMAIRE96]. In brief, the standard includes two types of specifications. One for the case of ad-hoc topology, while the other is in the case of infrastructure topology. For the ad-hoc topology if a station is in power-saving mode it wakes up for only short predefined periods of time to hear if there is data sent to it, in which case it exits from the power-saving mode and starts to receive. In the case of the infrastructure topology, the AP stores frames for stations that are in power-saving mode. The station that is in the power-saving mode wakes up periodically to listen to selected Beacons sent by the AP. These Beacons contains information about the queued traffic in the AP. If the station finds that there are data queued for it in the AP, it sends a special polling frame to the AP informing it that it may send the data.

Security and authentication is another important point to be mentioned here. It is out of the scope of this thesis and can be found in the standards documentation, but a brief description of it will follow [AMUNDSEN96].

The IEEE 802.11 specifies an optional data encryption algorithm. The algorithm aims at achieving a Wired Equivalent Privacy (WEP). This algorithm is only for station to station privacy and not for end to end privacy and only implements confidentiality function. The algorithm is based on an encryption technique called RC4 PRNG algorithm, which is defined by RSA Data Security, Inc [LAMAIRE96].

On the other hand, for authentication the IEEE 802.11 adopts two mechanisms. Only one of these mechanisms is defined in the standards. This mechanism depends on an *open* or *shared key* authentication. As the name suggests, this mechanism depends on that the authentication of a station/user is based on the communicating stations having knowledge of a shared secret key.

The other mechanism is a proprietary mechanism that depends on the WLAN manufacturer. This means that there are no rules set for this mechanism and it differs from one vendor to another.

Mobility and roaming is also an essential function that is required to be performed by the MAC layer. In brief to allow for mobility a continuous connection must be maintained between the moving Station (STA) and its basic Service Set (BSS). This is guaranteed, without any extra additions to the already described MAC functions, in both ad-hoc and infrastructure LANs when the STA is moving within the coverage range of its BSS.

In the case of ad-hoc networks, STA is not allowed to move outside of its BSS range and cellular architecture is not achievable.

On the other hand, in the case of infrastructure networks the concept of ESS allows for multiple BSS to be joined in one network. The standard allows for any STA to move between different BSSs in the same ESS without losing connection to the ESS. To allow for this the standards defines a Scanning and an Association-Reassociation-Disassociation mechanism that goes as follows:

1. STA is initially associated with a certain BSS by sending an Association request packet at the nominal start of operation.
2. Upon moving away from its BSS, STA can actively scan available channels by sending a probe packet on different channels, or it can passively scan the channels by listening to transmission activities in different channels.
3. When STA finds the current channel, it sends a Reassociation request packet to the AP.

4. If the AP receives the Reassociation request packet successfully, it sends a Reassociation response packet to the STA and roaming is successfully performed.

5. The DS configuration table is updated by the new BSS of this STA.

6. The previous BSS is informed by the Disassociation of the STA from it via the DS.

7. If the STA receives no Reassociation responses within a time-out it continues scanning next channels.

3.4. ETSI HIPERLAN Wireless LAN Standards

The High Performance LAN (HIPERLAN) group started its work in standards for wireless LANs in 1991. This group was formed by the European Telecommunications Standards Institute (ETSI).

Unlike the IEEE 802.11 standards, this committee was not driven by available products in the market. In this case, a set of required functions were first defined and then the committee set the rules to satisfy these functions [LAMAIRE96].

The standards here was mainly concerned with achieving high data rates, and minimum power consumption. This led to a limited range (10-100 m).

The standards allows for multi-hop routing which was prohibited by the IEEE 802.11. It also allows for time-bounded services, and has a part for power management.

3.4.1. Architecture

HIPERLAN is designed to work without the need for any infrastructure but it can support connection to infrastructure networks. The only supported WLAN topology is the ad-hoc topology, i.e. there are no centralized architecture [PAHLAVAN97]. Since one of the major requirements from a WLAN is the ability to support cellular operations to allow for range extension, then the ad-hoc technology alone would not be sufficient for such a requirement. To achieve this requirement, the HIPERLAN committee has adopted the so called *Multi Hop Routing*.

In brief, *Multi Hop Routing* is the ability of a station to act as a relay for the packet sent from a source station to another destination station (Fig 3-22). The operation can be described as follows:

If a station (STA1) wants to communicate with a station (STA2) that is a distance d from it, and d is larger than the coverage range of STA1, then:

- STA1 will choose a station STA3 which is in radio contact with it and possibly in radio contact with STA2 and will send the data to it with an information telling STA3 that the data is not destined to it.
- If STA3 is in radio contact with STA2 it will forward the data to it.
- If STA3 is not in radio contact with STA2 it will repeat the previous operation with a new station STA4 and so on till the data reaches STA2.
- ? To allow for an adaptive operation, the packet holds its routing information, which is kept in each station for faster resolving of future paths.
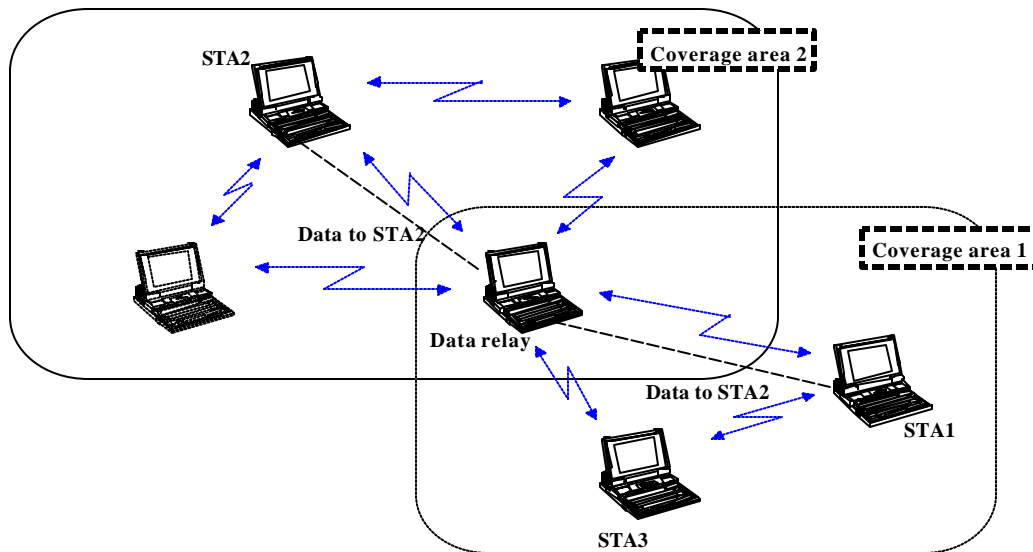
Fig 3-22 Mlti-hop routing and architecture of HIPERLAN

3.4.2.   Physical layer

The high data rates required by the HIPERLAN committee needs a reasonably wide spectrum in the range of 150 MHz in bandwidth. Unlike IEEE 802.11 physical specification which supports more than one physical medium, the HIPERLAN standards support only one physical layer which has the following characteristics [LAMAIRE96]:

- The transmission medium used is narrow band RF transmission.
- The current HIPERLAN standards operate in the RF frequency band range 5.15-5.30 GHz. with intentions for future usage of the band 17.1-17.2 GHz.

- The operating band is divided into five adjacent channels each with bandwidth equal to 23.529 MHz. The center frequency of the first channel is at 5.176468 GHz, and separation between center frequencies are equal to 23.5294 MHz (Fig 3-23).
- The operating bit rate is 23.529 Mb/s for transmitting data and training sequences.
- The bit rate for transmitting physical layer headers and for acknowledgments is 1.4706 Mb/s [WEINMILLER97].
- The modulation technique used is Gaussian Minimum Shift Keying (GMSK) which reduces the adjacent channels interference.
- To overcome Inter Symbol Interference (ISI) resulting from multi-path dispersion Decision Feedback Equalizer (DFE) is used at the receiver with a training sequence of 450 bits for every data packet.
- The channel coding scheme used is a BCH (31,26) interleaved across 16 code words. This means that for 16 code words each of 26 bits, i.e. a block of 416 user bits, data is transmitted using 16 code words each of 31 bits, i.e. a block of 496 bits.
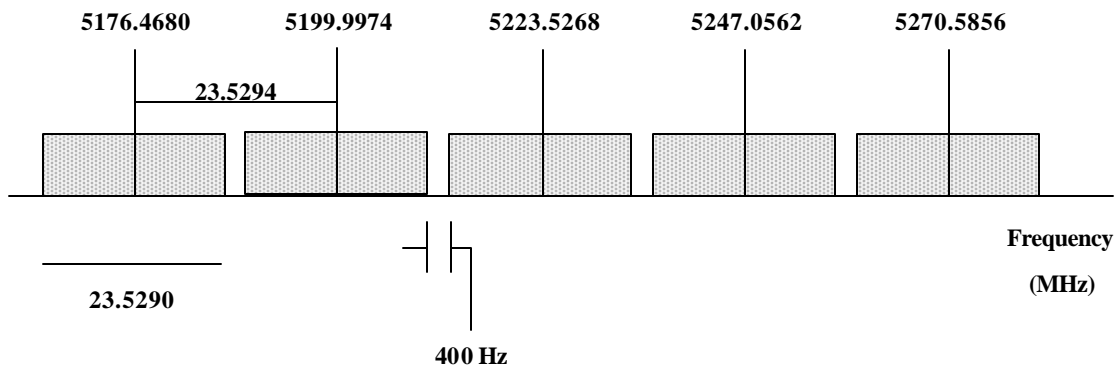
Fig 3-23 Frequency bands for the 5.15 GHz band of HIPERLAN

3.4.3. Medium Access Control (MAC)

The MAC is also based on CSMA but with different specifications than these for the IEEE 802.11. The scheme used is called Non Preemptive Multiple Access (NPMA) [WEINNILLER97] and [LAMAIRE96]. The scheme has the advantage of lower probability of collisions, but it suffers from longer delays due to its complexity. The scheme goes as follows (Fig 3-24):

Prioritization phase	Elimination phase	Yield phase	
Has 1-5 slots	Has $n+1$ slots, n<=12	Has m slots, m<15	
Slot is 256 bits long	Slot is 256 bits long	Slot is 256 bits long	

| Transmission ends, followed by one synchronization slot of 256 | For each contending user, the probability of transmitting a burst of i slots, $i<12$, is 0.5^{i+1} | Each contending user defers transmission for j slots, $j<14$, with a | Transmission starts. |

Fig 3-24 Channel access for HIPERLAN

- The medium is sensed for a relatively long period (specified to be 1700 bit period).
- If the medium is free transmission starts immediately.
- If the medium is busy, the contending nodes enter a contention resolving mode consisting of three phases. The aim of the phases is to reduce the number of contending stations till one of them is granted the access to the medium. These phases are respectively *prioritization*, *elimination*, and *yield* phases
- In the prioritization phase the aim is to allow only nodes with the highest priority packets to continue contending for the medium. This is done by having a slotted period, and making each node having a packet with priority p to transmit a burst (special sequence or only the carrier) at slot $p+1$ as long as it did not hear a transmission from a higher priority node. After the end of the first burst on the channel this phase ends and the elimination phase starts.
- In the elimination phase nodes that succeeded in transmitting a burst in the previous phase now contend for the channel. This contention is resolved by allowing the nodes to transmit a burst for a geometrically distributed number of slots and listening to the channel in only one slot. If a node hears in the listening slot burst from another node, it stops contending for the channel. This leaves only a small number of nodes contending for the channel. These nodes are those with the longest burst assuming that there is no hidden terminal problem. At the end of the longest burst this phase ends and the yield phase starts.
- In the yield phase the remaining nodes in contention defer their transmissions for a random geometrically distributed number of slots, while listening to the channel. If a transmission is heard all nodes defer their transmissions. The result of this scheme is that the probability of actual collisions is reduced to the range of 3 percent.

As it can be seen from the previous discussion that the elimination phase brings the number of contenders to a small number and then in the yield phase only one of these is granted access to the channel. This reduces the chances of actual collisions.

3.4.4.   Additional specifications in the HIPERLAN standard

Quality of Service for packet delivery (QoS) is another thing that is supported by the HIPERLAN standards. The QoS is supported via two schemes. Parameters for both of these schemes are provided by the application using the HIPERLAN specifications.

The two parameters controlling this process are the *packet priority*, and the *packet life-time*. These two parameters are used to determine the channel access priority for the packet. It is very important to note that in contrary to the IEEE 802.11, there is no other explicit mechanism (as the time-bounded mechanism) used to achieve the required QoS.

As multi-hop routing is supported here both the priority and life-time of the packet are transmitted with it. If a packet is not delivered to its destination during its life-time it is immediately discarded. We can see from the previous argument that although the HIPERLAN specifications was aiming at a statistically independent transmission rate for different traffic characteristics, this was not achieved because of the need to support multi-hop routing and ad-hoc topologies. An important issue to mention here is that as the multi-hop routing is supported; nodes are required to keep a data base of the routing information. This is also specified in the standards with the addition that multi-hop routing is optional for all the nodes. This means that if a packet arrives at a node and needs to be routed to another station, the node has the choice to either forward it to the other station or discard it and the sender will have to try again later.

Power management methods are specified here also. Two mechanisms are supported for power saving:

- In the first mechanism a node announces that it only listens during short periodical times keeping most of its circuits powered down at the remaining times. Other nodes wishing to transmit to this node will have to wait till the time this node wakes up. These nodes know this time from the broadcast made by the node entering the power save mode. For the case of broadcast packets nodes who wish to broadcast or multi-cast schedules its

broadcasting times allowing others to enter power-saving mode unless when they expect to hear a broadcast.

- The second method for power saving is through the use of the innovative two-speed transmission method. In this method packets have a short, low bit rate header (LBR) with 1.4706 Mb/s rate which informs the node whether it needs to listen to the packet or not. In this way even though the node is listening it can keep the error correction, the equalization, and other circuits powered off until it is informed by the LBR that there is a packet for it. Hence the station powers up all its circuits.

Encryption and authentication are also functions of the MAC layer of the HIPERLAN standards. While the standards keep away from describing an encryption technique, it describes methods for informing different station of the encryption key used. The standard defines a set of such keys and how they can be stored in the nodes. These specifications are not available now, as the work in security is not completed yet.

A global look at the situation here will certainly result in some interesting questions that need to be answered. First, about channel selection: How will all of the nodes belonging to the same logical LAN decide on a common channel? Second, the effect of CSMA: What will be the effect of hidden terminals, and how the throughput will be affected? Third, as all of the nodes use only one channel: What will be the achievable user traffic density especially for applications requiring limited delays?

These questions could only be answered through building prototype systems as suggested in [LAMAIRE96].

3.5. WINFORUM Frequency Etiquette Standards

WINFORUM was initiated by Apple Computer to form an alliance in the industry to obtain a frequency band from the FCC for the data-PCS [PAHLAVAN95]. The objective of this standard is to achieve fair use of frequency-time resources of channels which are unlicensed and used by different applications. In order to achieve this objective they are developing the so-called "*spectrum etiquette*". The etiquette does not intend to preclude any common air interface standard. The etiquette demands the use of the well known listen-before-talk (LBT). This means that any device does not have the permission to transmit unless the spectrum it occupies is not in use by any other device within the same range. An access time is specified for each device after which it should free the channel for other devices to use it. Power is also

required to be limited to keep ranges in short distances. Both the power and connection time are related to the bandwidth used to equalize the interference and keep fair access for of the channel.

WINFORUM standard defines two classes of information generation, the *asynchronous* and the *isochronous* transmissions. The asynchronous transmission is bursty, begins transmission within milliseconds, uses short bursts that contain a large amount of data, and releases the channel quickly. An example of this is the wireless LAN nodes. The standard allows the sub-bands for this transmission ranging from 50 kHz to 10 MHz. On the other hand, the isochronous transmission holds the channel for a long time. It is a periodic transmission, and has flexible link access times that may be extended up to a second. An example for this is the wireless PBX network. The transmission band may be divided into 1.25 MHz segments. The WINFORUM standards states that these two types are technically contrasting and cannot share the same spectrum.

3.6.   Conclusions

In this chapter, we have given a detailed overview of the two main standards used for wireless LANs; namely the IEEE 802.11 and the ETSI HIPERLAN standards. In this overview, we covered both the physical layer and medium access control techniques used in these standards. To conclude the chapter we gave an example of other standardization organizations that work in coordination with the two WLAN standardization organizations. This organization is the WINFORUM. Such organizations as WINFORUM, FCC, DECT and others are necessary to coordinate with WLAN standards organizations to guarantee wireless medium reliability.

From this overview we found that:

Both 802.11 and HIPERLAN standards are actually complementary to each others with their current situation. This can be seen by noting that while HIPERLAN supports multi-hop routing, 802.11 supports access point scheme. On the other hand, HIPERLAN works in the 5 GHz band and 802.11 works in the 2.4 GHz band. Also, HIPERLAN works with narrow band transmission while 802.11 works with spread spectrum transmission. Finally, 802.11 supports IR transmission and HIPERLAN does not.

Both types of standards have not yet reached the mature level for applications, as can be seen by visiting their web sites. Currently, the two organizations are cooperating to guarantee compatibility between products supporting both standards. This effort is still not yet completed.

It is worth noting that the IEEE 802.11 is having now two new sub-committees to study higher data rate support in the RF (2.4 GHz and 5 GHz). Another sub-committee is formed to study what they call Home-RF using Shared Wireless Access Protocol–Cordless Access (SWAP-CA) [LANSFORD99] which works in establishing a standard for home wireless networks.

It was noticed during the study of the standards that there is a lack of enough information about the IR physical layer of 802.11 and actually there are very few products in the market supporting this technology, while most wireless vendors focus on the RF technology and specially the spread spectrum supported by IEEE 802.11.

Chapter 4

Analysis of CSMA/CA Using Single Station Superposition (SSS)

Chapter 4
Analysis of CSMA/CA Using Single Station Superposition (SSS)

4.1.   Introduction

In this chapter we introduce an analytical mathematical model to represent the performance of a famous wireless LAN MAC protocol called Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). The mathematical model of such a technique is very difficult to achieve using direct Markov models, because of the existence of the random exponential back-off algorithm in this protocol. This algorithm complicates the Markov analysis, as it imposes a huge number of states in the Markov model of the network. This makes the model mathematically intractable. In order to overcome this problem two steps are taken. The first step is to simplify the Markov model itself by using a closed queuing model to represent the protocol instead of a direct multi-queue model, while the second step is to use an approximate technique to solve the Markov model and get the performance equations.

After establishing the mathematical equations that represent the performance parameters of the protocol numerical methods are applied to solve these equations and get numerical results. In the numerical results the effect of changing arrival rates and number of users on normalized throughput and packet delay will be demonstrated. The effect of changing the back-off algorithm parameters such as the number of retrials on a packet transmission before discarding it on the throughput and delay will also be plotted.

In order to achieve the above goals the chapter will start by a detailed description of the CSMA/CA MAC technique. Afterwards, an analysis for this technique is performed using a technique called Single Station Superposition which is described in [WOODWARD91]. Although this MAC technique is widely used in WLAN products it was found that most of the work done on this technique was based on simulation such as in [WEINMILLER97]. Even the few analytical materials on the subject were avoiding to include a major behavior of the technique which is the *random exponential back-off* such as [CHHAYA96] where it is mentioned that the back-off process is avoided to simplify the model.

Before starting to describe the MAC scheme some major definitions must be given first. These definitions are necessary to be able to evaluate the performance of a MAC in a quantitative way.

4.2. <u>Performance Evaluation Parameters for WLAN</u>

To be able to give an analytical performance evaluation for WLANs, the parameters according to which the performance is evaluated must be defined first.

The major parameters that are used to evaluate any data communication system including WLAN system are the *message throughput S* and the *message delay D*. These two parameters are going to be referred to as *throughput* and *delay* omitting the *message* part. These parameters are defined as follows:

Throughput *S* is the total number of successfully transmitted messages, from the source to the destination per unit time, divided by the channel capacity, calculated in number of messages per unit time, between the source and the destination. This parameter is sometimes referred to as *normalized throughput*, because its value could never be greater than one.

Delay *D* is the total number of time units taken by a message from its arrival at the source buffer, till it reaches its destination. This includes the queuing time and the transmission time of the message.

In the course of calculating these parameters, some terms and expressions will be used. Two of these terms are very important to be stated here. These two important terms are the o*ffered load G* and *Little's formula*.

Offered load *G* is the total number of messages arriving at the source per unit time, divided by the channel capacity, calculated in number of messages per unit time, at the source.

Little's formula is a mathematical formula that relates the *messages arrival rate at the queue input l*, the *number of messages waiting in the queue* called *queue length l*, and the *waiting time in the queue w*. The Little's formula states that:

$l = l \cdot w$

This formula holds even if the *number of messages being currently serviced* is included in *l*. However, it must be noted that in this case, *w* will include both *waiting time in the queue* and *service time at the server*.

One more thing to be mentioned here is that in the analysis that will follow, some general assumptions are made for all the models. These assumptions are as follows:

1. All network nodes are assumed to be in Radio contact with each other in case of ad-hoc architecture, and with the base station in case of infra-structure architecture. This means that the effects of hidden terminal problems and captures are not considered in the given analysis.

2. The effect of channel errors, such as fading effects, white Gaussian noises, and other channel error sources, on the performance are all assumed to be independent of the network state. This means that the effect of these parameters on the transmission performance can be modeled by introducing a constant term called *probability of no error* P_{nerr}. This term is multiplied by the probability of successful transmission for the scheme, assuming ideal channel, to get the overall performance.

3. The number of nodes (customers) in the network is assumed to be finite and equal to a value called *U*. This number does not include the base station in case of infra-structure networks, as the base station is chosen to be dedicated to traffic control only, i.e. it is not treated as a node itself.

4. Packets are assumed to be of fixed size.

4.3. CSMA/CA Protocol

CSMA/CA is a variation of CSMA/CD used for Ethernet. The problem with CSMA/CD is that bandwidth efficient collision detection in radio channels is difficult to achieve. This inefficiency is a result of the high dynamic attenuation of radio signals. This high attenuation makes it practically very difficult for a radio transceiver to listen to other signals while transmitting. This listening while talking procedure is essential for the collision detection part of CSMA/CD. To be able to overcome this problem and still achieving an acceptable performance, a technique that minimizes the probability of collision is used. Accordingly, in this technique the collision detection part is replaced by a collision avoidance part [STALLING97A]. The CSMA/CA has many variations. These variations differ only in the collision avoidance algorithm. Two of these techniques are the one used for IEEE 802.11 distributed access, and the one used for HIPERLAN access. Here, the protocol used for IEEE 802.11 will be described in details as it is adopted in a product existing in the market. This product is the WaveLAN RF WLAN of AT&T [TUCH93].

Due to its random access nature, CSMA/CA is robust with respect to protocol level, bandwidth sharing, and radio channel characteristics. It has the advantage of high reliability in terms of network survival as it is a distributed protocol i.e. no central hub (central point of failure) exists. It has also the advantage of ease of reconfiguration, because it is a broadcast technique which makes it easy to insert a station in the network or remove it from the network during the network operation.

4.3.1. Network architecture

As already stated, CSMA/CA is a distributed access scheme. This means that the scheme relies on a broadcast data flow configuration. Consequently, the network architecture for this scheme is of ad-hoc nature, and the connection between stations is a peer-to-peer connection. This means that all stations are equivalent in terms of access functions that they can perform (Fig 4-1).
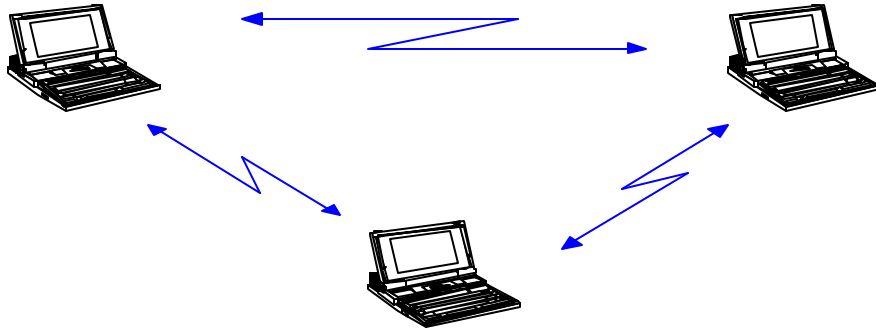
Fig 4-1 CSMA/CA network architecture

4.3.2. Protocol operation description

The protocol only uses one frequency and single spreading code for transmitting from any station to others. This means that it uses single channel access. The technique goes as follows (Fig 4-2):
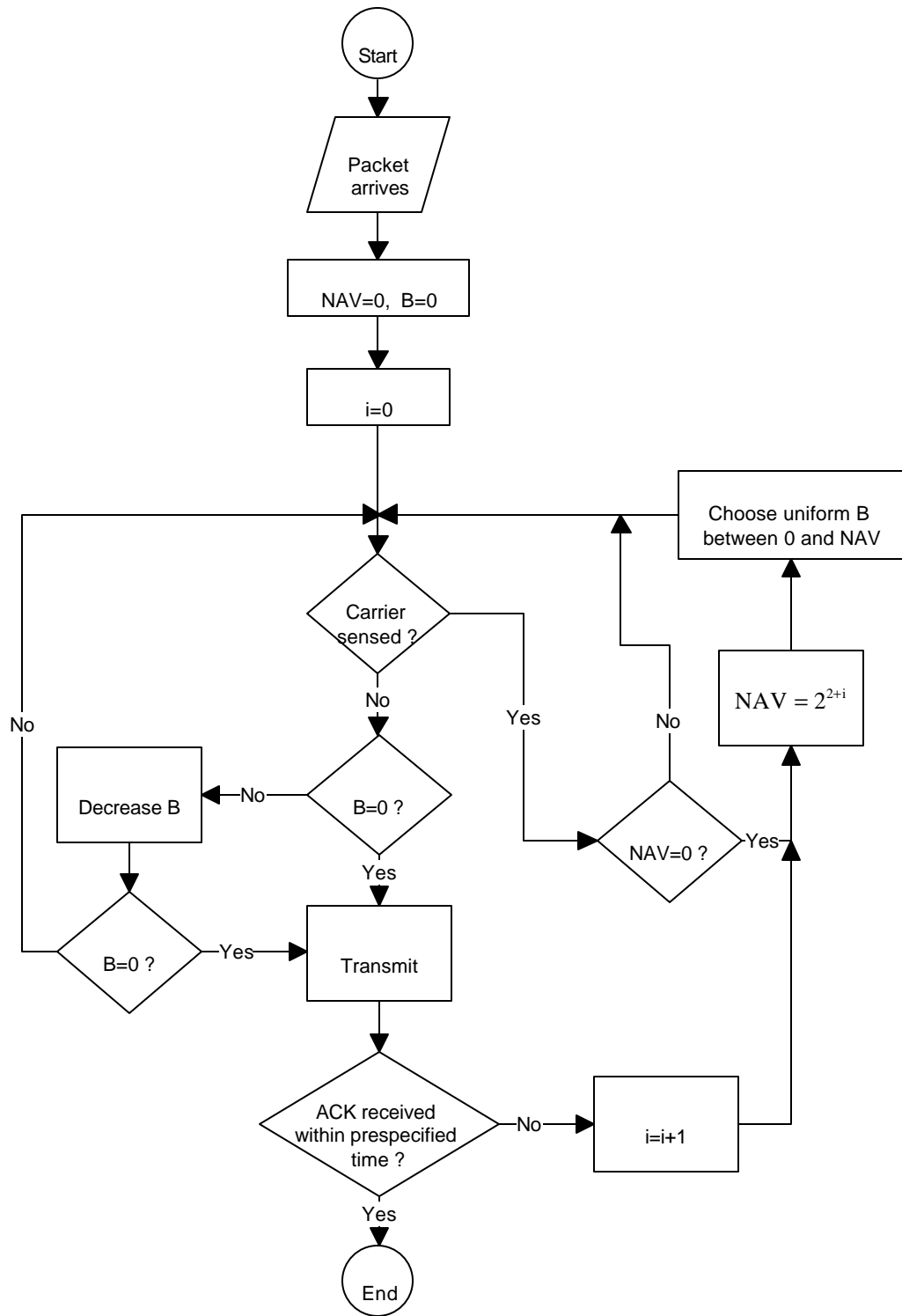
Fig 4-2 Flow chart for the MAC procedure of CSMA/CA

When a message arrives at a station, the station performs the listen before talk function by listening to the medium to see if any transmission is in progress. If the medium is found idle, the station starts its transmission immediately and continues till the end of the packet even if collision happens. On the other hand, if the medium is found busy, the station defers the transmission for a random period, called Network Allocation Vector (NAV), after which it retries transmission again. The random waiting period value is only decreased when the medium is free. This means that while deferring the transmission, the station listens to the medium. If the medium is found busy, the station stops decreasing the waiting period until the medium becomes free again. When this happens, the deferring period is decreased again.

A collision is detected, as well as any other communication error, via the acknowledge scheme. This means that if a collision happens, the receiving station will either not receive the packet correctly, or not receive it at all. Consequently, it will return either NACK, or no ACK at all. In both cases, the sending station will retransmit the packet again.

The scheme is a time slotted scheme, which means that new events can take place only at starts of slots. This time slotting nature increases the throughput of the scheme by reducing the probability of collisions.

In order to allow for more efficient collision avoidance scheme, the random number chosen for deference is chosen from a different range of values, whenever a collision or any other transmission error happens. This means that the probability of retransmission is different than the probability of transmitting for the first time. The random value is taken to be of binary exponential nature, and is rounded to be an integer number of time slots. The range from which the random defer waiting time is chosen is taken to be $[0,\ 2^{2+i}]$, where i is increased when a transmission fails.

Another important issue that must be considered when using carrier-sensing schemes with radio medium is that there is no guarantee that an intended transmitter is in radio contact with its receiver. Hence CSMA/CA although ensures that only one node gains access to the medium, it doesn't guarantee that the intended receiver will be able to get the data sent by the transmitter. Therefore, an additional handshaking procedure must be used to guarantee that the two partners are in radio contact. This addition also avoids the well-known *hidden terminal* problem that appears in CSMA protocols in case of radio channels. This problem is

that two stations may be trying to connect to a third one and both are in radio contact to that third station, but are out of range with respect to each other. This will result in a situation where these terminals will not sense carriers from each other and hence may send at the same time to the third station, which will result in an increase in the collisions. This problem will be avoided in the following handshaking procedure by using the receiver busy packet generated from the intended receiver if it is not ready to receive. This handshaking procedure proceeds as follows (Fig 4-3):

Whenever a Work Station (WS) needs to send a frame, it first sends a short *Request To Send (RTS)* control message to the destination WS using CSMA/CA. At the same time it starts a timer count (referred to as [1] in Fig 4-3) to allow for error recovery via timeout process. The RTS control message contains the address of both the source and the destination. On receipt of this, provided that the intended destination receives the request and is ready to receive a frame, it broadcasts a *Clear To Send (CTS)* reply message with the same pair of addresses, but with their order reversed within it. Alternatively, if the destination is not prepared to receive data, it returns a *Receive Busy (RxBUSY)* reply. This RxBUSY is the one used for avoiding hidden terminal problem. If the reply is positive then the transmitting WS transmits the *data frame*, and restarts the timeout timer. If the data frame is received correctly, the receiving WS returns an *acknowledgment ACK* message to the transmitting WS. Upon receiving this acknowledgement, the transmitting WS stops the timeout timer count (referred to as [2] in Fig 4-3). On the other hand, if the frame is not correctly received due to collision or transmission error, a *negative acknowledge NACK* is returned by the receiving WS. Upon receipt of a NACK the transmitting node retransmits the data again and restarts the timeout timer. These retries are repeated up to a predefined number of times then stops assuming loss of radio contact. It must be noted that all the transmissions here are using CSMA/CA techniques, even for control packets.
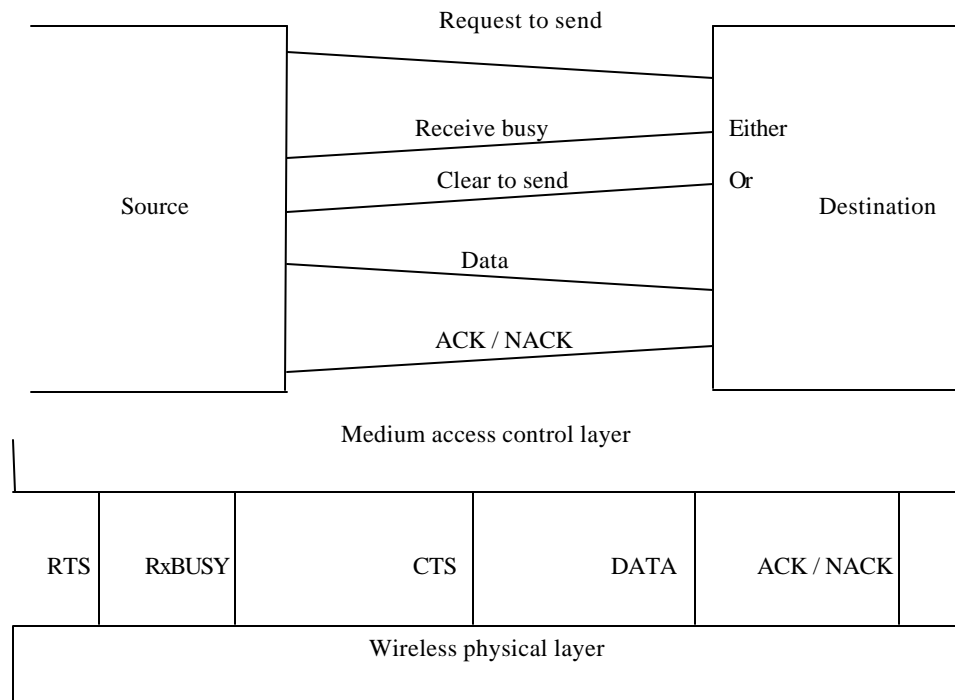
(a) Time sequence diagram of channel events

Primary (Source)   **Secondary (Destination)**
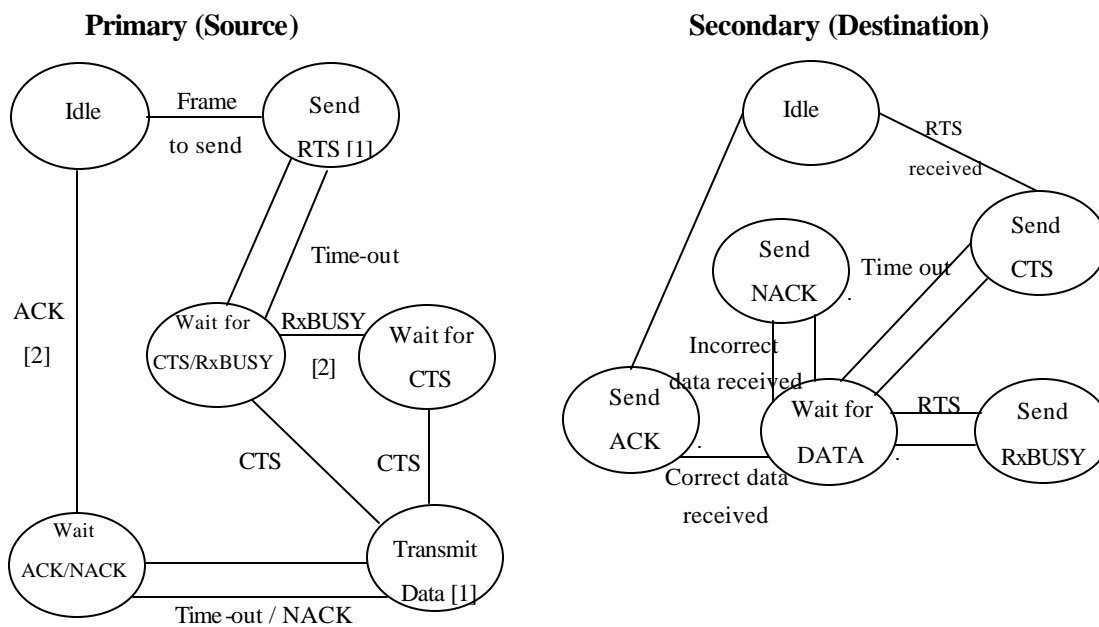
[1] = Start timer

[2] = Stop timer

(b) Source state transition diagram   (C) Destination state transition diagram

Fig 4-3 Handshaking technique used in CSMA/CA protocol

Another addition to make this technique more efficient, is the use of different inter-frame spaces. What is meant by inter-frame spaces is mandatory periods of silence on the medium between frames during which no station can issue a transmission. This concept was described before in Chapter Three in details. These spaces are given two different values. The shorter one is used to separate the CTS signal from the RTS, and to separate the data packet from the CTS. While the longer one is used to separate different complete transmissions (data plus handshaking associated with it) from each other. This technique results in minimizing the probability of hidden terminal effects, by ensuring that collisions due to this problem can take place only with RTS packets. These packets are much shorter than data packets. Hence, the overall performance is enhanced.

It was found from simulation analysis performed over CSMA/CA protocols that the addition of the handshaking technique does not increase the throughput in a significant way in case of having all stations in radio contact with each other (no hidden terminal problem). On the other hand, the use of this technique obviously increases the transmission delay of messages. Hence, in case of network configuration where all nodes are in radio contact with each other, CSMA/CA is used solely without applying handshaking. In the contrary, in case of having network configurations where hidden terminal problem may occur, the use of handshaking does imply a significant enhancement to the overall message throughput. This increase in the throughput compensates the relatively slight increase in the message delay. Hence, handshaking is applied in these configurations.

Finally, a phenomenon that also occurs in CSMA when used with radio transmission is the so-called *capture effect*. In brief, this phenomenon is the ability of a radio receiver to extract and receive correctly one message from a collision situation. This means that although more than one transmission can occur simultaneously, the receiver might be capable of receiving one of these transmissions correctly. Again, this phenomenon happens because of the dynamic nature of radio channels, and also because of the adaptive gain control in radio receivers.

It must be noted that in the following analysis we assume complete radio contact between stations. This means that there is no hidden terminal problem. Since there are no hidden terminal problem happening, handshaking will not be used, as it will complicate the analysis while not improving the performance as mentioned before.

4.3.3. Performance evaluation

An analysis for CSMA/CA that includes random back-off in the model will be given here using Markov chains and closed queuing modeling technique. The solution of the model for the steady state parameters is made by an approximation technique called Single Station Superposition (SSS). A description of this technique can be found in [WOODWARD91].

In order to be able to analyze the access technique, both a state transition diagram (Fig 4-4) and an open queuing model (Fig 4-5) of the access scheme will be given. These diagrams are of illustrative purposes. The use of direct open queuing models to find the analytical results of this technique will be mathematically intractable. To give an idea about the problem, it can be seen from Fig 4-5 that the queues are not connected directly to the server. On the contrary, the connection to the server is channel state, and queue state dependent. Hence direct cascading or parallelism between queues cannot be assumed. Also, the analysis of each queue separately depends on the channel state which depends on the other queues. All these factors make it very difficult to analyze this model mathematically. The open queuing model of this scheme only contains a main server that represents the CSMA/CA channel, the queues of the network stations, and a triggered back-off timer. The triggered back-off timer gives a logic '0' on the output as long as it is counting down till its value reaches zero, then it gives a logic '1' at the output. The timer is preset to a randomly selected value whenever its queue is not empty and an input trigger is given to it by the channel state input in case of busy for the first time or a collision happens. The timer then starts counting down whenever the channel state is not busy till it reaches zero in which case it closes the switch connecting the queue to the server. It must also be noted that the service here is not completed until the acknowledge of the transmitted message is returned to the transmitting node. This means that the service time includes the time waiting for acknowledgment and the time for transmitting acknowledgment.
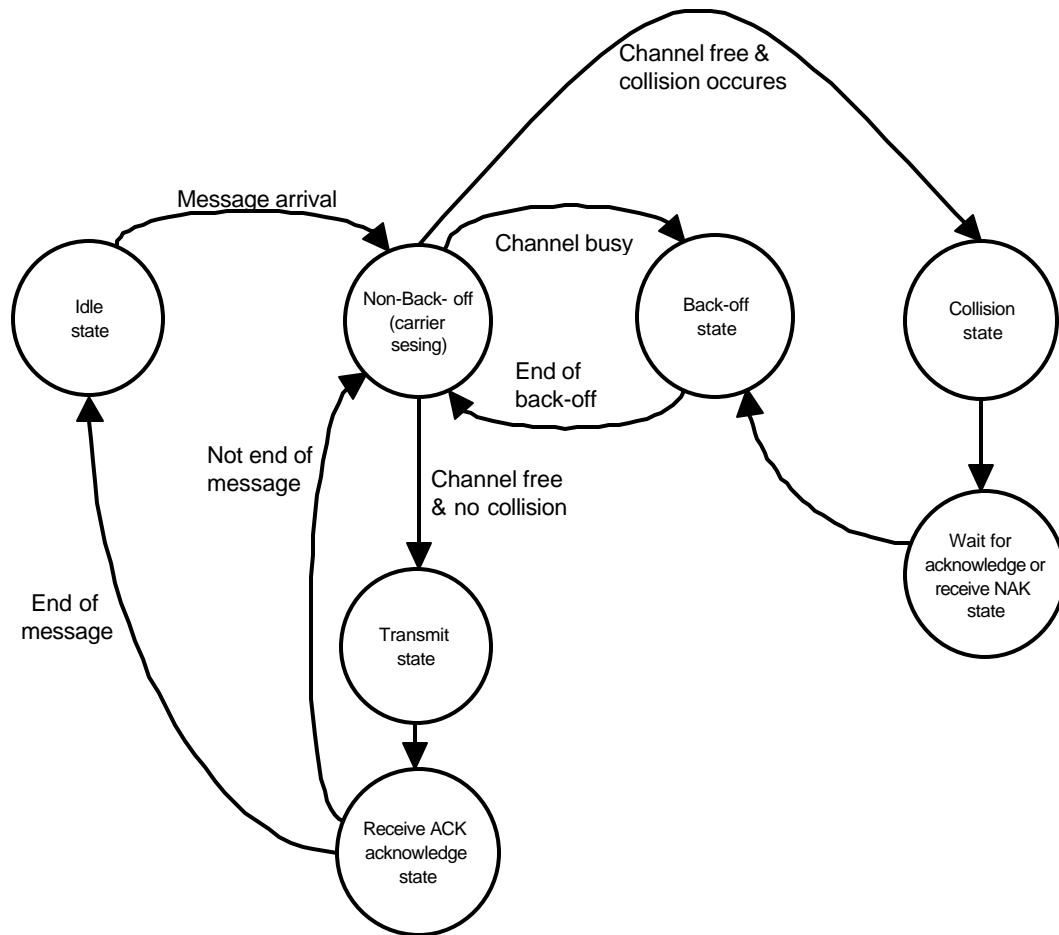
Fig 4-4 State transition diagram for CSMA/CA

Station queue

Buffer length=1 msg.

Active stations

(U-I)

Channel busy

state

Channel

.
.
.

Queue

occupation state

Idle stations

(I)

Triggered

back-off timer
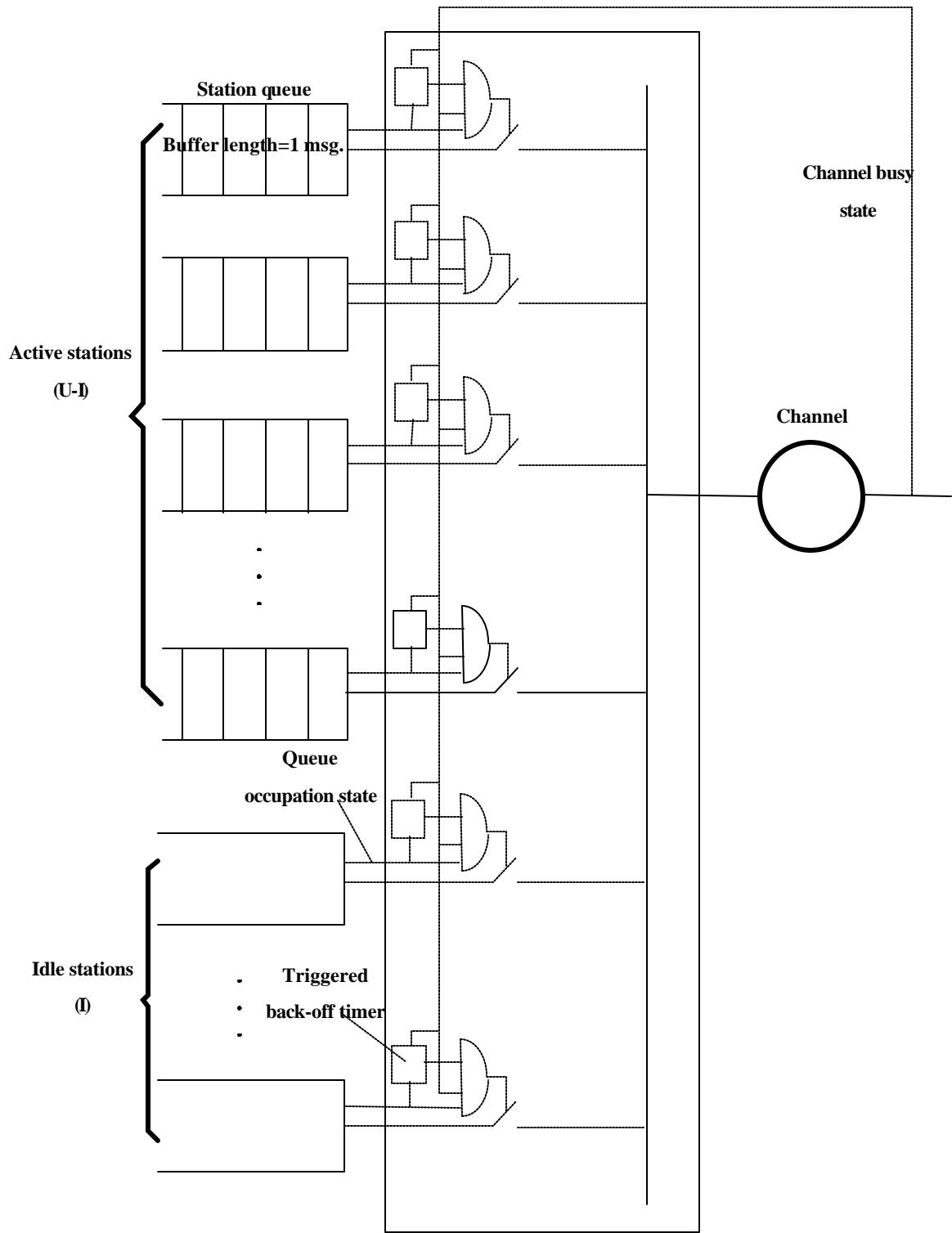
Fig 4-5 Open queuing model for CSMA/CA

The state transition diagram (Fig 4-4) of this technique is more descriptive as it shows all the possible stages a station passes through till it succeeds in transmission. Using this state representation, the closed queuing Markov model is derived. The closed queuing model is more mathematically tractable, and is constructed for one station with all possible states for this station.

A closed queuing Markov model for the system is derived by assigning a state for each distribution of users through all the possible states for any user. For example if the possible states for any user are idle (i), transmitting (t), and colliding (c), and the number of users in the system is (U), then the state of the system $X(I,T,C)$ is the case where I users are idle, T users are transmitting, and C users are colliding. Since, $I+T+C=U$ then, I and T are enough to represent the system and the states vector becomes $X(I,T)$. In this case we will have a two-dimensional Markov chain. If we represent the system as the description above, we will have a Markov chain where the nodes (queues) of this chain are all the possible states for any user and then let the nodes (queues) be occupied by the users according to the current state of the system. In other words, for the specific case described above we will have three nodes (queues) in the Markov chain idle state (i), transmitting state (t), and collision state (c). These nodes are queues occupied by the users of the system according to the state of each user. The state vector of the system $X(I,T)$ at any instant of time will be a snap shot of this model at this time. In this snap shot we will have I users in idle state, T users in transmit state and C users in collision state, and the condition $U=I+T+C$ holds.

In the above representation a direct solution to the Markov chain will require the calculation of state probabilities which is very difficult to achieve mathematically. One way to get around this is to use the system Markov model to represent only one station in isolation and solve this model for the state probabilities of this station. The effect of the other stations on the studied station states is represented by global parameters such as the probability of channel busy and probability of collision. After solving for one station and getting the performance parameters from the solution of the Markov model, the performance parameter for the whole network is derived by proper superposition of the same parameter for all nodes in the network. This technique for solving Markov chains for network protocols is referred to as Single Station Superposition (SSS) in [WOODWARD91] and is more general and physically

understandable than the Equilibrium Point Analysis (EPA) mentioned in [WOODWARD93], and [WOODWARD91]. The described above technique is the one used in this chapter to analyze CSMA/CA.

It must be noted that in Fig 4-4 the back-off state may perform one of two functions. These functions are in brief; assigning a random waiting time for the first trial, or assigning random waiting times for retransmission trials. The process decides which function to perform by determining from which state it entered. The function to be performed is selected in accordance with the following transitions respectively; *channel busy* which happens at carrier sensing, and *transmission occurred and no ACK received* which happens in case of collision or channel error. In the first case the new back-off assigns a back-off wait for the first time, while in the second case it recalculates the back-off wait according to number of retries.

To complete the description of this diagram it must be noticed that whenever a carrier is sensed while in back-off, the decrease of the random time stops till the medium is free. In the carrier sensing, the station remains there till the medium is free again, which means no decrease of random time will occur till medium is free.

The detailed Markov process queuing model for this protocol is given in Fig 4-6 and is based on closed queuing techniques of modeling with the following main assumptions:

? Time axis is slotted with time slot equal to t.

? Total number of users is finite and is assumed to be U.

? Idle users generate messages with probability l per slot, per user.

? Active users do not generate messages until they become idle.

? Message lengths have geometrical distributions with mean equal to T_{av} slots. Hence, the probability of reaching message end after transmitting a certain number of packets is taken to be s, where $s=l/(T_{av}/L)$, and L is the fixed packet length. This also means that a message consists on the average of $1/s$ packets.

? Each user can have at most one message waiting for transmission, i.e. user's buffer length is equal to one message.

? A user that senses the medium finds it busy with probability b.
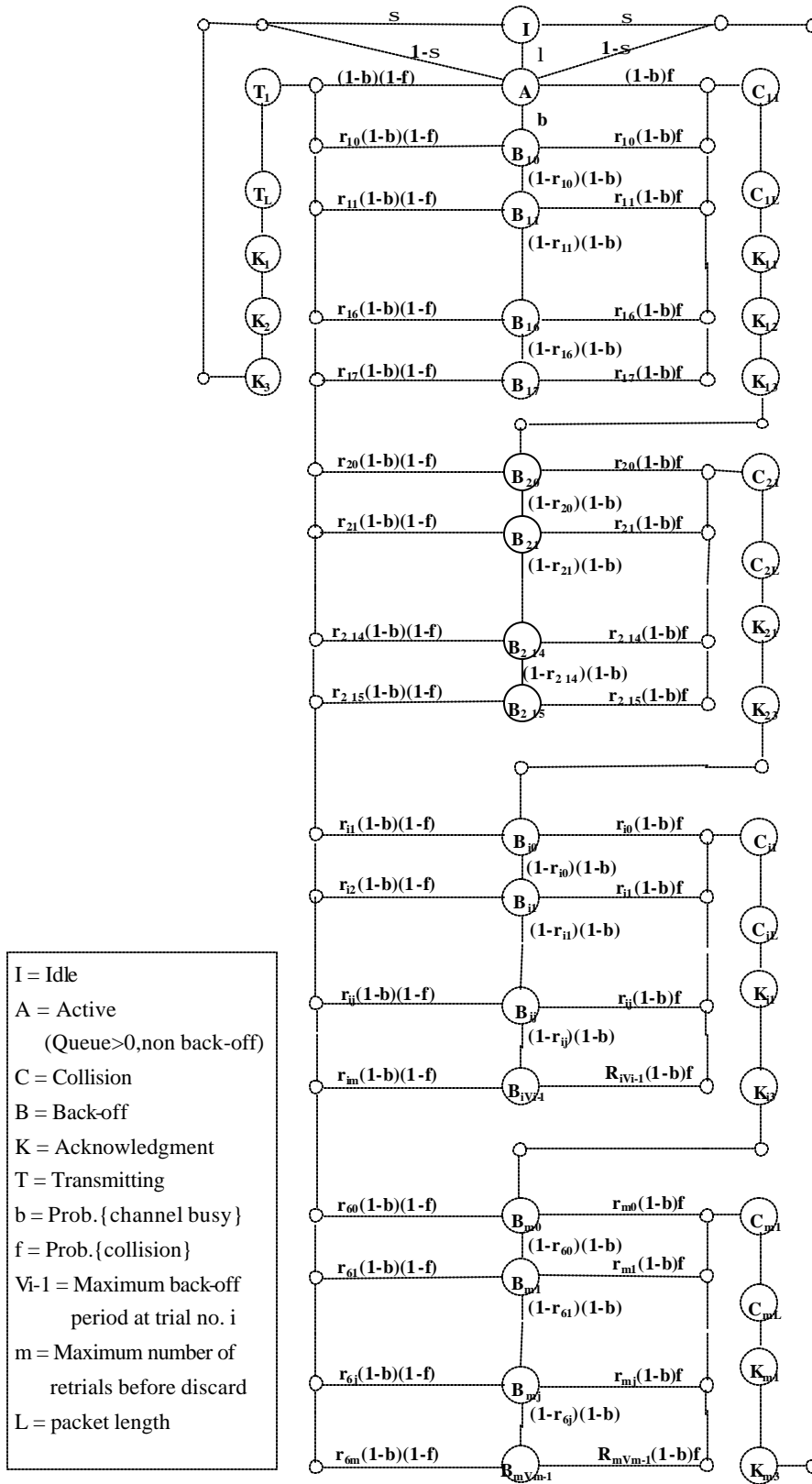
Fig 4-6 Discrete -time Markov queuing model of CSMA/CA

I = Idle

A = Active
  (Queue>0,non back-off)

C = Collision

B = Back-off

K = Acknowledgment

T = Transmitting

b = Prob.{channel busy}

f = Prob.{collision}

Vi-1 = Maximum back-off
  period at trial no. i

m = Maximum number of
  retrials before discard

L = packet length

95

- ? A user transmitting for the first time that finds the channel busy decides to wait for a random duration in the range *[0,V_1-1]* channel free slots with uniformly distributed probability density function before trying transmission again.

- ? A user coming from collision state waits for a random duration back-off in the range *[0,V_i-1]* channel free slots with uniformly distributed probability density function before trying transmission again. Where *i* is the number of transmission retrials.

- ? Collision is detected by the receiver and informed to the transmitter through acknowledgment.

- ? Acknowledgment transmission time is assumed to be 3 slots.

- ? Collision occurs with probability *f*.

- ? Packet length is constant and is assumed to be equal to *L* time slots.

- ? Channel is noise free and the only source of error is collision.

- ? Time slot is very small compared to the packet transmission time and also very small compared to mean time between arrivals for messages.

All users bebng to the same probability distribution class, i.e. the distribution of packet arrivals is the same for all users.

Considering Fig 4-6, it can be seen that nodes in queue *I* are idle, and have no messages to send. With probability *l* new messages arrive at an idle node. This node moves directly to active queue *A* just before the end of the current time slot. Then, either it finds the channel busy with probability *b* or finds it free with probability *1-b*. If the channel is found busy, the node goes into a back-off random delay operation represented by queues B_{ij}. In the back-off operation the node decides to wait for a number of time slots with channel free state (channel free slots) that is random and uniformly distributed in the range *[0,V_i-1]*, where *i* is the number of transmission retries and *1<i<m*, V_i-1 is the maximum back-off period at trial number *i* in terms of channel free slots, and *m* is maximum number of retrials on a packet before discarding it. This random property is shown in the model by the probabilities r_{ij} which are the probabilities of transmitting given that the node is at queue B_{ij} of back-off. Also, because in this technique the decrease of the random back-off counter stops when the channel becomes busy, the transition from queue B_{ij} to $B_{i\ j+1}$ includes the factor *(1-b)* which is the probability of free channel. When a node at queue *A* or queues B_{ij} finds the channel free and decides to transmit at this time, two scenarios are possible: The first one is a successful transmission, which means that this node is the only one transmitting with probability *(1-f)*. Hence, the node

goes through queues T_i and K_i which represent the time needed to transmit a packet of L time slots length and an acknowledgment of three time slots length. After a successful transmission is completed, the node either goes to I queue, or queue A with probabilities s, and 1-s respectively. Going to queue I means that the last packet in the message has been transmitted, i.e. end of message is reached. On the other hand, if more packets of the message are still in the buffer, the node moves back to queue A to continue. The other possibility is collision with probability f. This is the case when more than one node attempts to transmit at the same time. In this case, the channel becomes busy for the whole period of L slots transmissions plus three slots of acknowledgment. This waste is because collision detection is not performed and it is detected by *negative ACK*, or *no ACK at all* situations. This is represented by states C_{ij}, K_{i1}, K_{i2}, and K_{i3}. After finishing this collision state, the node goes to the start of the next back-off state, in which the same back-off procedure is performed as before, but with i increased by one. If this collisio n status is repeated m times, i.e. i reaches the value m, and no successful transmission happens, the back-off is reset and the node moves to the next packet in the message if any is remaining. This is represented by moving into state A with probability 1-s. On the other hand, if this packet was the last packet in the message (an event that has a probability equal to s) the node moves into state I.

It must be noted that in the queuing model shown in Fig 4-6, nodes at state B_{ij} stay in the same state as long as the channel is busy, i.e. with probability b. Also, the number of back-off states is V_i, where V_i is given by the expression $V_i=2^i.CW_{min}$ for a random binary exponential back-off, and CW_{min} is the minimum back-off window. Another important issue is that states B_{i0} represent the case where a station in back-off has chosen to transmit as soon as the channel becomes free.

The values of the probabilities r_{ij} of transmitting for a node given that the node is at state B_{ij} must be calculated for the model to be completed. To calculate these values, it must be noted that the probability of choosing to wait till state B_{ij} is the probability of waiting j channel free slots in the ith retransmission retry. As the waiting time is uniformly distributed, this probability is given by $1/V_i$. But this is the unconditional probability of transmitting at state B_{ij}. What is needed here is the conditional probability of transmitting given the system is at state B_{ij}. It can be easily shown (Fig 4-7) that this probability is given by $r_{ij}=1/(V_i-j)$. This expression can be

logically deduced from the shown queuing system by noting that, given that a node is at state B_{ij}, this node has only V_i-j possible scenarios.

Markov Closed Queuing Model Equilibrium Equations

The diagram given in Fig 4-6 is a closed queuing model for the users in the network. This means that each node of it is actually a queue containing a number of users being in a certain condition. Consequently, the states of the system are the vectors containing the number of users in each state.

In other words if we assume that X_n is the state of the system (state of Markov chain) at time slot n, then $X_n = \{ x^1_n, x^2_n, \ldots, x^H_n \}$ where H is the total number of criteria that are used to group users.

The model diagram given in Fig 4-6 is not a direct mapping of the protocol description given before. In fact some statistical simplifications were made to get to this model. The modifications made are mainly to the representation of the random binary exponential back off process representation in different stages. The actual and the equivalent simplified representation for a random binary exponential back off process in the case of second trial ($i=2$) is given in Fig 4-7.

(a) Direct representation model   (b) Equivalent simplified model
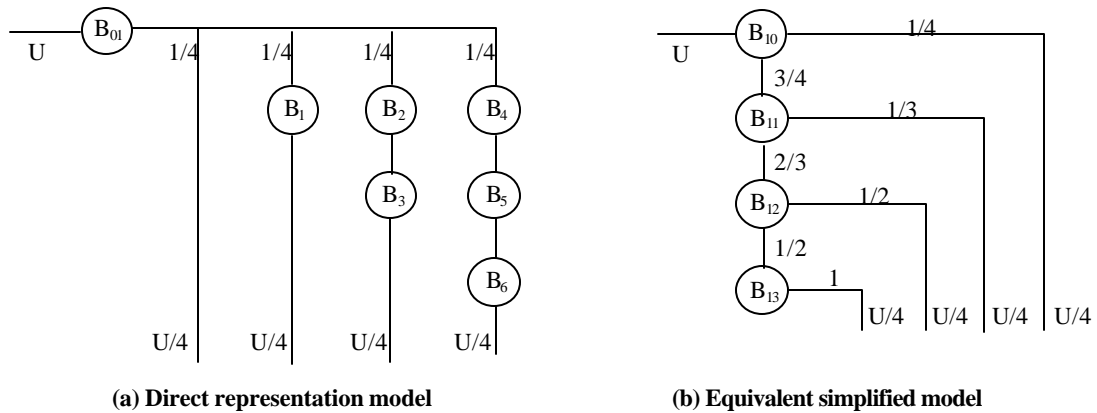
Fig 4-7 Random exponential back-off Markov model

In this case the waiting time represented by the states could be in the range $[0, 2^i{-}1]$ channel free slots with uniform distribution. This is shown in Fig 4-7a by the different possible branches each with probability $1/2^i$ that a node in back off could go through. In this model, there are 7 different states. This model can be simplified to reduce the number of states without affecting the statistical characteristics of the model. The simplified model is shown in Fig 4-7b. In this model, we have only 4 states. The statistical equivalency of the two models can be verified by noting the total probabilities across each branch of the output. These probabilities multiplied by the total number of users U at the input gives the average users flow across each branch. It can be seen that in both models this number is the same at the output of each branch. This means that the statistical characteristics at the input and the output of these two models are the same. Hence, they are statistically equivalent as long as we are not interested in the flow of users inside each one, but only interested in the users flow at the input and outputs.

The equilibrium equations for the model in Fig 4-6 are the equations based on the law of flow conservation. This law can be applied to the number of users in the queue as well as the probability of being in a queue (state). This is because both are equivalent with only a scaling factor difference. In the following we will use probabilities and solve the model for one station only then use superposition to get values for U stations. This is the so called Single Station Superposition (SSS) technique. The law of flow conservation applied to probabilities states that "At steady state the probability of entering to a state (a queue) is equal to the probability of exiting from this state (queue)". Using the label of the state to represent the probability of being in the state, the equilibrium equations are as follows:

$$B_{ij}\left(1-r_{ij}\right)=B_{i\,j+1} \qquad , \qquad 1\le i\le m, \qquad 0\le j\le V_i-2 \tag{4-1}$$

$$(1-\boldsymbol{s})K_{m3}+\boldsymbol{1}.I+(1-\boldsymbol{s})K_3 = A \tag{4-2}$$

$$\boldsymbol{s}.K_3+\boldsymbol{s}.K_{m3} = \boldsymbol{1}.I \tag{4-3}$$

$$K_{i3} = (1-b)B_{i+1\,0} \qquad , \qquad 1\le i\le m-1 \tag{4-4}$$

$$(1-b)(1-f)\left[A+\sum_{i=1}^{m}\sum_{j=0}^{V_i-1}r_{ij}.B_{ij}\right]=T_1 \tag{4-5}$$

$$(1-b)f\left[A+\sum_{j=0}^{V_1-1}r_{1j}.B_{1j}\right]=C_{11} \tag{4-6}$$

$$(1-b)f\sum_{j=0}^{V_i-1}r_{ij}.B_{ij}=C_{i1} \quad , \qquad 2\leq i\leq m \tag{4-7}$$

$$T_1=T_2=...T_i=...T_{L-1}=K_1=K_2=K_3 \tag{4-8}$$

$$C_{i1}=C_{i2}=...C_{ij}=...C_{iL-1}=K_{i1}=K_{i2}=K_{i3} \quad , \qquad 1\leq i\leq m \tag{4-9}$$

The equilibrium equation at point B_{10} will be linearly dependent on all the others, hence it is replaced by the normalizing equation which states that the sum of the probabilities of all possible states is equal to one. This is given by equation:

$$I+A+\sum_{i=1}^{m}\sum_{j=0}^{V_i-1}B_{ij}+\sum_{i=1}^{L}T_i+\sum_{i=1}^{3}K_i+\sum_{i=1}^{m}\left[\sum_{j=1}^{L}C_{ij}+\sum_{j=1}^{3}K_{ij}\right]=1 \tag{4-10}$$

The normalized throughput S can be defined as:

The probability of a slot carrying a successful transmission.

S = Prob{1 slot carrying successful transmission}

  = Prob{Slot successfully transmitting for station 1

       $\cup$ Slot successfully transmitting for station 2 ….

       $\cup$ Slot successfully transmitting for station i ….

       $\cup$ Slot successfully transmitting for station U}

  = Σ Prob{Slot successfully transmitting for station i}

  = U.Prob{One station is in successful transmission state}

The probability of a station being in a successful transmission is the sum of the probabilities of successful transmissions per station. This can be expressed from the above model by the sum of the probabilities of states T_i, which can be expressed mathematically as follows:

$$P_s=\sum_{i=1}^{L}T_i \tag{4-11}$$

where P_s is the probability of being in a successful transmission state for a station. From the definition above, the throughput has the following expression:

$$S = U.P_s \qquad (4\text{-}12)$$

The delay can also be obtained by Little's formula. This formula states that:

> $l=a.w$
>
> where l is the queue length, a is the arrival rate of customers, and w is the average waiting time in the queue (delay).

To be more careful in the above formula the throughput is used instead of the arrival rate. As in the case of systems with possibilities of retransmission and losses, the throughput includes the retransmission traffic in the arrival rate. This is more accurate, because this retransmission traffic will induce more delay in the system that should be counted for in calculating the total delay. The normalized delay (delay in terms of unit message transmission time) can be defined from above as follows:

> Delay per station (assuming no more than one packet in station at any time).
>
> D_{norm} = Avg. No. of packets in station / Rate of service per packet.
>
> $\qquad = (0 \text{ x } Pr\{\text{no packet}\} + 1 \text{ x } Pr\{\text{one packet}\}) / Pr\{\text{serving a packet}\}$
>
> $\qquad = (0 \text{ x } P(I) + 1 \text{ x } [1\text{-}P(I)]) / \Sigma \ P(T_i)$

The formula for normalized delay (in terms of unit message transmission time) according to the above definition is then as follows:

$$D_{norm} = \frac{1-I}{P_s} \qquad (4\text{-}13)$$

The definition of the normalized delay for U stations is the same as that for one station. This is because the way the delay is measured from definition is per station. What is remaining is the expression for the non-normalized delay D (in terms of time slots). This is simply the normalized delay multiplied by the packet length in time slots.

$$D = D_{norm}.L \quad = \frac{1-I}{P_s}.L \qquad (4\text{-}14)$$

It can be seen from the previous equations that the values of the equilibrium solution of these equations depend on the channel state represented by the probabilities b and f. To overcome this problem the solution of these equations will go in the following steps:

1- An expression for the collision probability f is derived from the state probabilities in the system model and the number of users U in the system.

2- A channel Markov model is proposed that represents the channel states as seen by any node that wishes to transmit.

3- From the channel model an expression for the busy probability b is derived in terms of state probabilities of system model and number of users U in the system. The need for the channel model to calculate the busy probability b is because the expression of this probability cannot be deduced directly from the states of other nodes as in the case of collision probability. This is because by definition the busy probability is the probability of one station only transmitting successfully or two or more stations are in collision. Since no single station alone can be in collision, the collision probability on the medium is not easy to calculate from the system Markov model for a single station as it is not an independent probability for each station. In fact it implies that there is at least one more station that is in a collision state too. Consequently, the use of the channel Markov model allows for counting for this independence with a simpler mathematical solution.

4- The equations for the system state probabilities are solved together to get an expression for the probability of the first state of successful transmission T_1 in terms of b, f and the system traffic and configuration parameters.

5- Both expressions for f and b are dependent on each other and on T_1.

6- Expressions for the throughput S and the delay D are found in terms of T_1, f, b.

7- A numerical solution is applied to get b, f, T_1.

From the numerical values of b, f, T_1 values for both S and D can be found.

The solution steps are shown in the following flow chart (Fig 4-8) to make it easy to track through the mathematical steps that are followed to get to model solutions.
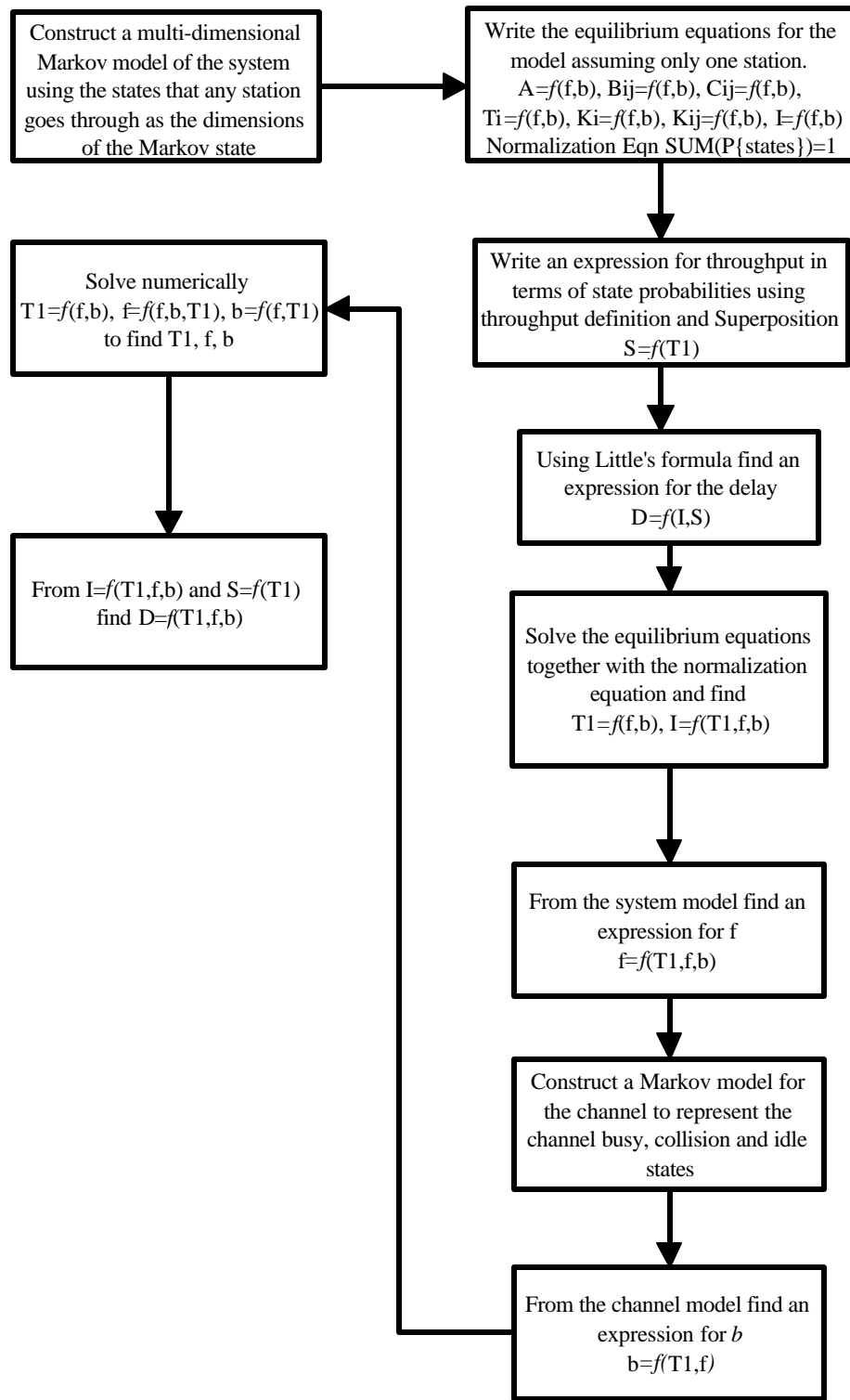
Fig 4-8 Flow chart of SSS solution method

The mathematical solution steps for the above system of equilibrium equations are given in Appendix A. Here, only the final results of these steps are given. These results are as follows:

$$G = U.L.\frac{l}{s}$$ (4-15)

$$S = U.L.T_1$$ (4-16)

$$D = \frac{1 - \frac{s}{l}T_1\left(1 + \frac{f^m}{1-f^m}\right)}{T_1}$$ (4-17)

$$T_1 = \frac{1}{\frac{s}{l} + (L+3) + \frac{1}{(1-b)(1-f^m)}\left[1 + \frac{2CW_{min}-1}{2}.b + \frac{s}{l}f^m(1-b) + f\left(\frac{2CW_{min}-2CW_{min}(2f)^{m-1}}{1-2f} + \frac{1-f^{m-1}}{2-2f}\right) + f(L+3)(1-b)\left(\frac{1-f^m}{1-f}\right)\right]}$$ (4-18)

The only step left to evaluate the system performance is to find expressions for f and b. These expressions will be used to get the value of T_1 in terms of system parameters. Having found T_1 in terms of the system parameters both throughput S and delay D can be evaluated.

The steps for finding expressions for f and b are given in Appendix A. The final results for these steps are as follows:

$$f = 1 - \left(1 - \frac{T_1}{(1-b)(1-f)}\right)^{U-1}$$ (4-19)

$$b = 1 - \frac{1}{1 + (L+3)\left(1 - \left[1 - (L+3)T_1 - (L+3)\left(\frac{f}{1-f}\right)T_1\right]^{U-1}\right)}$$ (4-20)

Equations (4-18), (4-19), and (4-20) are three non-linear equations in three variables T_1, f, and b. Assuming that all the system parameters like arrival rate l and number of users U are taking arbitrary values, these equations can be solved together numerically to get numerical values for T_1, f, and b. These values can be used in the throughput equation (4-16) and the delay equation (4-17) to obtain numerical results for the performance of the CSMA/CA protocol. This is the goal of the next section.

4.3.4. Numerical results

As mentioned before the equations (4-17), (4-18), and (4-19) will be solved together using numerical techniques. The numerical solution is obtained using the software package MathCad (version 8 Professional Academic Edition) and the function used in this package to obtain the numerical solution is called Minerr(). Having found numerical values for T_1, f, b, one can obtain throughput S and delay D from equations (4-16) and (4-17).

The results of this numerical solution are plotted in the following figures and commented on thereafter. These results are obtained assuming $s=1$ which means that on the average a message contains only one packet, and also the time slot $t = 3$ μsec. Before going into our analysis result Fig 4-9 that contains the normalized throughput against the normalized fresh offered load is given for both our analysis and the simulation done in [WEINMILLER97] to verify the analysis.
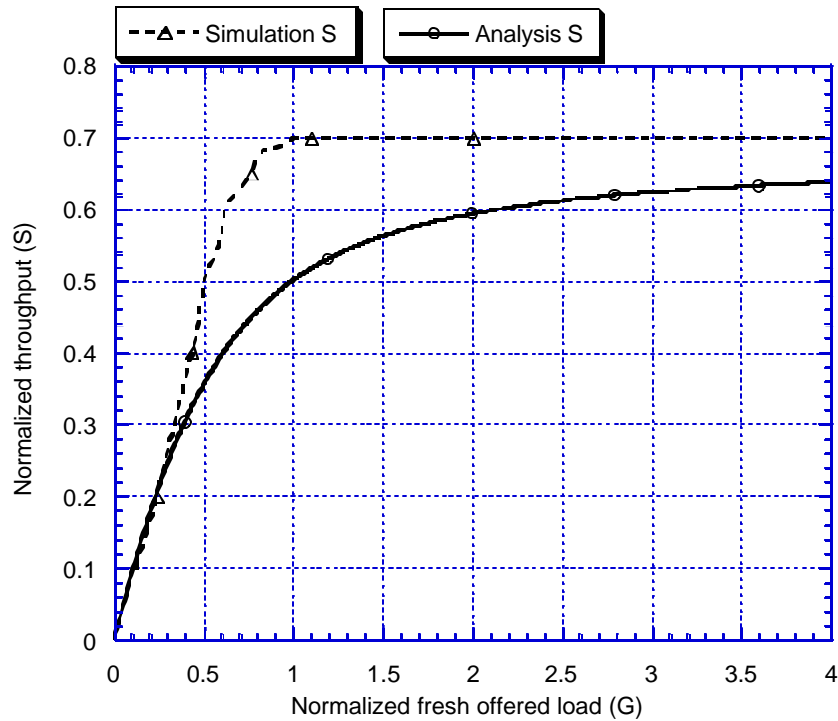
Fig 4-9 Normalized throughput *S* from analysis and simulation against normalized fresh offered load *G* for *L*=100 slots and *U*=2
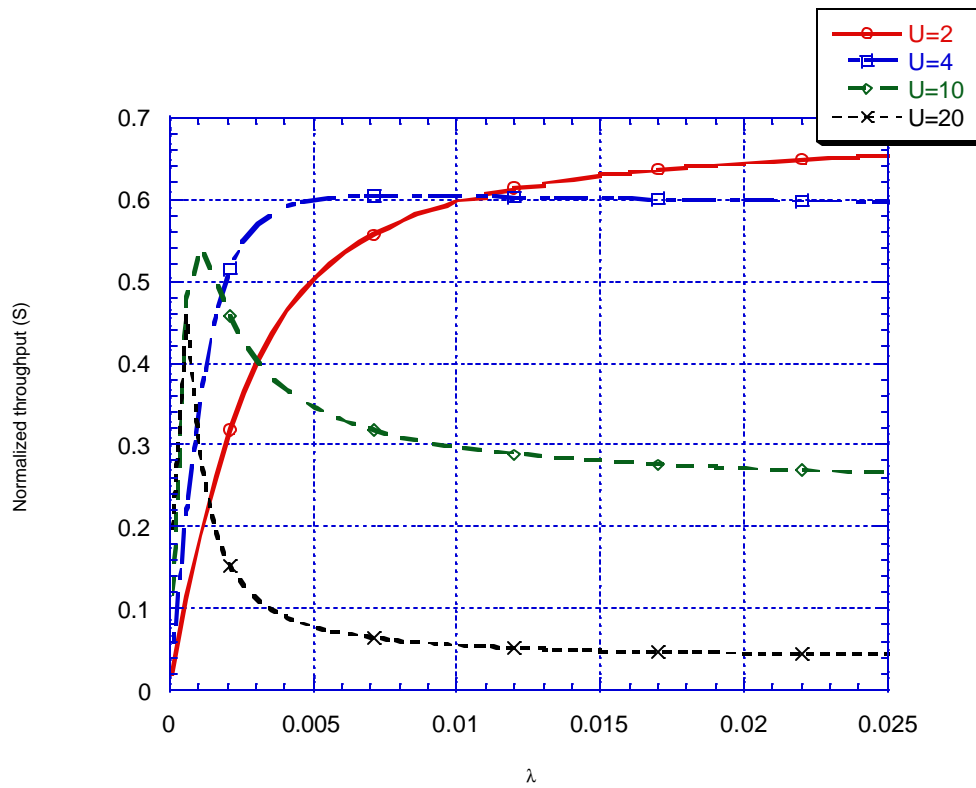
Fig 4-10 Normalized throughput S as a function of probability of arrival λ for different values of U and for m=2, and L=100 slots
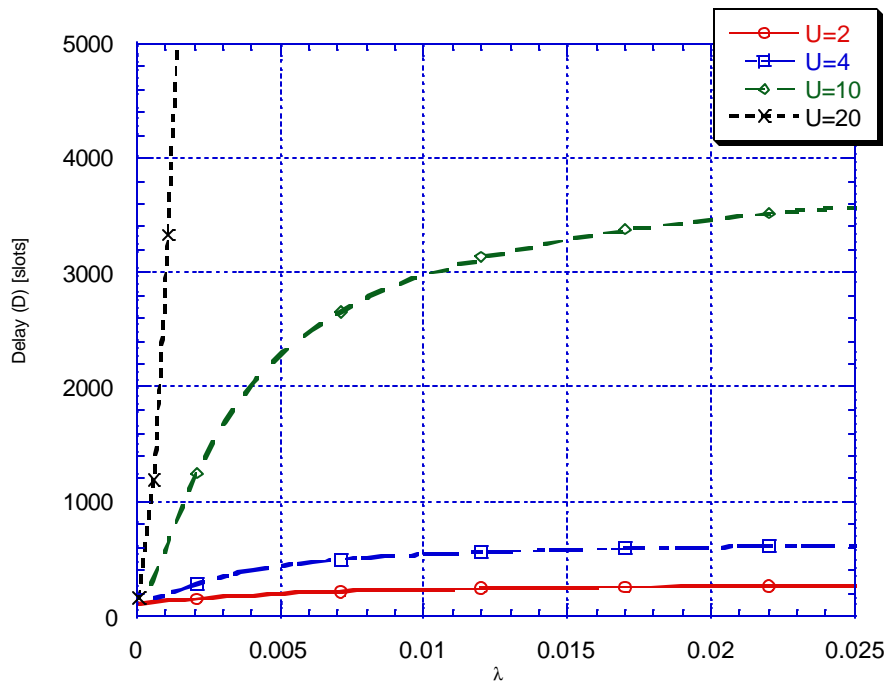
Fig 4-11 Delay D in slots as a function of probability of arrival λ for different values of U and for m=2, and L=100 slots
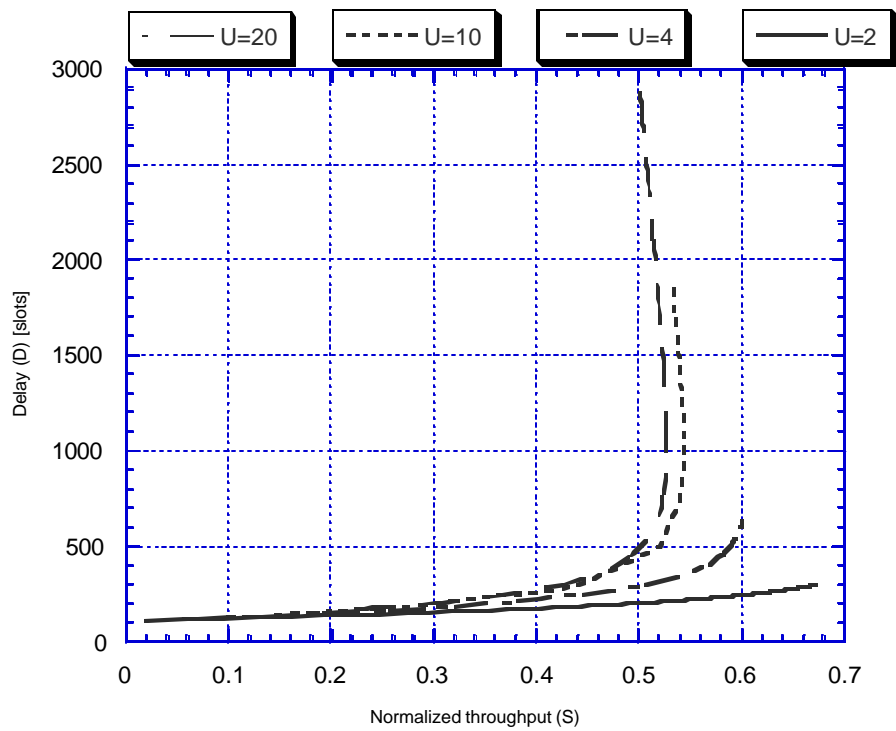
Fig 4-12 Delay *D* in slots against throughput *S* for different values of *U* and for *m*=8, *L*=100 slots
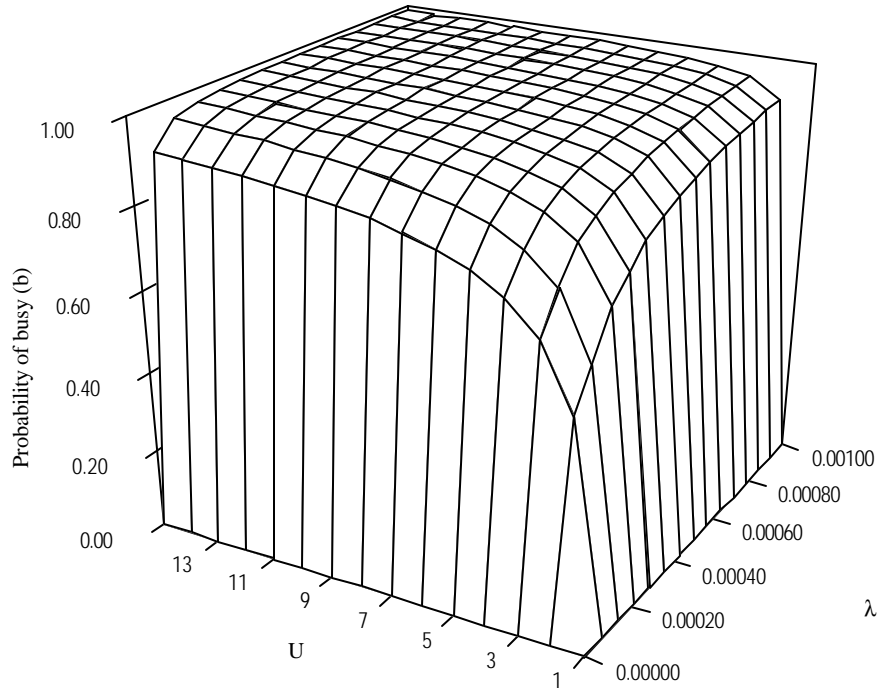
Fig 4-13 Probability of channel sensed busy *b* as a function of *U* and *l* for *L*=100 slots and *m*=2
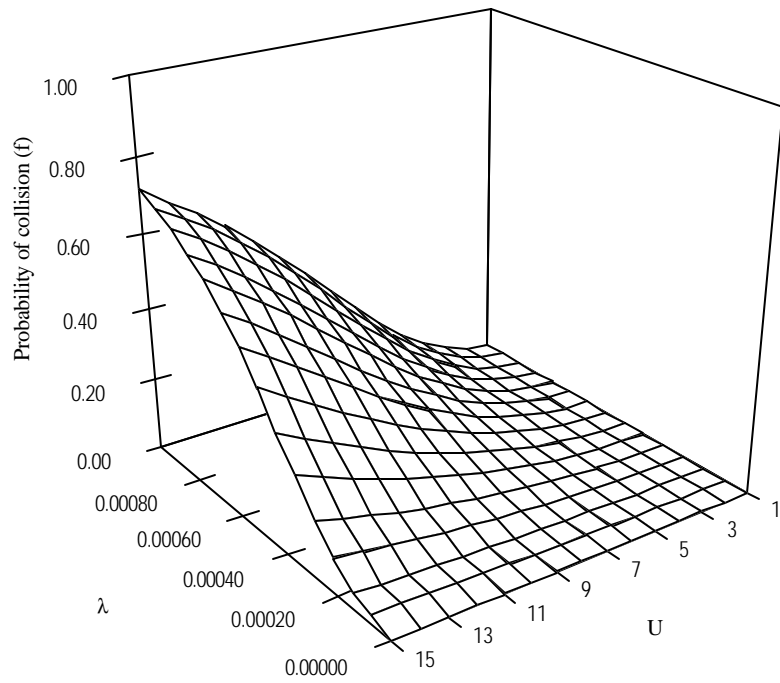
Fig 4-14 Probability of collision _f_ as a function of _U_ and _l_ for _L_=100 slots and _m_=2
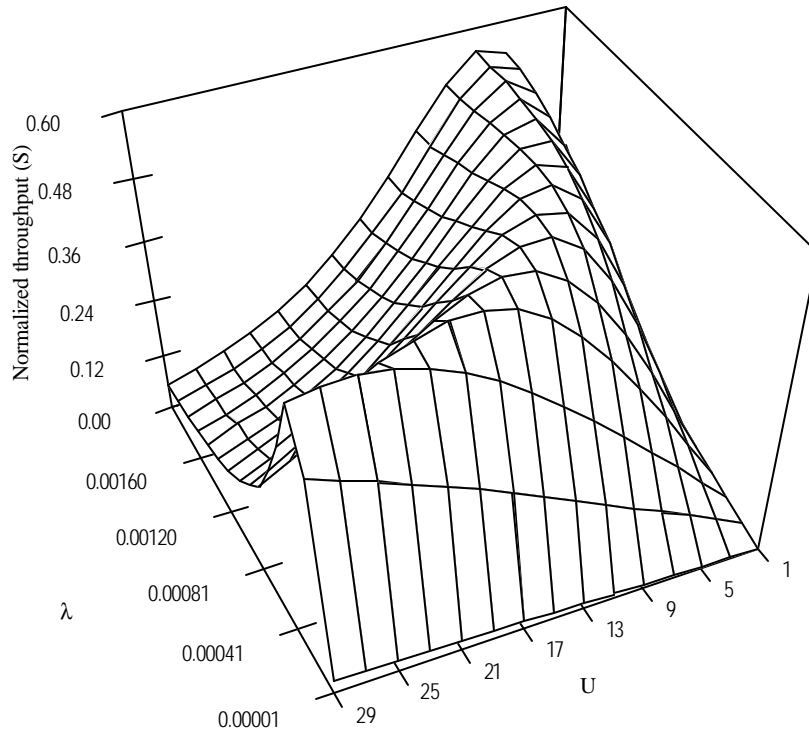
Fig 4-15 Normalized throughput _S_ as a function of _U_ and _l_ for _L_=100 slots and _m_=2
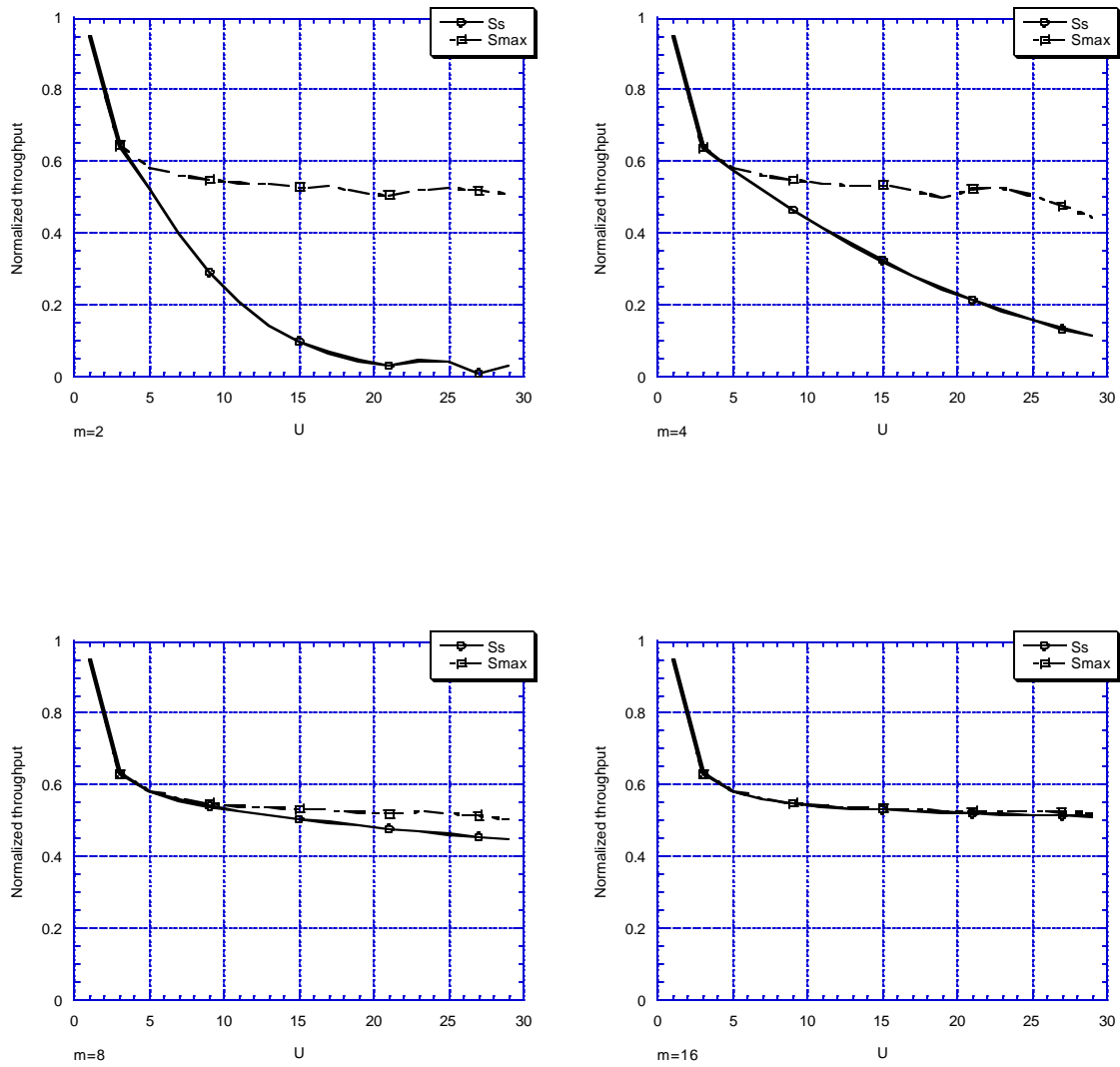
Fig 4-16 Maximum throughput S_{max} and steady state throughput S_s against number of users U for L=100 slots and different values of m
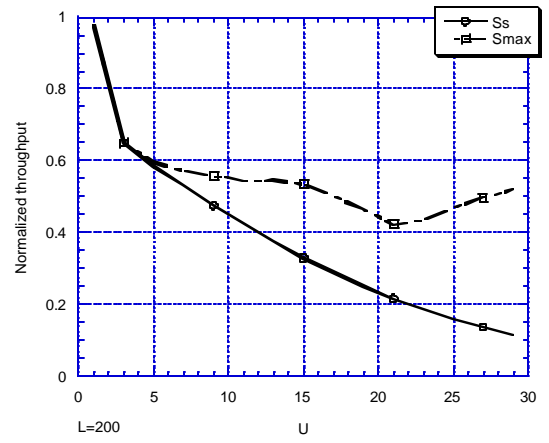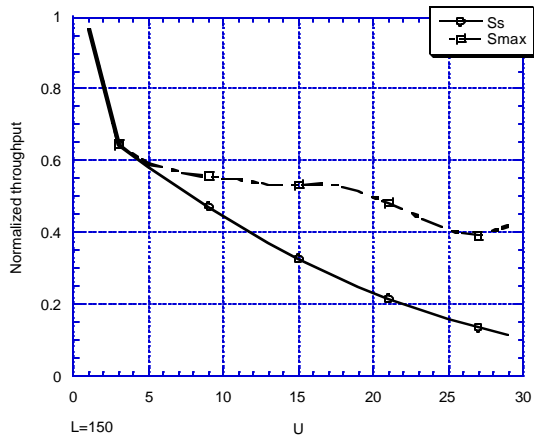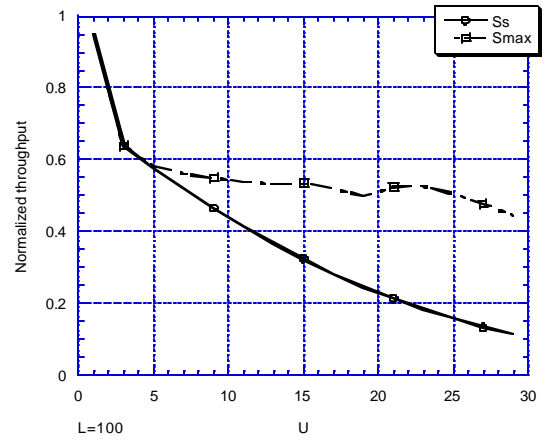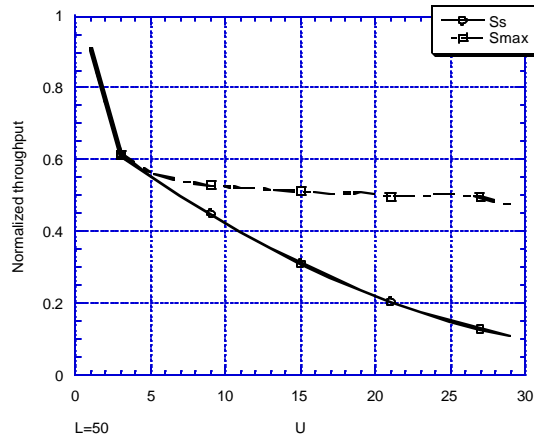
Fig 4-17 Maximum throughput S_{max} and steady state throughput S_s against number of users U for m=4 slots and different values of L

4.4.   Comments on Numerical Results

In this chapter a queuing model was proposed to allow for modeling the behavior of CSMA/CA MAC protocol with its back-off feature. A mathematical model was derived from the Markov queuing model using a technique called Single Station Superposition (SSS) which solves the model for one station while introducing the effect of the other stations using global parameters. After finding the steady state solution for the model the throughput and the delay for the system of U users are derived from their values for one user by properly applying a superposition for U users. Finally, numerical techniques are applied to solve the established performance equations and achieve numerical values.

In Fig 4-9 the numerical values for the throughput are compared with those calculated by simulation in [WIENMILLER97] for verification. From this comparison we can notice that the analysis results for throughput are always equal to or slightly less than that of the simulation. This deviation is a result of considering the use of inter-frame space, which slightly enhances the system throughput, in the simulation results.

Fig 4-10 is plotted by changing the value of probability of arrival λ and calculating the corresponding throughput and delay. This represents a single point in the figure. From this figure it is noticed that the throughput is increasing with increasing the arrival rate as this means that more slots of the channel capacity are used. This increase continues till it reaches a maximum value, which depends on the number of users in the system. This is because the more users are in the system the more the probability of collision (Fig 4-14) which reduces the throughput of the system.

On the other hand, one can notice that in Fig 4-10 the throughput for larger values of number of users drops below the maximum value it reaches as arrival rate increases. This criterion is due to the discarding of colliding packets after a certain number of retrials. To study this criterion in more details Fig 4-16 is given. In Fig 4-16 we notice that as the number of retrials before discarding increases the steady state value and the maximum value of the throughput for different number of users are approaching each other. In fact they not only approach each other but they reach higher values also for different values of number of users. This symptom is happening because as the number of retrials increases the probability of discarding a

message decreases and hence the throughput increases. This increase in the throughput is more significant at higher number of users, as the probability of collision is higher. This is why the two curves for maximum throughput and steady state throughput approach each other.

In Fig 4-11 one can observe that the delay reaches a saturation value with increasing the probability of arrival. This saturation behavior is a result of the consideration of finite number of retrials on colliding packets, which gives a bounded delay. On the other hand, in Fig 4-12 it is also noticed that increasing number of users always increases the delay regardless of the value of the probability of arrival. Looking at Fig 4-10 it is noticed that increasing the number of users enhances the throughput up to a certain value of probability of arrival after which it decreases the throughput. To combine these two figures in a comprehensive graph Fig 4-12 is given. In this figure the delay is plotted against the throughput for different values of arrival rate and for different number of users. It is noticed in the figure that increasing the probability of arrival increases the throughput and the delay up to a certain value of throughput. After this value the delay continues to increase while the throughput either settles or decreases then settles. It is also noticed that after a while the delay also settles and increasing probability of arrival changes neither throughput nor delay. This is observed in Fig 4-12 by the fact that all the points overlap on the same spot and the curve line does not go any longer.

From Fig 4-13 it is noticed that the probability of finding the channel busy reached high values with increasing both probability of arrival and number of users. This means that this protocol imposes large delay values at higher loads. This explains why the delay values are high in Fig 4-11.

This means that CSMA/CA is not suitable for time-bounded applications such as voice or video applications.

From Fig 4-15 it can be seen that for a value of number of retrials equal to 2 the system performs adequately in terms of throughput with number of users up to 10 users regardless of the value of the probability of arrival. Increasing the number of users above this value increases the probability of collision hence decreases throughput remarkably.

Finally, from Fig 4-16 it can be noticed that increasing the number of retrial times m affects both the maximum and steady state throughput values remarkably and makes them approach each other. On the other hand, in Fig 4-17 increasing the packet length L affects slightly the

maximum throughput values but does not result in making them approach the steady state value and does not affect the steady state value. This means that the major factor causing the steady state value of the throughput to differ from the maximum value is the operation of discarding colliding packets after a finite number of retransmissions.

4.5.   <u>Conclusions</u>

From the above comments and results the following conclusions can be obtained:

CSMA/CA is a protocol suitable for low number of users and low arrival rates (i.e. low offered traffic) which is expected for a contentions-based access scheme.

The protocol has a large delay value that makes it unsuitable for time-bounded applications. This is why it is accompanied by another controlled access scheme in the IEEE 802.11 standards.

The value of the number of retrials in the back-off scheme affects the protocol performance remarkably till a value of about 16 retrials where it acts as if it has infinite number of retrials.

Finally, using Single Station Superposition (SSS) technique with disrete time Markov analysis, we were able to establish a mathematical model for the performance parameters of a complicated protocol such as CSMA/CA. This model took into consideration the effect of the random binary exponential back-off algorithm with finite number of retransmissions, which is an important feature of CSMA/CA. The numerical results from the analytical model were close to those produced by simulation, which indicates that the model produced has good accuracy.

Chapter 5

Performance Analysis of Wireless/Wired LAN Interconnected via Wireless Access Point

Chapter 5

Performance Analysis of Wireless/Wired LAN Interconnected via Wireless Access Point

5.1. Introduction

With the advent of wireless technology and standards, a situation that is likely to be encountered is to have an existing wired LAN to which a wireless LAN extension is added. Another common situation is the need to connect a group of wireless LANs through a wired backbone. In all these situations, medium access schemes for the new hybrid (wireless-wired) LAN must exist to resolve contentions in both LANs.

In this chapter, an integration of two protocols is proposed as a protocol for such situations. This medium access scheme consists of integrating the Packet Reservation Multiple Access (PRMA) scheme [Appendix B] with the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) scheme. The first one, PRMA is used for wireless access to the access point, which is connected to the wired part. The second one, CSMA/CD on the other hand, is used for resolving contention on the wired medium between different access points and other wired nodes.

PRMA is chosen because, it is a centralized access scheme. This centralization feature is more suitable for the studied case as we use access points. It also has the advantage of stable operation under both heavy and light load conditions because of its adaptive nature and because it combines both controlled and contention based access schemes [MITROU90]. Meanwhile CSMA/CD is used because it is one of the most efficient and widely used techniques in wired LANs.

In order to study this kind of networks, we first present a description of the integrated protocol, focusing on the modifications needed at the base station to integrate the two protocols. Afterwards, equivalent queuing models are proposed. These queuing models are accompanied by some assumptions aiming at separating the two networks into two equivalent networks that has already been analyzed before. Achieving this goal, we reuse the analysis

already done for both equivalent networks and add to it the necessary modifications to come to analytical models that represent the performance parameters of the integrated system. Finally, numerical results for this proposed model are calculated, plotted and commented on.

5.2. <u>Integrating PRMA and CSMA/CD Protocols</u>

The proposed integrated protocol is aimed at achieving a wireless LAN access to a wired LAN through a dedicated access point. There are many challenging points in this situation. First, as wired LAN protocols are already more established than wireless LAN protocols, and because the main objective is to connect a wireless LAN to an already existing wired one, there should be no modifications made to the wired LAN MAC. Also, the analysis of such integration cannot be carried straightforward by establishing a Markov chain model for the entire network and solving it. This is because we have two different data source categories and two different data destination categories, and tremendous number of possible states for each source. These considerations will make a direct detailed Markov chain model for this situation mathematically intractable.

Integrating PRMA protocol for the wireless LAN access described in Appendix B and the well-known CSMA/CD for the wired LAN access needs some modifications and assumptions to allow for the use of this integration. Before describing the integrated protocol itself, the network architecture must be presented.

5.2.1. Network architecture

The network is assumed to consist of a wired bus type, CSMA/CD LAN connected through an access point to a centralized access wireless PRMA LAN (Fig 5-1).

The access point will have to act as a converter between these two different types of networks. On the other hand, since the PRMA is a centralized access scheme that already has a base station, it is assumed here that the base station for the wireless PRMA will act additionally as the access point between the wireless and the wired LANs. Due to this situation, message transmission can occur according to one of the following scenarios:
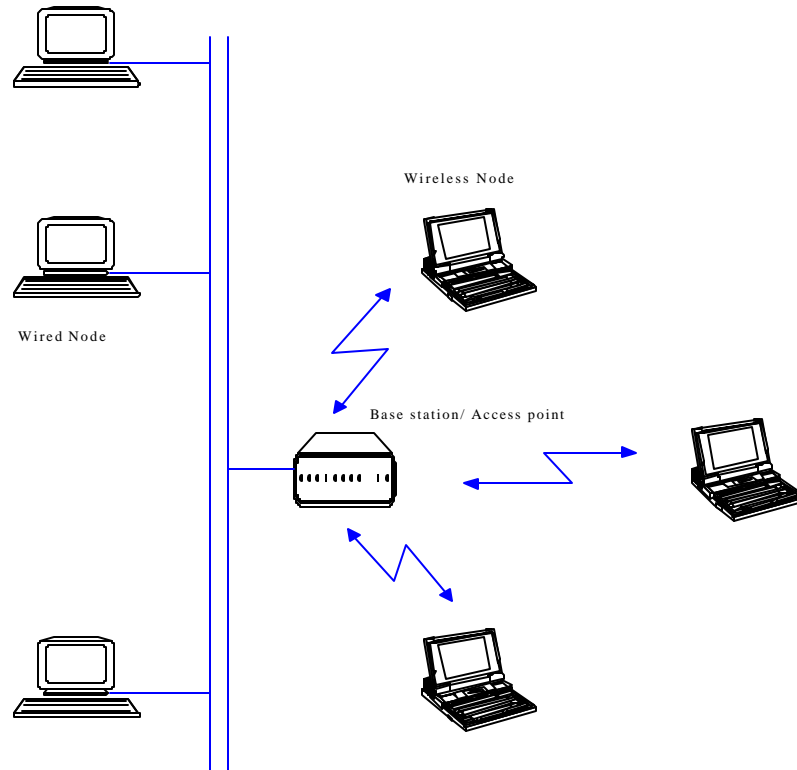
Fig 5-1 Topology of hybrid wireless/wired LAN

- Scenario 1: a message is transmitted from a wired user to a wired user.
- Scenario 2: a message is transmitted from a wireless user to a wireless user.
- Scenario 3: a message is transmitted from a wireless user to a wired user.
- Scenario 4: a message is transmitted from a wired user to a wireless user.

Before proceeding to describe what happens in each scenario of the proposed hybrid access scheme, we must state the modifications needed to allow for this integration. These modifications are chosen such that they allow for the required functionality, with an acceptable performance.

As mentioned before, these modifications should not affect the wired LAN MAC procedure which means that the analysis of the wired to wired part will not change in principle. Also, the modifications are chosen such that they will have a small effect on the wireless to wireless access technique. Finally, to allow better performance and yet, analytical isolation the access point is required to act as a bridge between the two networks and as a base station for the

wireless network. The bridging action is used to achieve the best possible performance. This bridging action implies that packets are not passed from one sub-network part to the other unless the intended receiver is in the other part.

Mainly, the modifications made are in the base station (access point) architecture and operation technique. These modifications are described in the following section.

5.2.2.   Access Point architecture and operation

Reviewing the access scenarios stated before and keeping in mind the independent operation of each of the two protocols integrated here, it can be seen that the base station should contain mainly two separate message queues (Fig 5-2). The first queue will hold messages from wireless stations to wired stations till they are served by the CSMA/CD protocol. The second queue will hold messages from wired stations to wireless stations till information slots are found for them in the outbound channel of the PRMA. These queues are needed to absorb the differences in traffic characteristics between the two protocols used. Moreover by assuming that these queues are of infinite length, traffic in each part of the network (wireless and wired) is isolated from the other part. On the other hand, messages coming to the base station that are coming from wireless stations and transmitted to other wireless stations are treated in the same way as for the case of PRMA only. This means that they are not queued in the access point, because they are not transmitted from the source till a slot is already reserved for them.

To allow for the integration of the two protocols, some changes must be made to the reservation queue and technique of the PRMA to accommodate the new hybrid topology. The most dramatic change is that in contrary to the case of PRMA protocol, in this protocol the wireless inbound channel traffic is not the only user of the outbound channel. Also, not all of the inbound channel traffic needs to be transmitted in the outbound channel. These facts are because of the existence of wired to wireless traffic, and wireless to wired traffic, respectively. This implies that separate reservation queues for wireless generated traffic and wired generated traffic are required. These queues control the inbound and a part of the outbound channel services, and the rest of the outbound channel, respectively. This is different from the case of PRMA where only one reservation queue controls both inbound and outbound channels.
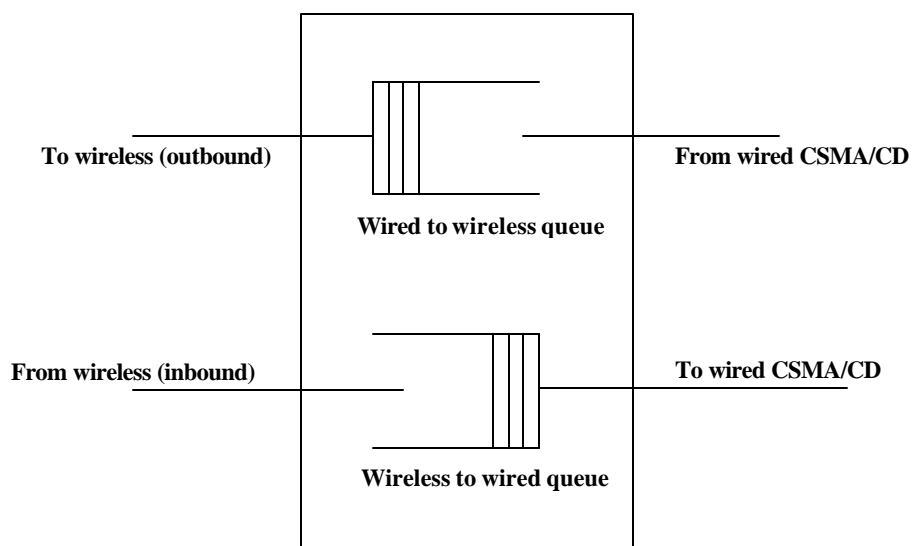
Fig 5-2 Message queues at the access point

The use of these queues will be such that, the wireless to wired traffic will control only a part of the inbound channel. Likewise, the wired to wireless traffic will control only a part of the outbound channel. Finally, the wireless to wireless traffic will still have to control both channels, since a source will not begin transmission unless an information slot is reserved for its message in both inbound and outbound channels. The access point queuing model and channel utilization will be as shown in Fig 5-3.

The path of the wireless to wireless reservations is shown in Fig 5-3 by the line which enters the ws-to-ws control box only, as the server of the wireless generated traffic reservations determines which controller is signaled according to the destination of the message. It must be noted that it is assumed that messages are not transmitted from the source till an information slot is reserved for them in both inbound and outbound channels to guarantee stability of the protocol and eliminate the need for queuing wireless to wireless messages in the access point.

To make the protocol more efficient and not disturb the wireless to wireless technique, which makes the analysis more tractable and simpler, it is assumed that a predetermined part of the inbound channel is used for wireless to wireless traffic. The rest of the channel is used for wireless to wired traffic. Moreover, the same fixed portion is used in the outbound channel for wireless to wireless traffic, and the rest is used for wired to wireless traffic. Due to these assumptions, the scenario of having (in the case of wireless to wireless transmission) only

118

inbound or outbound free slots will not happen, because the use of both channels equivalently is guaranteed again. This resembles exactly the case of PRMA alone. This allows for reusing the effort done in analyzing PRMA in case of wireless to wireless traffic without any modification to the analysis technique. On the other hand, the scenarios of wireless to wired and wired to wired will be simple integration of PRMA and CSMA/CD analysis results.

Fig 5-3 Hybrid PRMA protocol access point queuing model to accommodate wireless-to-wired and wired-to-wireless traffic in addition to wireless-to-wireless

Looking at Fig 5-3 we will notice that for the wired to wireless traffic a reservation is done. This is actually not needed and it only adds more delay to the system. This reservation can be eliminated because data are already there at the AP, and the AP is the only user of the outbound channel and the controller of the system resources. On the other hand, data arriving from wired CSMA/CD is of separate packet nature and not grouped into messages in the message queue inside the AP. This means that instead of serving all packets coming form a certain source to a certain destination using only one slot per frame, we will be able to use all the available slots in the outbound wired to wireless sub-channel for servicing the messages. Hence, the outbound wired to wireless traffic would be just First Come First Serve (FCFS) queue with multiple servers (information slots). This means that packets queued in the wired to wireless queue are served by the wired to wireless portion of the outbound channel with no need for reservation operation. Accordingly, the access point queuing model will be simplified by removing the wired generated traffic reservation queue, and the wd-to-ws control of the outbound channel as no reservation process is done for this traffic. Also, because of the assumption that fixed portion of the inbound traffic is used only by wireless to wireless traffic and an equivalent portion is also kept only for the same traffic in the outbound, the servers representing the inbound and outbound channels are separated each into two servers. One of these servers is for the wireless to wireless traffic and the other is for wireless to wired or wired to wireless traffic in the inbound and outbound channels, respectively. The modified access point queuing model is shown in Fig 5-4.

5.2.3.   Protocol description

The main changes made to the two protocols involved in the proposed hybrid protocol are mainly concerning the reservation queues, the message queues, and the access point architecture, all of which belong to the PRMA part. Hence, the wired to wired messages access technique, which uses the CSMA/CD is not affected. Accordingly, in the first scenario stated before the access scheme is purely a CSMA/CD algorithm with no changes (Fig 5-5).

In the second scenario; namely wireless to wireless transmission, a station wishing to transmit uses the Aloha technique to send a reservation request, which includes the address of the destination, to the access point. The access point detects from the destination address that the

transmission is a wireless to wireless one. Accordingly, the access point tries to reserve a slot in the wireless to wireless portion of both the inbound and outbound channels for this message. This is done in the same manner explained in Appendix B. Upon success of reservation the node starts transmitting the message packets in the reserved slots in the inbound and outbound channels till the message ends. Upon the message end, both slots are freed and could be reserved to any other message transmission (Fig 5-6).
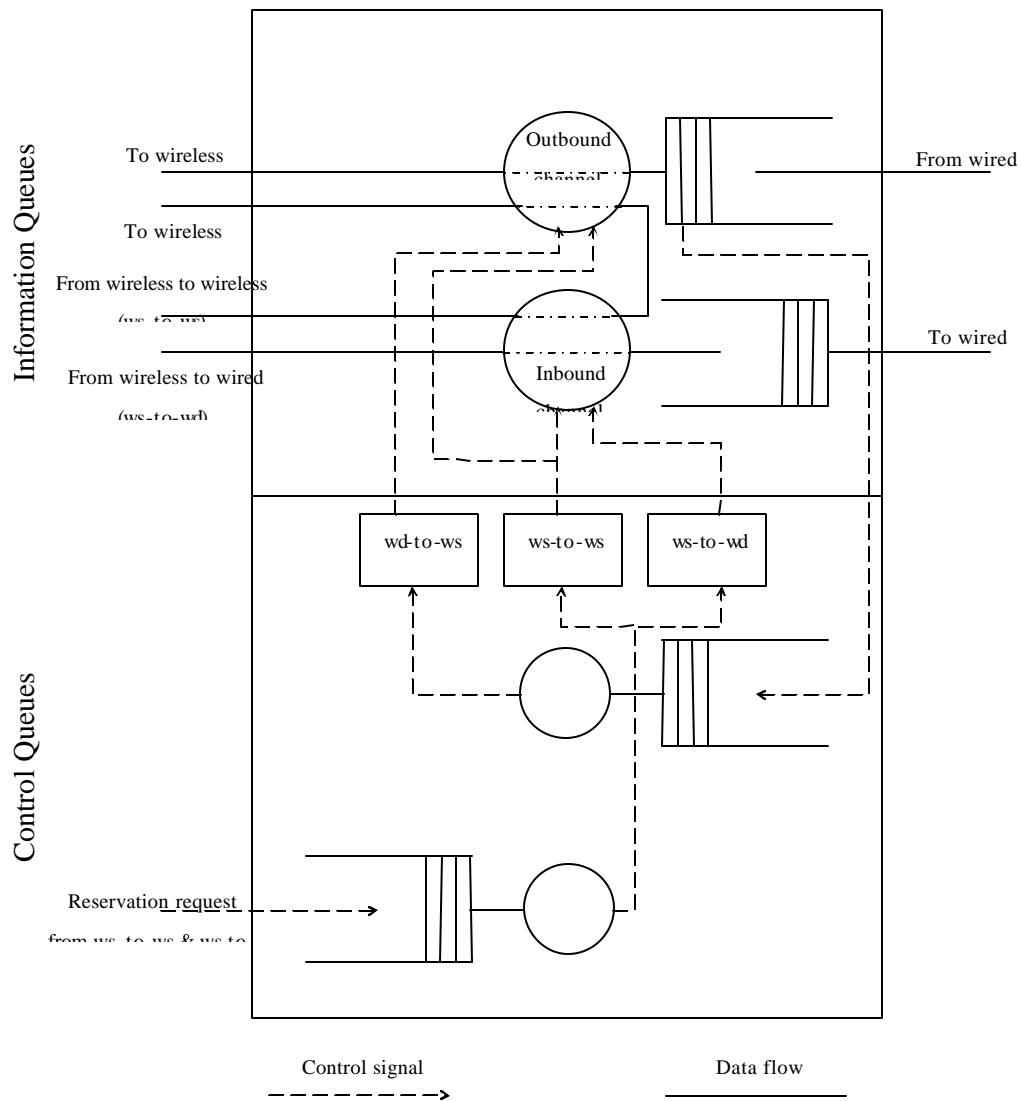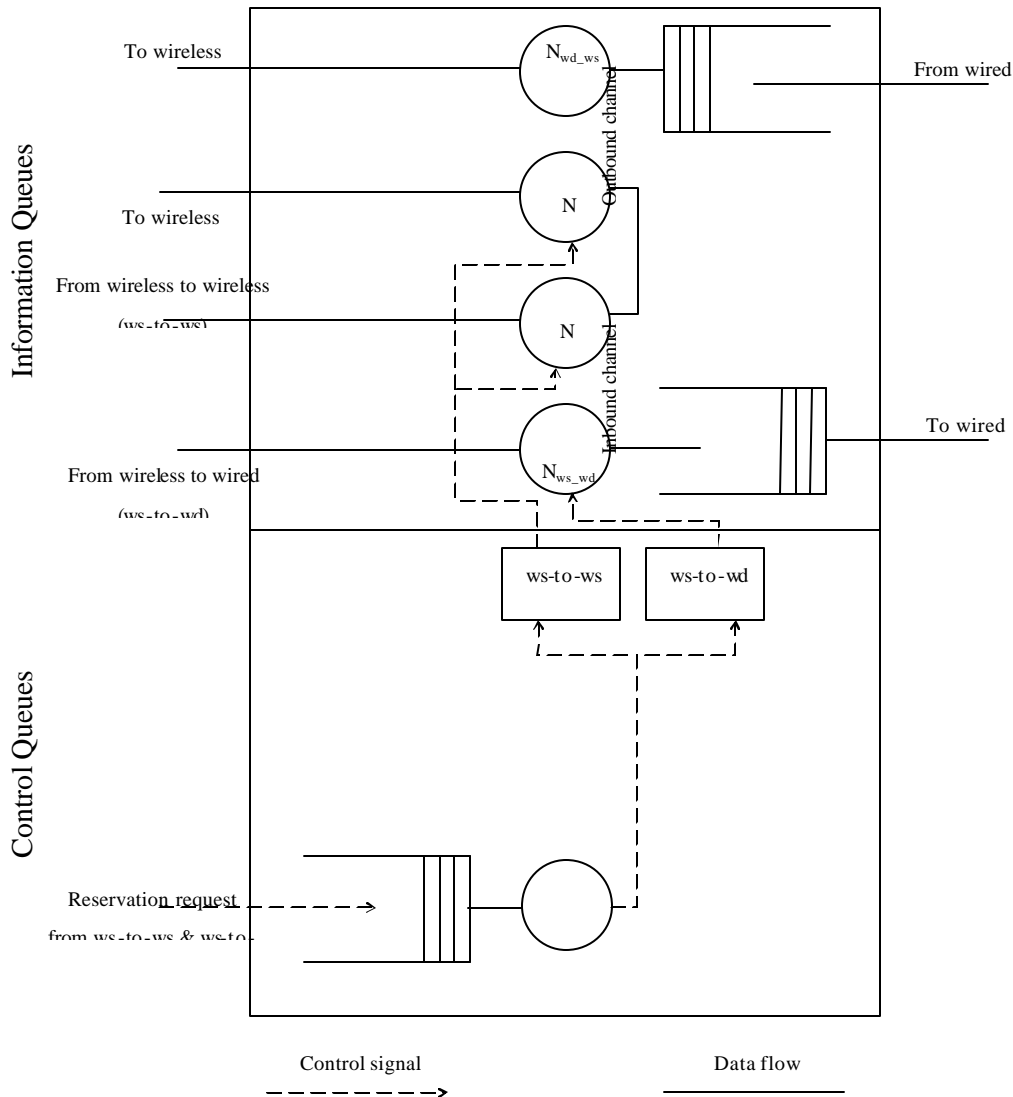

Fig 5-4 Modified hybrid PRMA protocol access point queuing model to accommodate wireless-to-wired and wired-to-wireless traffic in addition to wireless-to-wireless
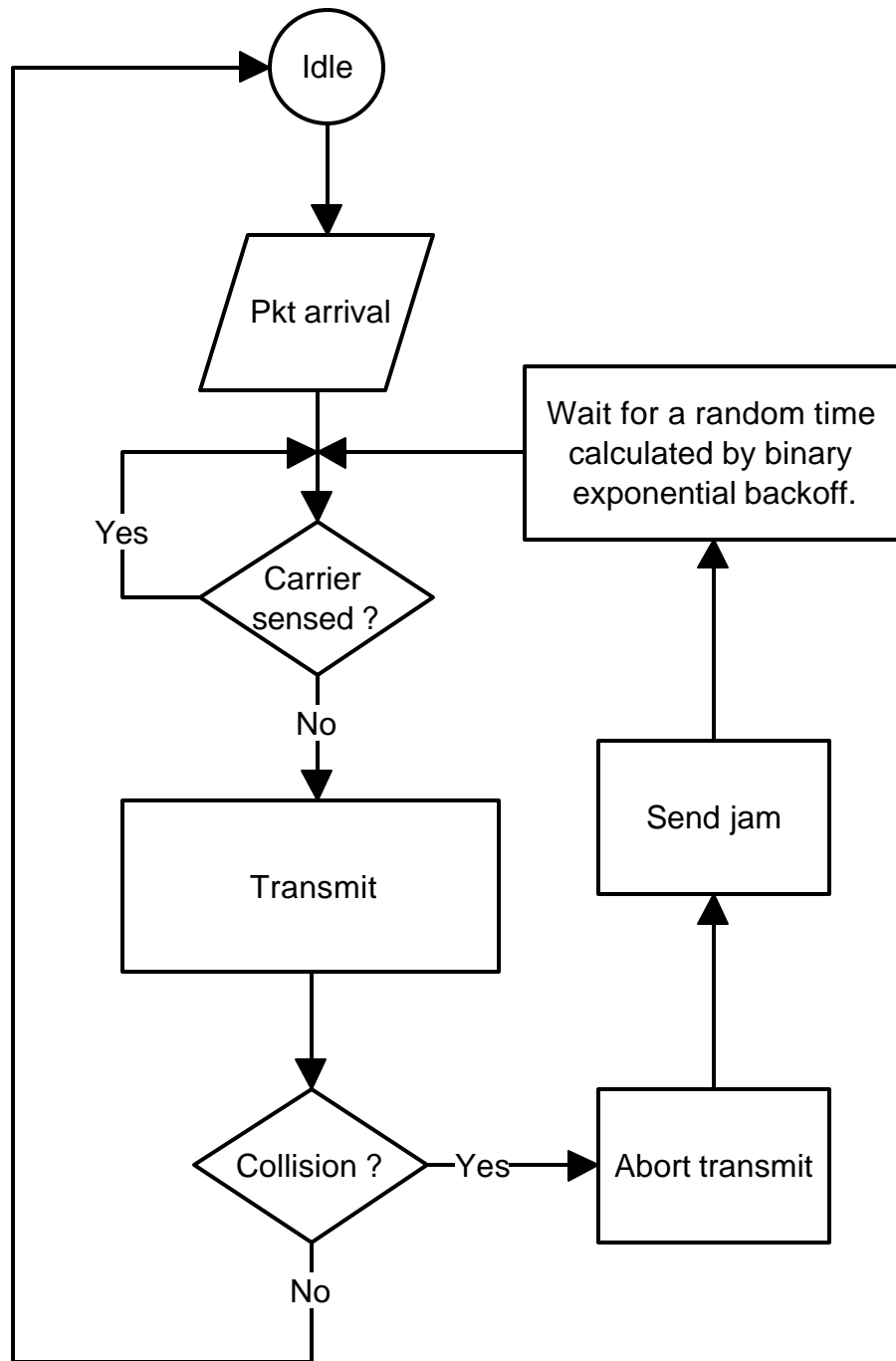
Fig 5-5 CSMA/CD protocol flow chart used for wired to wired traffic (scenario 1)
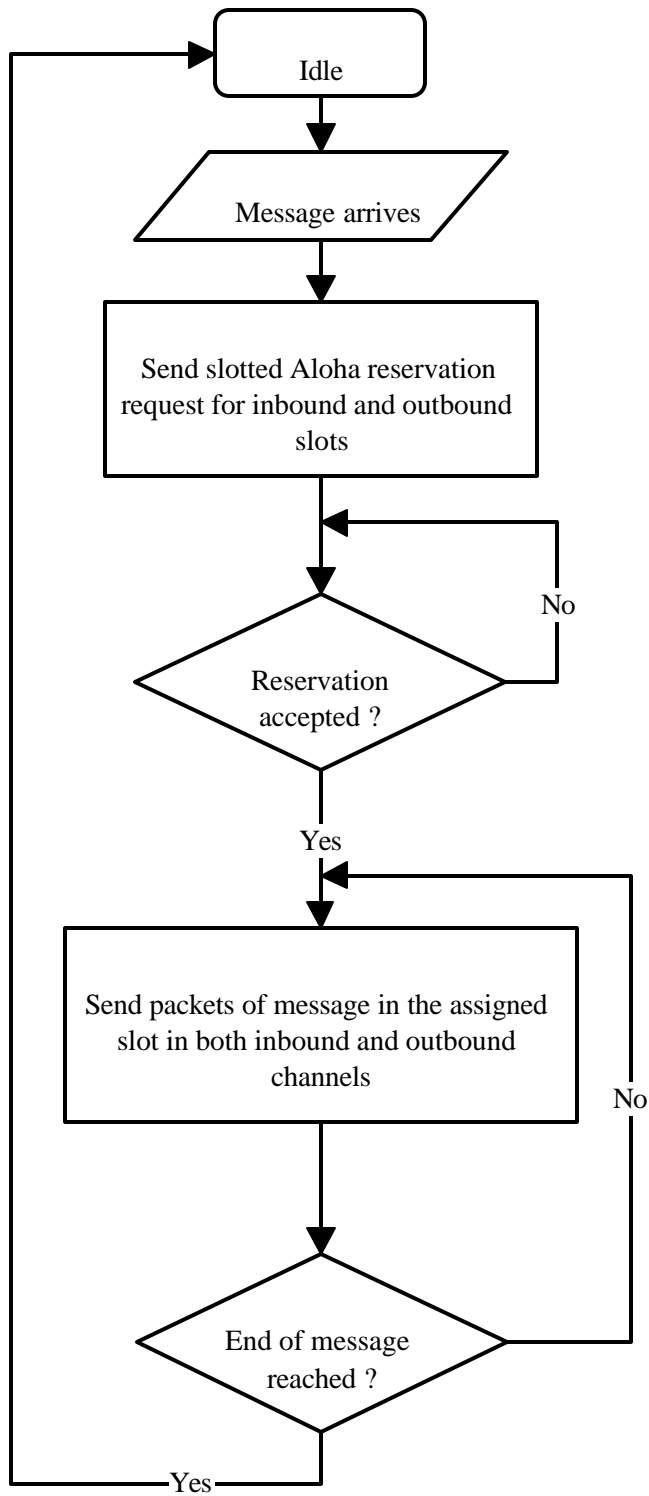
Fig 5-6 PRMA protocol flow chart used for wireless to wireless traffic (scenario 2)

123

In the third scenario; namely wireless to wired transmission a station wishing to transmit uses the Aloha reservation channel to send a reservation request to the access point with the destination address in it. The access point detects through the destination address that the transmission is a wireless to wired access. Accordingly, the access point only sends the request to the inbound reservation queue to request a slot in the wireless to wired traffic part of this channel. When the turn of the request comes and a free slot is found in the inbound channel, a reservation accept is sent to the requesting station with the number of the reserved slot included in it. Upon receiving the reservation accept, the transmitting station starts transmitting the packets of its message, one per frame during its assigned slots. These packets are queued in the access point in the wireless to wired message queue till they are serviced by the CSMA/CD service technique to get to their wired destination station. On the other hand, whenever a packet or more is found in the wireless to wired queue, the access point wired connection begins contending for the wired medium using CSMA/CD to transmit this packet to its destination. By assuming that this queue is large enough (approximately infinite) complete isolation is achieved between wireless access technique and the wired access one (Fig 5-7).

In the fourth scenario, a wired node wishing to transmit a packet to a wireless one transmits this packet to the base station using CSMA/CD. Packets are queued in the base station in the wired to wireless queue. Whenever a packet is received by the access point from the wired sub-network, it is assumed that the access point will assign the first available time slot in the wired to wireless part of the outbound channel to this packet, with no reservation process needed in this case (Fig 5-8).

This concludes the description of the newly proposed hybrid access scheme. The next section of this chapter will focus on establishing the mathematical models for this protocol. This establishment will try to reuse the effort already done in analyzing both the CSMA/CD and PRMA when each is used as a stand alone access protocol.
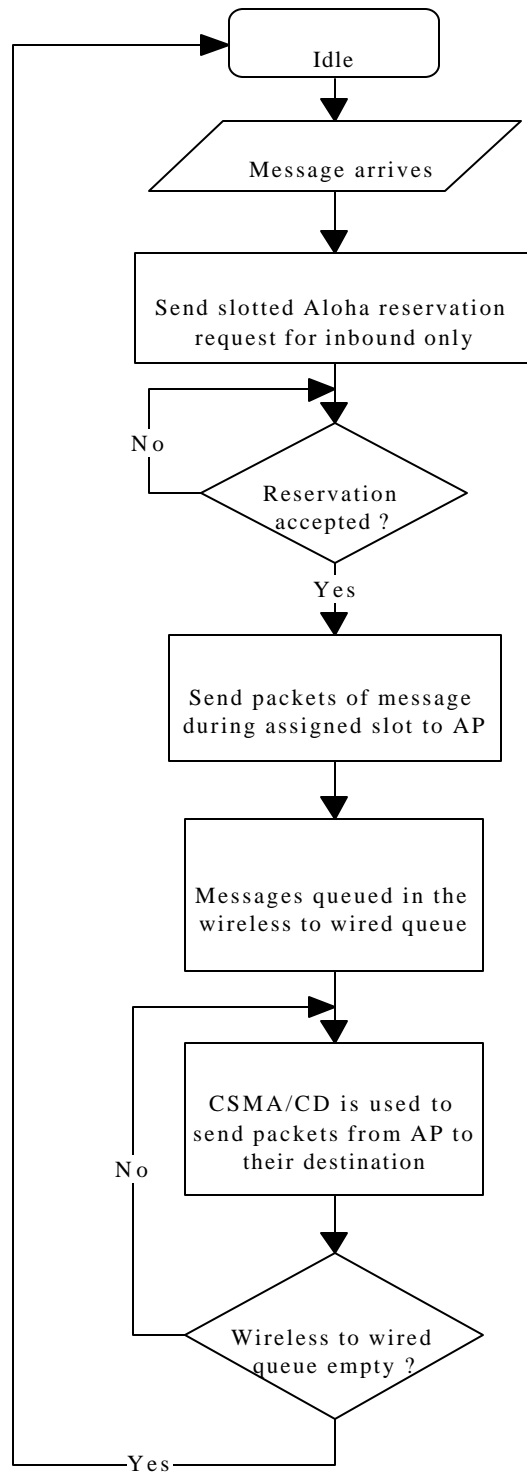
Fig 5-7 Wireless to wired traffic access algorithm (scenario 3)
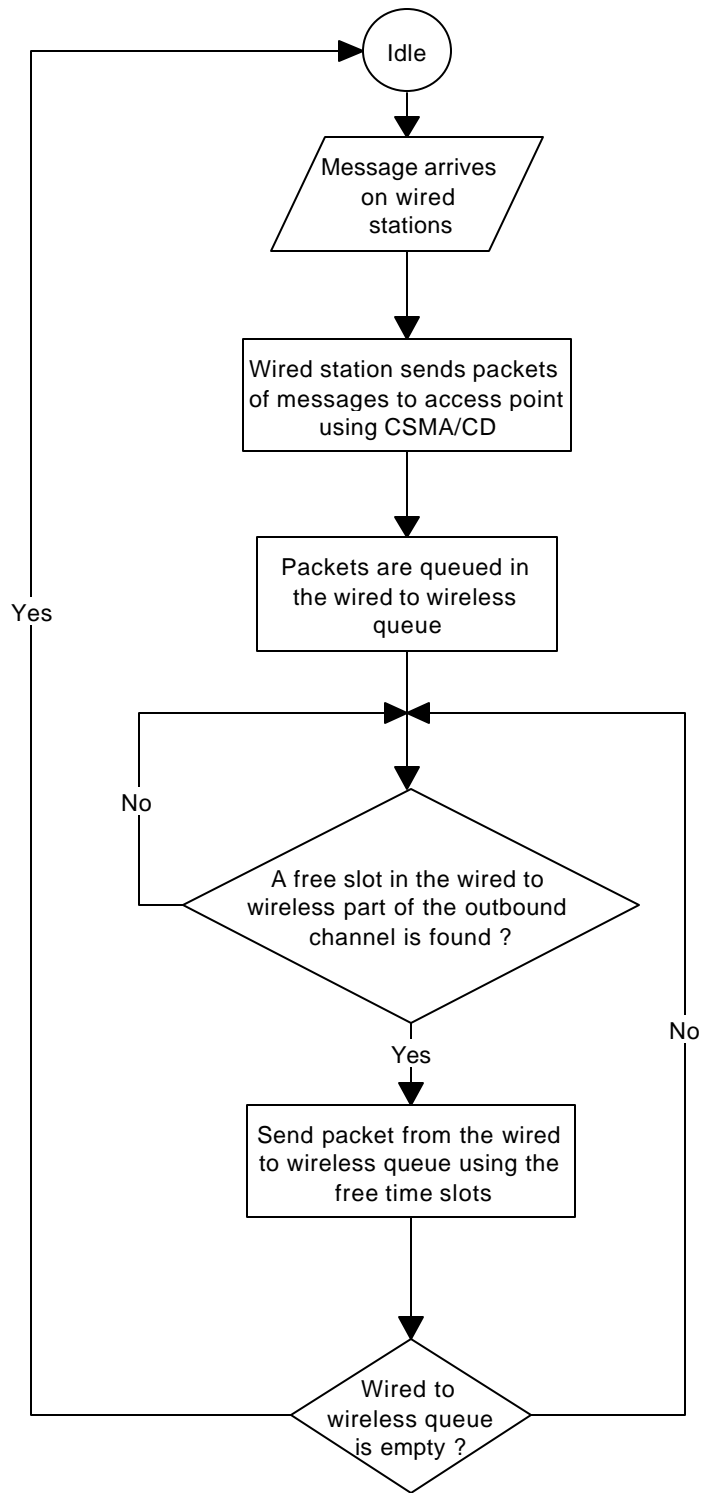
Fig 5-8 Wired to wireless traffic access algorithm (scenario 4)

5.3.   Analysis Strategy

As the protocol of interest here is a mixture of two protocols, the analysis will proceed in a somehow different way than the way used in the previous chapter. The reason for this difference is to make maximum use of the effort already done in analyzing each of these two protocols when used as a standalone protocol. The procedure will depend on the previously described scenarios to find out the needed modifications that should be made to the already established mathematical models for both PRMA (which is given in [MITROU90] and reproduced with more details in Appendix B) and CSMA/CD (which is given in [SCHWARTZ87]) to find out the mathematical model for the integrated protocol. The analysis of both the PRMA and CSMA/CD assume loss-less systems, which means that packets are retransmitted up to infinite number of times if necessary till they reach their destinations. One parameter that is going to be established mathematically and then analyzed quantitatively is the packet delay (D). Another important parameter to be investigated is the throughput (S), but this has to satisfy certain stability conditions. These stability conditions are a direct consequence of the infinite buffer and infinite number of retransmissions assumptions. The delay parameter is already established for each of the two protocols when used independently. In the following, the effect of using these two protocols in a hybrid scheme on the calculations of these two parameters will be stated verbally. Then, the new modified mathematical models will be established. As a first step towards this goal, three major characteristics of this proposed hybrid protocol should be stated. These are:

1. The traffic characteristics of the wireless part of the network are isolated from that of the wired part due to the use of infinite message buffers at the access point. This means that the access point can be treated as an extra wired node when analyzing the wired part and as an extra wireless node when analyzing the wireless part. This extra node in both cases will have a packet arrival rate characteristic that represents the effect of the traffic of the other part on the studied part.

2. The access point acts as a bridge between the two network parts. This means that it only passes a packet generated by a wireless station to the wired part when this packet has an intended receiver that is a wired station. The same rule applies to the wired to wireless traffic. This means that the wireless to wireless traffic will be totally isolated from the wired to wired traffic.

3. The wireless inbound and outbound channels are divided each into two separate parts. One part in each for the wireless to wireless traffic only. The other parts are the wireless to wired part in the inbound channel, and the wired to wireless part in the outbound channel. The division is made such that the wireless to wireless traffic occupies the same number of slots in both the inbound and outbound channels.

Accordingly, the analysis technique will follow the following steps:

(a) Divide the network (Hybrid Wireless-Wired network) into two separate equivalent sub-networks. One is completely a wired sub-network with CSMA/CD and the other is completely a wireless sub-network with the modified PRMA.

(b) Add to each sub-network an additional equivalent source that represents the equivalent traffic of the other separate sub-network that enters this sub-network.

(c) Find the values of the equivalent traffic represented by each of the equivalent sources on the studied separate network part. These values are calculated according to certain assumptions and approximations and should satisfy stability conditions.

(d) For each of the four possible scenarios described before, find the equivalent packet delay.

(e) The overall network average packet delay is the average of the four possible delays respectively. This average is calculated by multiplying the calculated delay for a certain scenario with the probability of having this scenario happening.

5.4. Model Assumptions

The mathematical solution technique carried in this chapter is based on the following assumptions:

1. We have finite number of stations in both wireless and wired sub-networks. This number is given by U_{ws} in the wireless sub-network and U_{wd} in the wired sub-network.
2. All the wireless stations are in radio contact with each other and with the base station.
3. The wireless channel is assumed to be ideal.
4. The wireless to wired and the wired to wireless message buffers at the access point are assumed to be of infinite length.
5. Fixed ratio of the traffic generated by wireless nodes represented by the factor R_{ws_wd} has wired stations as their destination.
6. Fixed ratio of the traffic generated by the wired nodes represented by the factor R_{wd_ws} has wireless stations as their destination.
7. All the arrival processes are assumed to be Poisson arrivals.

8. The wireless inbound channel is assumed to be divided into two separate sub-channels; one for the wireless to wireless traffic with number of slots N and the other is for the wireless to wired traffic with number of slots N_{ws_wd}.

9. The wireless outbound channel is assumed to be divided into two separate sub-channels; one for the wireless to wireless traffic with the number of slots N and the other is for the wired to wireless traffic with number of slots N_{wd_ws}.

10. The inbound wireless to wired sub-channel and the outbound wired to wireless sub-channel are assumed to have the same number of slots in order to keep the inbound frame length equal to the outbound frame length. This equality is essential to keep synchronization between the two channels. Hence, $N_{ws\_wd}=N_{wd\_ws}$.

11. The packet length in both wireless and wired parts of the network is assumed to be constant and to have the same value. This length is equal to m seconds in the wired network and at the same time N_{MS} time slots in the wireless network.

12. The time slot on the wired sub-network t is equal to one end to end propagation delay.

13. The time slot on the wireless sub-network is equal to the wireless to wireless sub-channel length in seconds F divided by the number of mini-slots in this sub-channel. The number of mini-slots is given by multiplying the number of slots in the wireless to wireless sub-channel N by the number of mini-slots per slot N_{MS}. Accordingly, wireless sub-network time slot is given by $F/(N.N_{MS})$.

14. The time slot in both sub-networks is assumed to be equal. This means that $t=F/(N.N_{MS})$.

15. Messages arrive at wireless idle users with rate l_{ws} (messages/sec/node).

16. The packet arrival rate at wired users is given by l_{wd} (packets/sec/node).

17. Wireless messages are assumed to have exponentially distributed length with average equal to T_{av} seconds. Accordingly, the rate of reaching message end s is given by $s=1/T_{av}$.

18. The equivalent source that represents the effect of each network on the other has an equivalent arrival process that is a Poisson process.

5.5. Queuing Models for Hybrid PRMA-CSMA/CD

Taking the aforementioned assumptions into consideration together with the previous description of the protocol, a mathematical model for the system can be established from the two mathematical models of the two protocols when each is used as a standalone protocol. Before going into math, an illustrative equivalent architecture to the network that uses the aforementioned assumptions to simplify the analysis would be given. This equivalent architecture is mainly dependent on assumptions (4), (5), (6), and (7) as will be seen. From

assumption (4), the network could be separated stochastically into two equivalent sub-networks; one of them is a wired sub-network while the other is a wireless sub-network. The effect of the wired to wireless traffic on the wireless sub-network and of the wireless to wired traffic on the wired sub-network is represented by an extra traffic source added for each equivalent sub-network.

From assumption (5) the equivalent source (or in this case node) added to the wired sub-network will have a message arrival rate λ_{ws_wd} equal to the throughput of the wireless to wired inbound channel $S^{in}_{ws\_wd}$ (Fig 5-9).

From assumption (6) the equivalent source (or in this case queue) added to the wireless network will have a message arrival rate equal to the traffic generated by wired nodes only $U_{wd}\lambda_{wd}$ multiplied by the ratio $R_{wd\text{-}ws}$ of the packets destined to wireless stations (Fig 5-10).

Station queue

Buffer length=1 msg.

Active wired stations (U-I)

Channel busyand collision state

Channel busy state

Channel

Wireless equivalent queue

$\mathbf{I}_{ws\text{-}wd}$

Queue occupation state

Idle stations (I)

Triggered back-off timer

Channel collision state

Control signal
– – – ➤

Data flow
·················
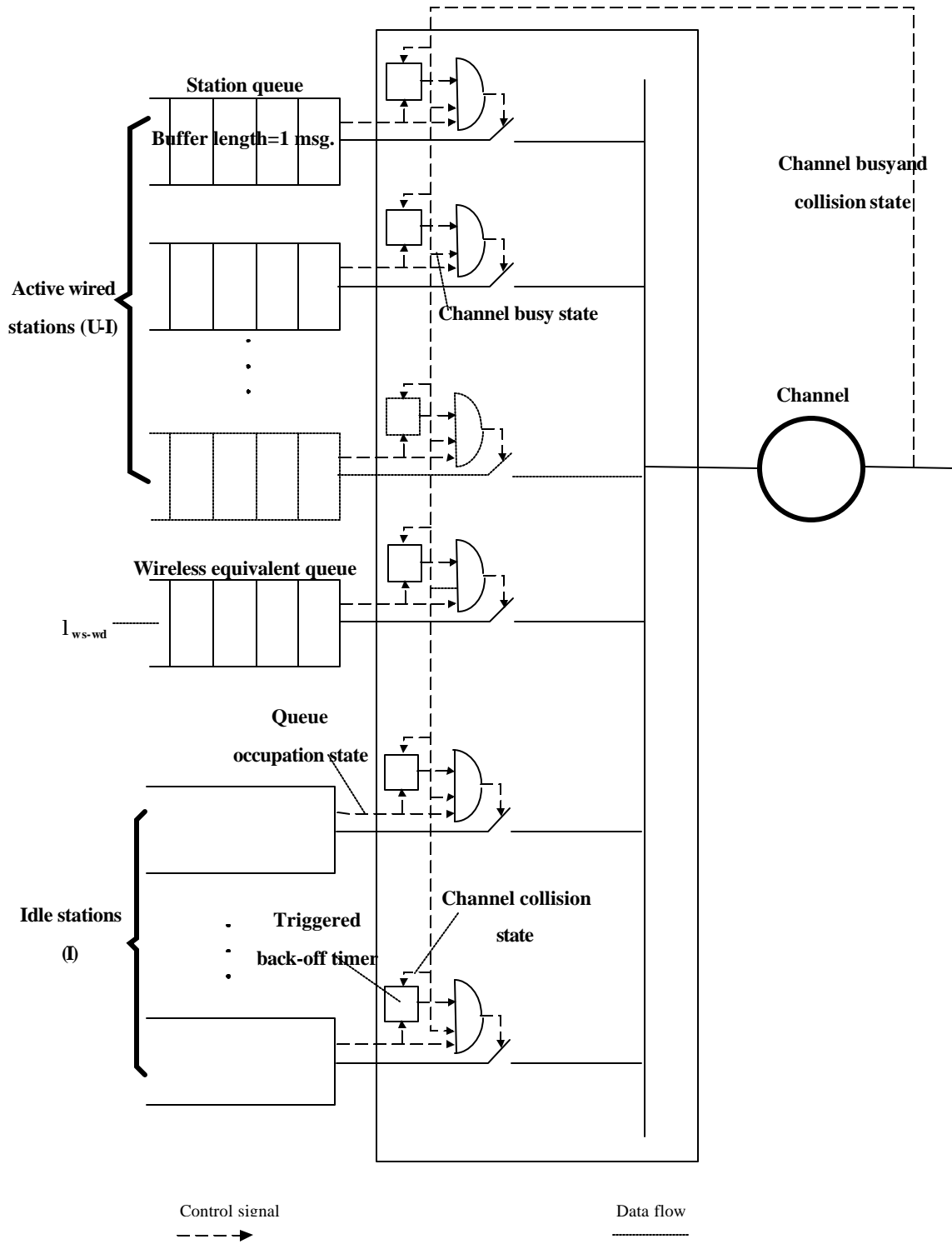
Fig 5-9 Wired separate equivalent sub-network queuing model

Wired to wireless queue

$I_{wd} \cdot U_{wd} \cdot R_{wd-ws}$

Wireless reservation queue

Aloha request queue

Aloha server

Station queue

Inbound control

U inputs to N-N$_R$ outputs matrix switches

Active stations

Outbound control

Information slots for inbound and

Idle stations

Control signal

Data flow
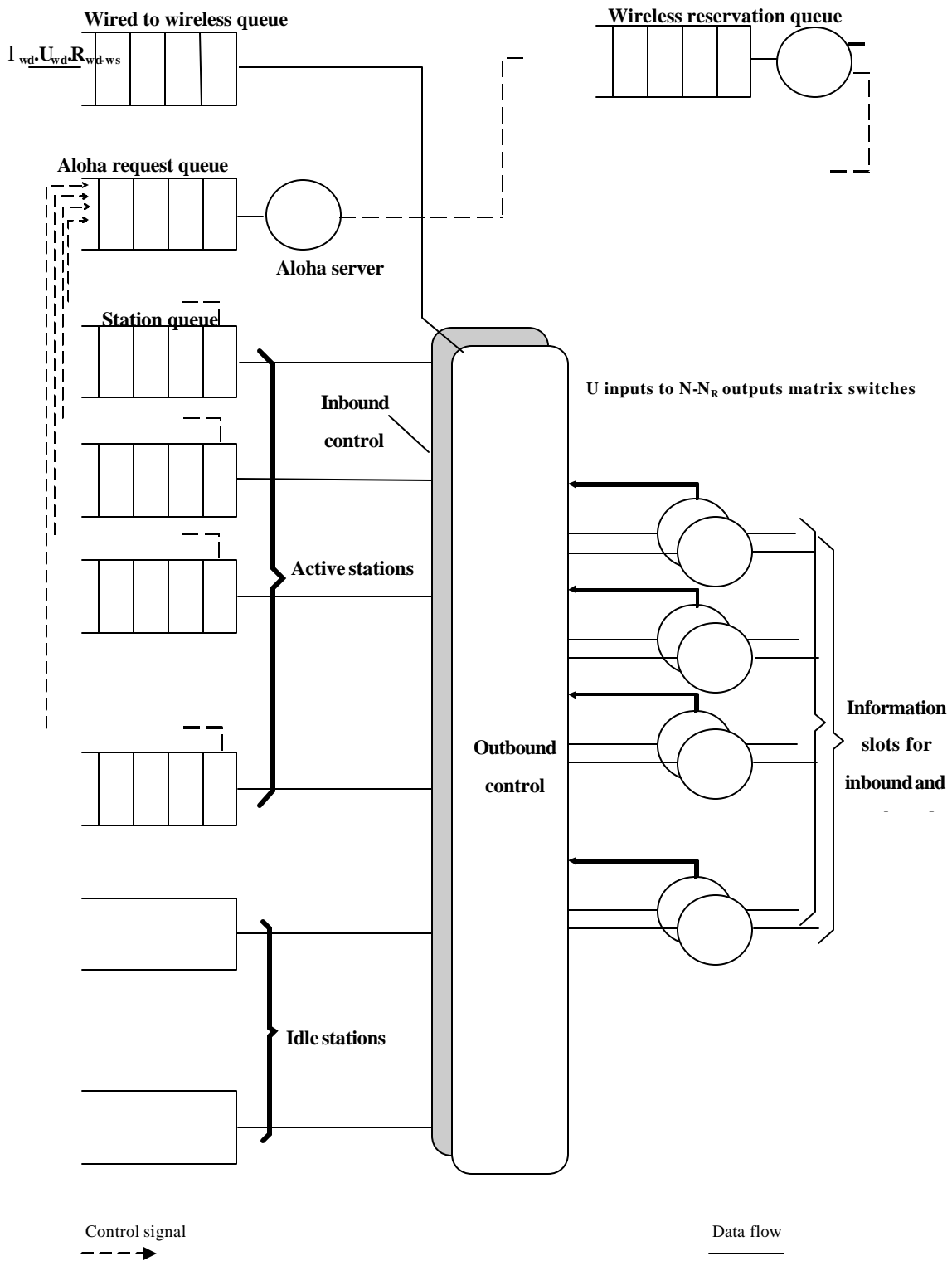
Fig 5-10 Wireless separate equivalent sub-network queuing model

132

Based on the above, we have a stochastically equivalent model for the hybrid network consisting of two separate sub-networks. One of them is a wired sub-network with bus architecture and uses CSMA/CD for MAC. This network consists of U_{wd} stations, each of which has an arrival rate equal to l_{wd} packets per unit time, and an extra wireless to wired traffic equivalent node with an arrival rate equal to l_{ws_wd} (Fig 5-9). The other separate sub-network is a wireless network with base station architecture and uses PRMA as a MAC protocol. This sub-network consists of U_{ws} stations, each with an arrival rate for idle users equal to l_{ws} messages per unit time, and an extra wired to wireless traffic equivalent source queue with an arrival rate equal to $l_{wd}.U_{wd}.R_{wd-ws}$ (Fig 5-10). It must be noted however that this extra source queue differs from the other wireless nodes in that it only uses the outbound reservation queue and channel. This means that instead of representing it as a node it is more convenient to represent it as an extra queue in the access point that is served by the outbound channel only.

Considering the wireless separate equivalent sub-network it is found that the reservation process is separated into two processes; one for the inbound channel and the other for the outbound channel. This is because the outbound information traffic is not totally composed of the inbound information traffic as in the case of PRMA alone. Here, the wired to wireless traffic is added to this channel. This implies that the inbound and outbound channels should be analyzed separately so as to accommodate the new situation. Fortunately, this will not mean that the analysis will start all over from the beginning, as will be seen later.

Moreover, using assumptions (8), (9) and (10) two goals are achieved:

(a) The wireless to wireless reservation in both inbound and outbound channels will be done simultaneously. This eliminates the possibility of having reservation completed in one channel and the information slots are reserved but not used till the other channel reservation is completed, which would have some performance degradation in terms of efficiency.

(b) The wireless to wireless analysis will be exactly the same as that for PRMA alone with the difference that the channel bandwidth allocated for the information packets is changed.

The only disadvantage of this assumption is that it might not be quite efficient in terms of channel utilization because of its static nature. This drawback can be overcome to a great extent by allowing for N and N_{ws_wd} to be calculated adaptively to minimize the delay and keep

the system stable. The base station according to the current traffic characteristics can do this calculation. However, for the analysis carried here it is assumed that these values are fixed and predetermined, but again the efficiency could be improved by choosing these parameters to minimize the delay according to expected network characteristics.

5.6. <u>Performance Analysis</u>

Let the number of wireless users be U_{ws}, the number of wired users be U_{wd}, and one access point is used which is also the wireless base station. The wireless channel contains two different frequency channels. One is called inbound and the other is the outbound. Each channel has a total length F_t seconds. Inbound channel is divided into two sub-channels; one for wireless to wireless traffic with number of slots N and length equal to F seconds, the other is for wireless to wired traffic with number of slots N_{ws_wd}. Outbound channel is also divided into two sub-channels; one for wireless to wireless traffic with number of slots N and length equal to F seconds, the other is for wired to wireless traffic with number of slots N_{wd_ws}. Since both inbound and outbound channels are equal in length, and as the wireless to wireless sub-channel in each are also equal, the wireless to wired sub-channel has to be equal to the wired to wireless sub-channel. This means that $N_{ws\_wd}=N_{wd\_ws}$.

The Aloha reservation process of the PRMA is done by the wireless users during the reservation slots of the inbound channel. These slots are identified to the wireless users by the base station. These slots has a number $N_R(k)$. Since any wireless node need not to know whether its destination is a wireless or a wired node, the Aloha contention and hence reservation slots for any type of traffic are assumed to occur only during the N slots wireless to wireless sub-channel. This means that $N_R(k)$ is always a portion of N. The reservation slots $N_R(k)$ happening in the inbound have mirrors in the outbound channel which occur at the same time with one packet slot delay. These slots in the outbound are called the acknowledgement slots and are used by the base station to inform the wireless nodes of the results of the reservation requests they sent. These slots are also with number $N_R(k)$ and are a part of the N wireless to wireless sub-channel in the outbound frame. The number of reservation slots and hence the number of acknowledgement slots $N_R(k)$ has a minimum value N_{RMIN} but can go higher depending on the wireless generated traffic characteristics. Any reservation slot is

further divided into N_{MS} mini-slots. This division is done to use slotted Aloha instead of pure Aloha for reservation contention process. This is because slotted Aloha has better performance (Fig 5-11).

The wireless generated traffic characteristics are defined by three state variables; namely i, h, and k. These variables are as follows:

i is the number of idle users.

h is the number of users contending in the Aloha reservation process.

k is the number of users succeeded in sending reservation requests; either waiting for the base station to assign an information slot for them or are sending packets of their message in the assigned information slot.
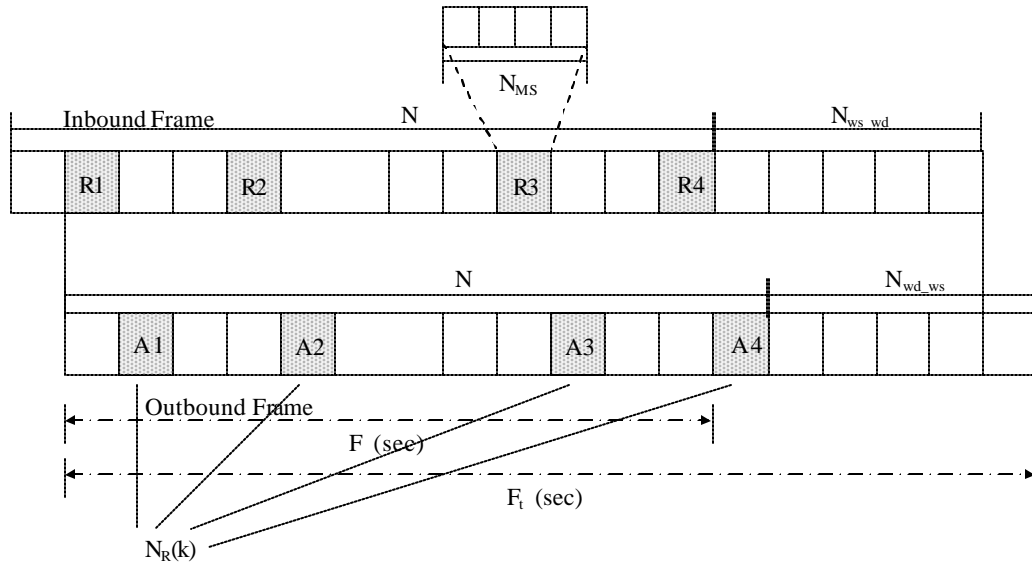
Fig 5-11 Inbound and outbound wireless channels and their sub-channels

From the analysis given in [MITROU90] and reproduced in Appendix B, the state probabilities of the wireless PRMA system are given by the following equations:

$$P(i) = \frac{\left(\dfrac{s}{l_{ws}}\right)^{i}\binom{U_{ws}}{i}}{\left(1+\dfrac{s}{l_{ws}}\right)^{U_{ws}}} \quad , \qquad 0 \le i \le U_{ws} \tag{5-1}$$

135

$$P(h \mid i) = P(0 \mid i) \prod_{j=1}^{h} \frac{b_{j-1}}{d_j} \quad , \qquad 1 \le h \le U_{ws} \tag{5-2}$$

where:

$$b_h = \left[1 - h.p_t.P_{nerr}(1-p_t)^{h-1} \frac{N_R(k)}{N} \right].i.\mathbf{1}_{ws}$$

$$d_h = h.p_t.P_{nerr}(1-p_t)^{h-1} \frac{N_R(k)}{N} \left(\frac{N.N_{MS}}{F} - i.\mathbf{1}_{ws} \right)$$

$$P(0 \mid i) = \frac{1}{1 + \displaystyle\sum_{h=1}^{U_{ws}} \prod_{j=1}^{h} \frac{b_{j-1}}{d_j}}$$

$\mathbf{s} = 1/T_{av} \qquad , \qquad p_t =$ probability of deciding to transmit in the Aloha

$P_{nerr} =$ probability of error free transmission.

From probability theory the joint probability $P(h,i)$ in terms of conditional probability $P(h/i)$ and $P(i)$ is given by:

$$P(h,i) = P(h \mid i).P(i) \tag{5-3}$$

Substituting (5-1) and (5-2) into (5-3) yields:

$$P(h,i) = \begin{cases} \dfrac{1}{1 + \displaystyle\sum_{h=1}^{U_{ws}} \prod_{j=1}^{h} \frac{b_{j-1}}{d_j}} . \displaystyle\prod_{j=1}^{h} \frac{b_{j-1}}{d_j} . \dfrac{\left(\frac{\mathbf{s}}{\mathbf{1}_{ws}}\right)^i \binom{U_{ws}}{i}}{\left(1 + \frac{\mathbf{s}}{\mathbf{1}_{ws}}\right)^{U_{ws}}} & , 0 \le i \le U_{ws} \,, 1 \le h \le U_{ws} \,, i+h+k = U_{ws} \\[3em] \dfrac{1}{1 + \displaystyle\sum_{h=1}^{U_{ws}} \prod_{j=1}^{h} \frac{b_{j-1}}{d_j}} . \dfrac{\left(\frac{\mathbf{s}}{\mathbf{1}_{ws}}\right)^i \binom{U_{ws}}{i}}{\left(1 + \frac{\mathbf{s}}{\mathbf{1}_{ws}}\right)^{U_{ws}}} & , 0 \le i \le U_{ws} \,, h = 0 \,, i+h+k = U_{ws} \\[3em] 0 & , \text{otherwise} \end{cases}$$

$$\tag{5-4}$$

This gives an expression for the probability of having i users idle, h users contending for Aloha reservation and k users transmitting or waiting for an information slot to be assigned to them. These are the state variables of the PRMA Markov model (Appendix B). It is now time to

find expression for delay for different scenarios in terms of i, h, and k and then average them over the state probabilities $P(h,i)$ to find the average delay for each scenario. Then using these delays the overall average delay is found by summing them with appropriate averaging probabilities of each scenario.

5.6.1. Wireless to wireless packet delay D_{ws_ws}

From the description given to this scenario it is obvious that it is exactly the same of PRMA when used alone. This means that the state dependent message delay $D_{ws\_ws}(h,i)$ will have the same expression as that given for PRMA in Appendix B with the difference that the total frame size is now F_t, where $F_t=[(N+N_{ws\_wd})/N].F$. This means that the total wireless to wireless message delay $D_{ws\_ws}(h,i)$ will consist of two parts; the message access delay $D_{acc}(h,i)$ and the message transmission delay D_{tr}. Referring to the analysis given in Appendix B, these are given by:

$$D_{acc}(h,i) = D_{aloha}(h,i) + D_{queue}(h,i) \tag{5-5}$$

where:

$$D_{aloha}(h,i) \cong [h/S_{aloha}(h,i)]\,F/(N.N_{MS}) + n_{tr}(h,i).Ack.F/N \tag{5-6}$$

$$S_{aloha}(h,i) = h.p_t.P_{nerr}.(1-p_t)^{h-1}\frac{N_R(k)}{N} \tag{5-7}$$

$$N_R(k) = \begin{cases} N-k \;, & k < N - N_{RMIN} \\ N_{RMIN} \;, & otherwise \end{cases} \tag{5-8}$$

$$k = U_{ws} - h - i \tag{5-9}$$

$$n_{tr}(h,i) = \frac{1}{(1-p_t)^{h-1}.P_{nerr}} \tag{5-10}$$

Here $n_{tr}(i,h)$ denotes the total number of retries before a successful transmission occurs, and Ack is the number of slots waiting for reservation acknowledgement. Now using (5-7), (5-8), (5-9), and (5-10) into (5-6) one obtains:

$$D_{aloha}(h,i) \cong \frac{F}{P_{nerr}.(1-p_t)^{h-1}}\left(\frac{1}{p_t.N_R(k)N_{MS}} + \frac{Ack}{N}\right) \tag{5-11}$$

$$D_{queue}(h,i)=\begin{cases}0.5F_t\ , & k\left(1-R_{ws\_wd}\right)<N-N_{RMIN}\\ F_t+\left[k\left(1-R_{ws\_wd}\right)-\left(N-N_{RMIN}\right)\right]\dfrac{1}{\mathbf{s}\left(U_{ws}-i\right)\left(1-R_{ws\_wd}\right)}\ , & otherwise\end{cases}$$

(5-12)

On the other hand, the message transmission time is given by:

$$D_{tr}\cong F_t\cdot\frac{\left(1/\mathbf{s}\right)}{\left(F/N\right)}=\frac{N.F_t}{\mathbf{s}.F}$$

(5-13)

Then the total message delay $D_{ws\_ws}(h,i)$ is the sum of both.

$$D_{ws\_ws}(h,i)=D_{acc}(h,i)+D_{tr}$$

(5-14)

The average message delay $\overline{D_{ws\_ws}(h,i)}$ is given by:

$$\overline{D_{ws\_ws}(h,i)}=\sum_{h=0}^{U_{ws}}\sum_{i=0}^{U_{ws}-h}D_{ws\_ws}(h,i).P(h,i)$$

(5-15)

Since a message consists on the average of $N/(\mathbf{s}.F)$ packets, the average packet delay is the average message delay divided by the number of packets per message. Accordingly:

$$D_{ws\_ws}=\overline{D_{ws\_ws}(h,i)}\frac{\mathbf{s}.F}{N}$$

(5-16)

5.6.2.   Wireless to wired packet de lay D_{ws_wd}

In the wireless to wired traffic the packet suffers from delays in the wireless inbound channel due to Aloha reservation, reservation queuing and transmission delay. After reaching the access point, the packet is sent to its final destination via CSMA/CD protocol, which adds further delay to it. The wireless delay will have the same formulas in the wireless to wireless scenario with the difference that the number of information slots this time is N_{ws_wd}. This means that the inbound message delay $D^{in}_{ws\_wd}(h,i)$ till it reaches the wireless to wired message buffer is the sum of the inbound message access delay $D^{in}_{acc}(h,i)$ and the transmission delay D_{tr}.

$$D^{in}_{ws-wd}(h,i)=D^{in}_{acc}(h,i)+D_{tr}$$

(5-17)

where:

$$D^{in}_{acc}(h,i)=D_{aloha}(h,i)+D^{in}_{queue}(h,i)$$

(5-18)

$$D_{queue}^{in}(h,i) = \begin{cases} 0.5F_t \quad , & k.R_{ws\_wd} < N_{ws\_wd} \\ F_t + [k.R_{ws\_wd} - N_{ws\_wd}]\dfrac{1}{\boldsymbol{s}(U_{ws} - i)R_{ws\_wd}} \quad , & otherwise \end{cases}$$

(5-19)

Note that $D_{queue}^{in}(h,i)$ is the reservation queue delay in the wireless to wired inbound sub-channel which is different than that of the wireless to wireless sub-channel because of different number of information slots available in each. On the other hand, the Aloha delay $D_{aloha}(h,i)$ is the same in both scenarios because it uses the same slots which are always part of the wireless to wireless sub-channel slots.

Again, the average inbound wireless to wired message delay $\overline{D_{ws\_wd}^{in}(h,i)}$ is given by:

$$\overline{D_{ws\_wd}^{in}(h,i)} = \sum_{h=0}^{U_{ws}} \sum_{i=0}^{U_{ws}-h} D_{ws\_wd}^{in}(h,i).P(h,i)$$

(5-20)

What is remaining now for this scenario is the CSMA/CD delay. This will be calculated using the expression given in [SCHWARTZ87] and derived originally in [BUX81]. In order to be able to use this expression, the wireless to wired equivalent traffic $\boldsymbol{l}_{ws\_wd}$ must be calculated. This traffic is the number of packets transmitted from the access point to the CSMA/CD network per second. This is equal to the number of packets successfully transmitted in the inbound wireless to wired sub-channel per second.

If we define $S_{ws\_wd}^{in}(h,i)$ as the number of messages successfully transmitted in the inbound wireless to wired sub-channel given that the system is at state *(h, i)* then from Little's formula on can write:

$$S_{ws\_wd}^{in}(h,i) = \frac{(U_{ws} - i)R_{ws\_wd}}{D_{ws\_wd}^{in}(h,i)}$$

(5-21)

From the definition of $\boldsymbol{l}_{ws\_wd}$ it is equal to the average number of messages successfully transmitted in the inbound wireless to wired sub-channel multiplied by the average number of packets per message. Accordingly, it is given by:

$$\boldsymbol{l}_{ws\_wd} \equiv \overline{S_{ws\_wd}^{in}(h,i)}\frac{N}{\boldsymbol{s}.F}$$

(5-22)

where:

$$\overline{S_{ws\_wd}^{in}}(h,i) = \sum_{h=0}^{U_{ws}} \sum_{i=0}^{U_{ws}-h} S_{ws\_wd}^{in}(h,i).P(h,i) \tag{5-23}$$

Total offered traffic to the CSMA/CD channel $\mathbf{1}^{tot}{}_{wd}$ is equal to the sum of that offered by the wired stations and the wireless to wired traffic. Hence, one can write:

$$\mathbf{1}_{wd}^{tot} = U_{wd}.\mathbf{1}_{wd} + \mathbf{1}_{ws\_wd} \tag{5-24}$$

Achieving an expression for the total offered traffic to the CSMA/CD, the delay on the CSMA/CD network D_{wd_wd} can be calculated from the following expression:

$$D_{wd\_wd} = \left[r \frac{\frac{\overline{m^2}}{\overline{m}^2} + (4e+2)a + 5a^2 + 4e(2e-1)a^2}{2(1-r(1+(2e+1)a))} + 1 + 2e.a - \frac{\left(1-e^{-2a.r}\right)\left(\frac{2}{r}+2a.e^{-1}-6a\right)}{2\left(F_p(\mathbf{1}).e^{-r.a-1}-1+e^{-2r.a}\right)} + \frac{a}{2} \right]\overline{m} \tag{5-25}$$

where:

$$r = \mathbf{1}_{wd}^{tot}.\overline{m}$$

$$a = \frac{t}{\overline{m}}$$

m = packet length in seconds

t = end to end propagation delay on the wire medium

$F_p(\mathbf{1})$ = Laplace transform of the packet length distribution

$$F_p(\mathbf{1}) = \int_0^\infty f(m).e^{-\mathbf{1}_{wd}^{tot}.m} dm$$

$f(m)$ = packet length distribution

$$\overline{m^2} = \text{second moment of the packet length distribution} = \int_0^\infty f(m)(m-\overline{m})^2 dm$$

$$\overline{m} = \text{first moment of the packet length distribution} = \text{mean} = \int_0^\infty f(m).m dm$$

Assuming a fixed constant packet length in our analysis, the following are true:

$$f(m) = \mathbf{d}(m-\overline{m})$$

$$\overline{m^2} = \overline{m}^2$$

$$F_p(\mathbf{1}) = \int_0^\infty \mathbf{d}(m - \overline{m}) e^{-I_{wd}^{tot}.m} dm = e^{-I_{wd}^{tot}.\overline{m}} = e^{-r}$$

Accordingly, equation (5-25) becomes:

$$D_{wd\_wd} = \left[\mathbf{r} \frac{1 + (4e+2)a + 5a^2 + 4e(2e-1)a^2}{2(1 - \mathbf{r}(1 + (2e+1)a))} + 1 + 2e.a - \frac{\left(1 - e^{-2a.r}\left(\dfrac{2}{\mathbf{r}} + 2a.e^{-1} - 6a\right)\right)}{2\left(e^{-r-r.a-1} - 1 + e^{-2r.a}\right)} + \frac{a}{2} \right] \overline{m}$$

$$(5\text{-}26)$$

This delay formula is valid only under a stability condition that is necessary for the steady state solution to exist. This necessity is a direct implication of the assumption that packets are retransmitted until they are successfully sent with no upper bound on number of retransmission retrials. The stability condition is as follows:

$$\mathbf{r} < \frac{1}{1 + a(1 + 2e)} \tag{5-27}$$

Finally, the wireless to wired average packet delay D_{ws_wd} is given by:

$$D_{ws\_wd} = \overline{D_{ws\_wd}^{in}(h,i)} \frac{\mathbf{s}.F}{N} + D_{wd\_wd} \tag{5-28}$$

5.6.3.   Wired to wired packet delay D_{wd_wd}

The expression for wired to wired packet delay D_{wd_wd} is already given in the previous section by equation (5-26).

5.6.4.   Wired to wireless packet delay D_{wd_ws}

In this scenario packets suffer from two types of delays. The first one is the CSMA/CD delay given by equation (5-26) till they reach the wireless base station. The second is the delay waiting in the base station wired to wireless queue till they are served by slots in the wired to wireless outbound sub-channel and the transmission delay for the packet. Since packets are already there in the base station after succeeding in the CSMA/CD process, there is no need to perform any reservation process here. Actually, the base station will service the packets using free slots in the outbound wired to wireless sub-channel in a first come first serve discipline. Accordingly, the system can be modeled as an M/M/n queuing model with n (the number of servers in the system) equal to the number of slots in the outbound wired to wireless

sub-channel. This means that $n=N_{wd\_ws}=N_{ws\_wd}$. The delay of an M/M/n queue is analyzed in many queuing books such as [BERTSEKAS87]. The expression for the delay in this queue D^{out}_{queue} is given by:

$$D^{out}_{queue} = \frac{1}{m_{MMn}} + \frac{p_0 (n.r_{MMn})^n}{n!(1 - r_{MMn})} \cdot \frac{1}{n.m_{MMn} - l_{MMn}}$$ (5-29)

where:

$$r_{MMn} = \frac{l_{MMn}}{n.m_{MMn}}$$

$$p_0 = \frac{1}{\sum_{i=0}^{n-1} \frac{(n.r_{MMn})^i}{i!} + \frac{(n.r_{MMn})^n}{n!}}$$

$$n = N_{ws\_wd}$$

m_{MMn} = service rate of the server in packets per second. Each server is a slot in the outbound wired to wireless sub-channel. Hence, the service rate is given by:

$$m_{MMn} = \frac{1}{F_t}$$ (5-30)

l_{MMn} = arrival rate at the input of the queue. This is equal to the portion of the wired generated traffic directed to wireless nodes. This portion is represented by the factor R_{wd_ws}. Accordingly, l_{MMn} is given by:

$$l_{MMn} = U_{wd} l_{wd} R_{wd\_ws}$$ (5-31)

The delay expression given in equation (5-29) is valid only under a certain stability condition. This stability condition is necessary for an existence of a steady state solution under the assumption of infinite queue length. This condition is given by:

$$r_{MMn} < 1$$ (5-32)

The total wired to wireless packet delay D_{wd_ws} is thus given by:

$$D_{wd\_ws} = D_{wd\_wd} + D^{out}_{queue}$$ (5-33)

Finally, we need to get the total average system delay regardless of the type of the traffic. This is the issue of the next section.

5.6.5. Total average system delay D

The total average system delay will be calculated by summing the delays of all the possible scenarios with each of them multiplied by the probability that the scenario occurs. These probabilities are the probabilities that the traffic generated from each source type (wireless and wired) is directed to a destination type (wireless or wired). These are represented by the ratios R_{ws_wd}, R_{wd_ws} and their complements. Accordingly, D is given by:

$$D = D_{ws\_ws}\left(1 - R_{ws\_wd}\right) + D_{ws\_wd} R_{ws\_wd} + D_{wd\_wd}\left(1 - R_{wd\_ws}\right) + D_{wd\_ws} R_{wd\_ws}$$

$$(5\text{-}34)$$

5.6.6. Throughput calculations

The throughput in packets per second S_{ws_ws} of the wireless to wireless scenario can be got from Little's formula noting that the queue length is the number of users intend to transmit to wireless stations. Hence, we have:

$$S_{ws\_ws} = \frac{N}{\boldsymbol{s}.F} \sum_{h=0}^{U_{ws}} \sum_{i=0}^{U_{ws}-h} \frac{(U - i)(1 - R_{ws\_wd})}{D_{ws\_ws}(h,i)}.P(h,i) \qquad (5\text{-}35)$$

The throughput of the wireless to wired scenario S_{ws_wd} is equal to the total offered load from the wireless nodes on the wired part. This is given by:

$$S_{ws\_wd} = \boldsymbol{l}_{ws\_wd} \qquad (5\text{-}36)$$

On the other hand, the throughput of the wired to wired scenario S_{wd_wd} is the same as the total fresh offered traffic to the CSMA/CD system from wired users. This is given by:

$$S_{wd\_wd} = U_{wd} \boldsymbol{l}_{wd}\left(1 - R_{wd\_ws}\right) \qquad (5\text{-}37)$$

Finally, the throughput of the wired to wireless scenario is the same as the total offered traffic to the wired to wireless message queue (the M/M/n queue). This is given by:

$$S_{wd\_ws} = \boldsymbol{l}_{MMn} \qquad (5\text{-}38)$$

In contrary to the case for the delay, the total average throughput S is the summation of these scenario throughputs directly with no weighting needed.

$$S = S_{ws\_ws} + S_{ws\_wd} + S_{wd\_wd} + S_{wd\_ws} \qquad (5\text{-}38)$$

5.7. <u>Numerical Results</u>

From the previous analysis we have arrived at closed form equations for the delay of the system. These closed form equations however depend on the states of the wireless sub-network represented by the variables h, i, and k. This dependency is removed by performing the averaging over the joint probability $P(h,i)$ and using the relation $k=U_{ws}\text{-}h\text{-}i$. Doing so, we achieve an equation for the total delay that depends on the system configuration (such as U_{ws}, and U_{wd}) and on the traffic characteristics (such as λ_{ws}, λ_{wd}, R_{ws_wd}, R_{wd_ws}). The numerical results for the delay can be computed from equations (5-16), (5-28), (5-26), (5-33), and (5-34). In conjunction with these equations there are two stability conditions that must be satisfied for the system to have a steady state stable solution. These conditions are represented by the inequalities (5-27), and (5-32).

In the following graphs we will show how the system performance and stability are affected by the different system parameters. We will also try to do some optimization for the system delay using some of those parameters.

In the following curves $U_{ws}=U_{wd}=10$ users, $\lambda_{wd}=1500$ packets/sec, $\lambda_{ws}=1500$ messages/sec, $F=2\times10^{-4}$ sec, $N_{RMIN}=2$ slots, $N_{MS}=100$ slots, $N=6$ slots, $s=2$, $p_t=0.5$, $P_{nerr}=1$, $Ack=3$ packets.
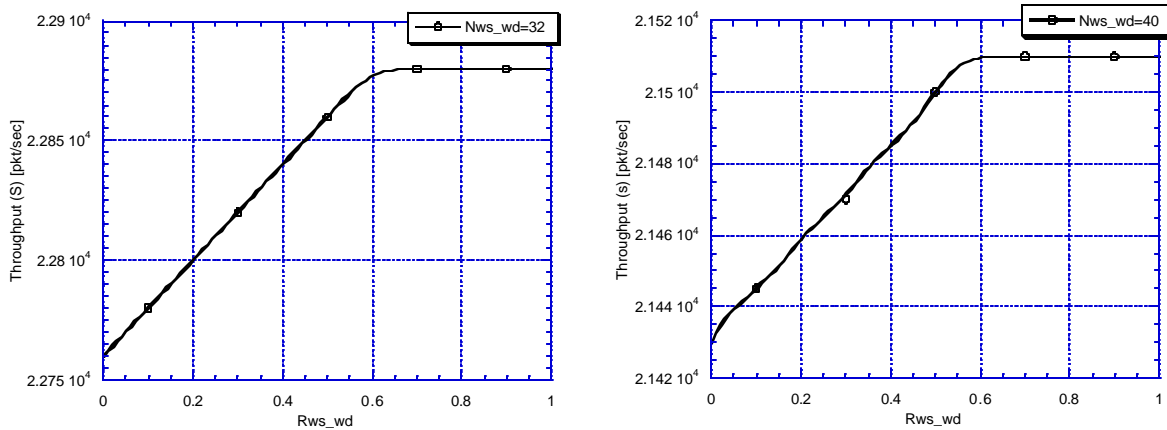
Fig 5-12 Throughput S in packets per second against R_{ws_wd} for $R_{wd\_ws}=0.5$ and two different values of N_{ws_wd}
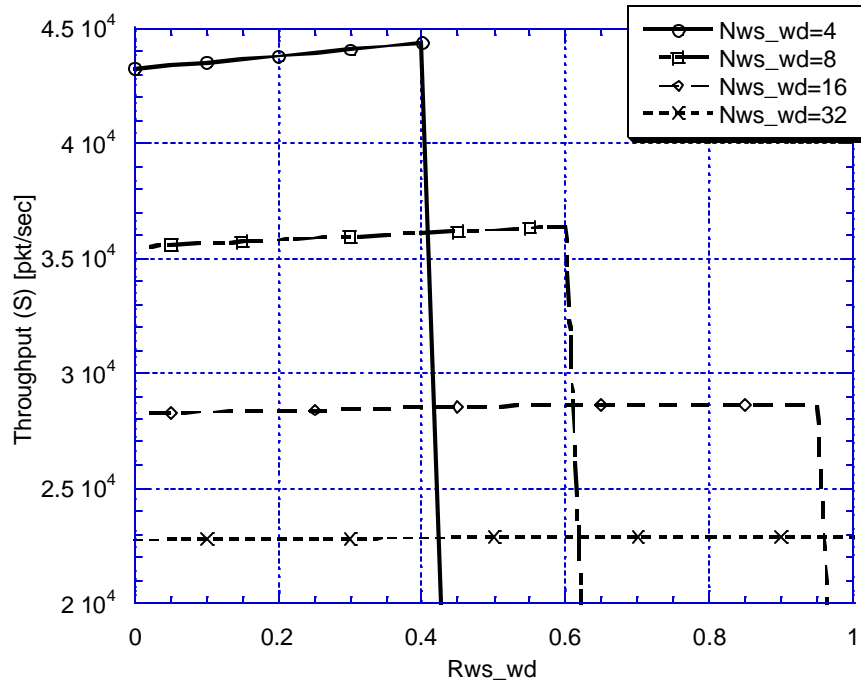
Fig 5-13 Throughput S in packets per second as a function of R_{ws_wd} for different values of N_{ws_wd} and for R_{wd_ws}=0.5
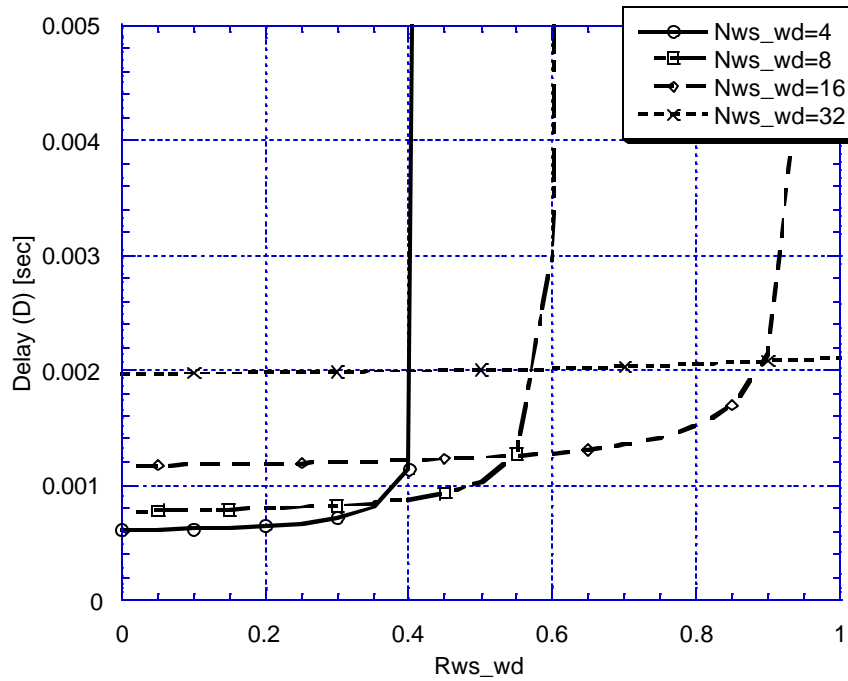
Fig 5-14 Delay D in seconds as a function of R_{ws_wd} for different values of N_{ws_wd} and for R_{wd_ws}=0.5

145
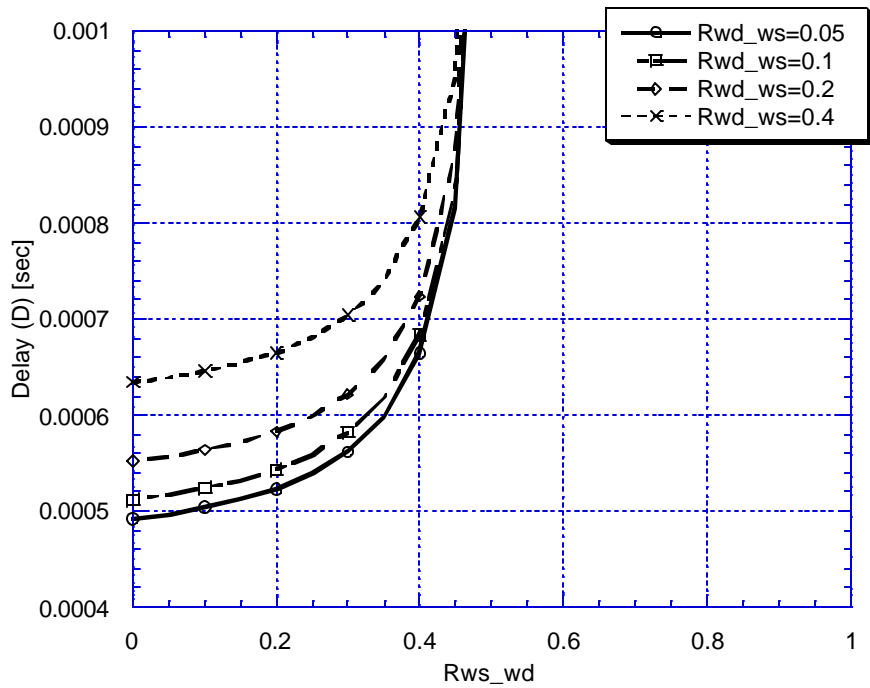
Fig 5-15 Delay D in seconds against R_{ws_wd} for different values of R_{wd_ws} and $N_{ws\_wd}=6$
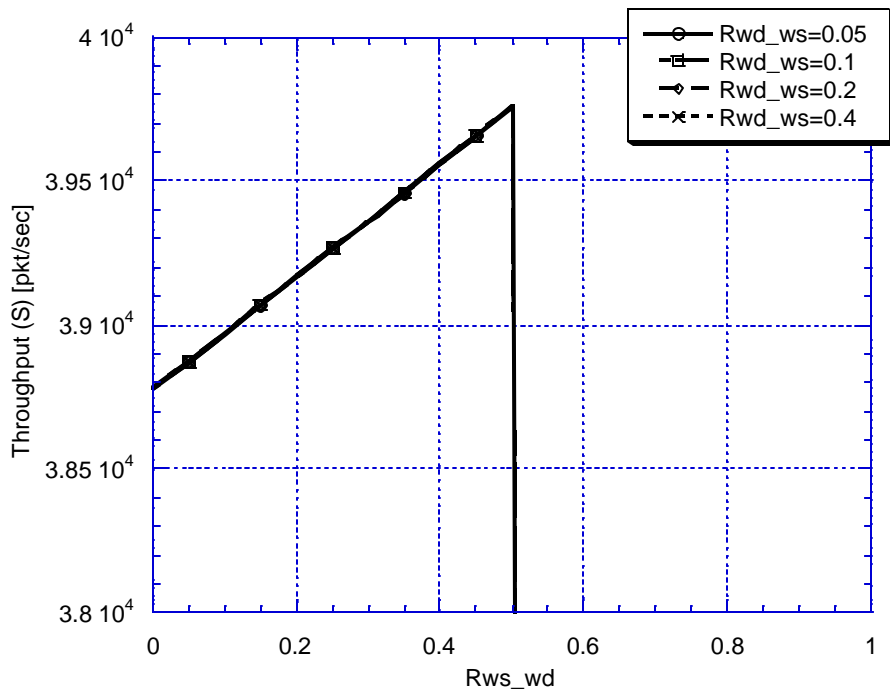
Fig 5-16 Throughput S in packets per second against R_{ws_wd} for different values of

R_{wd_ws} and $N_{ws\_wd}=6$
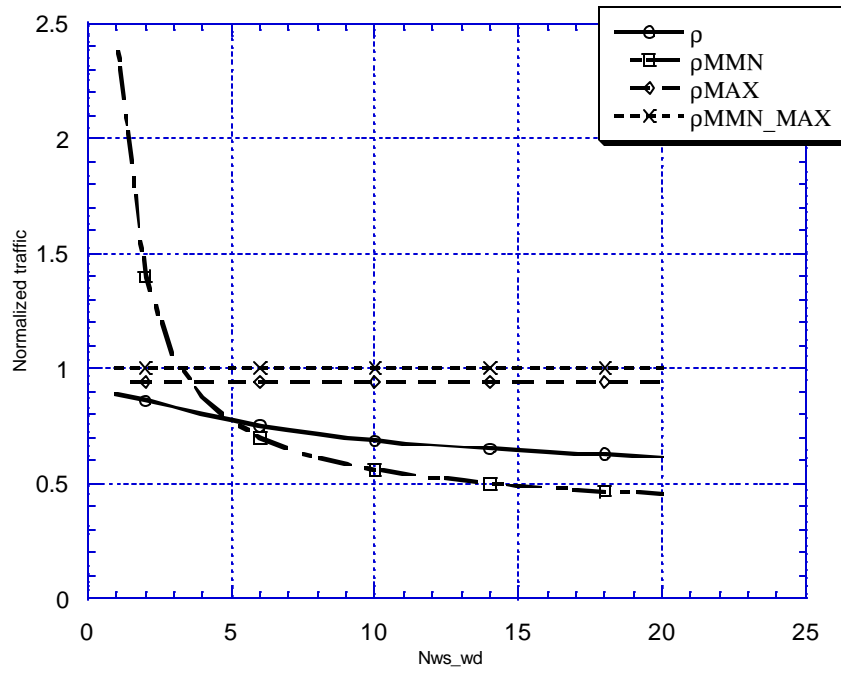
Fig 5-17 Normalized traffics offered to CSMA/CD *r* and offered to outbound wired to wireless sub-channel *r*$_{MMn}$ and their maximum values *r*$_{MAX}$ and *r*$_{MMn\_MAX}$ respectively, for stable operation against N_{ws_wd} for R_{ws_wd}=0.3, and R_{wd_ws}=0.7
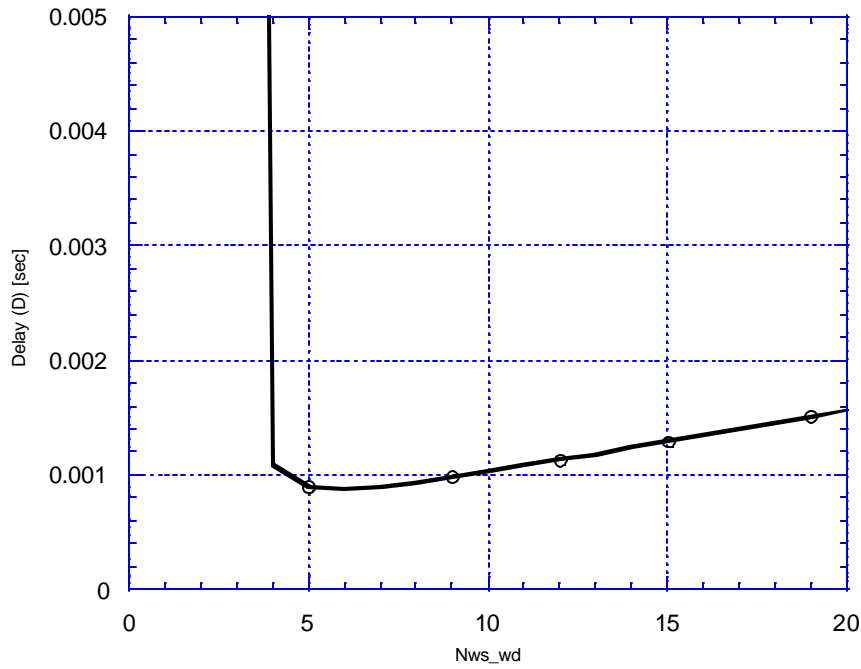
Fig 5-18 Delay *D* in seconds against N_{ws_wd} for R_{ws_wd}=0.3, and R_{wd_ws}=0.7
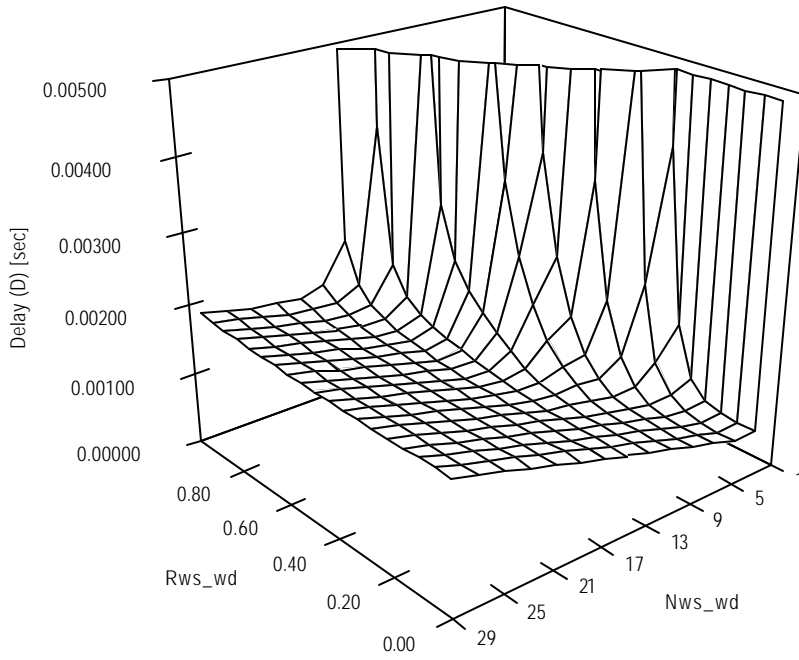
Fig 5-19 Delay *D* in seconds as a function of R_{ws_wd} and N_{ws_wd} for R_{wd_ws}=0.5
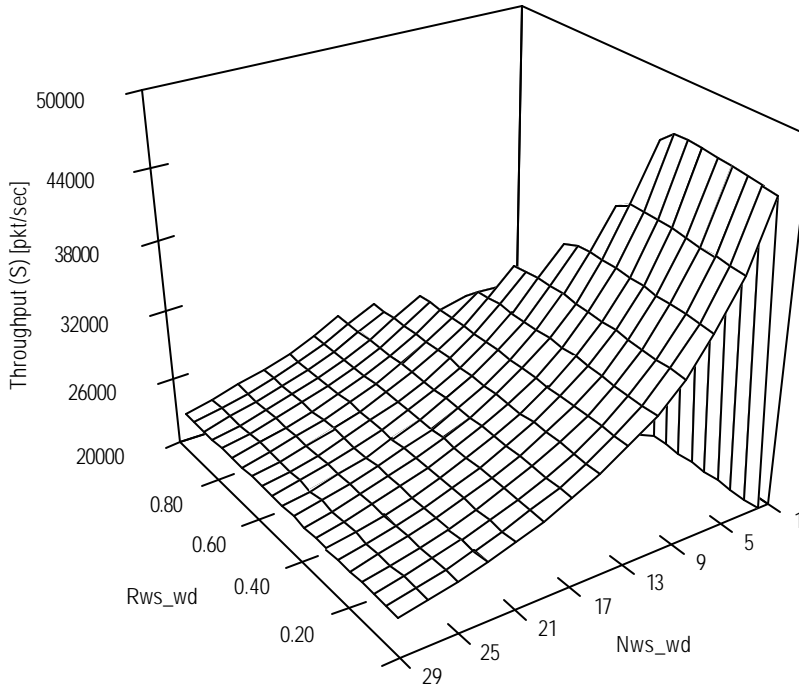
Fig 5-20 Throughput *S* in packets per second as a function of R_{ws_wd} and N_{ws_wd} for

R_{wd_ws}=0.5

148
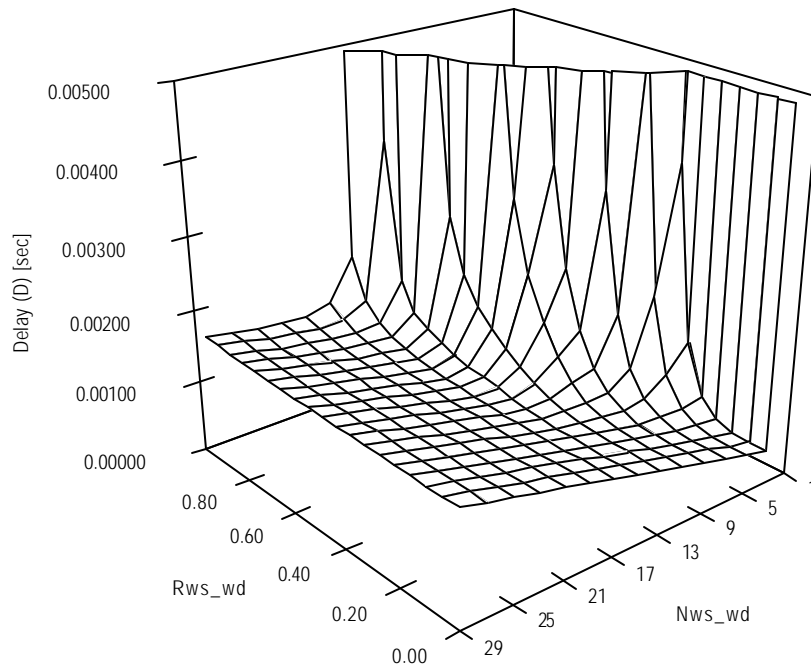
Fig 5-21 Delay *D* in seconds as a function of R_{ws_wd} and N_{ws_wd} for R_{wd_ws}=0.3
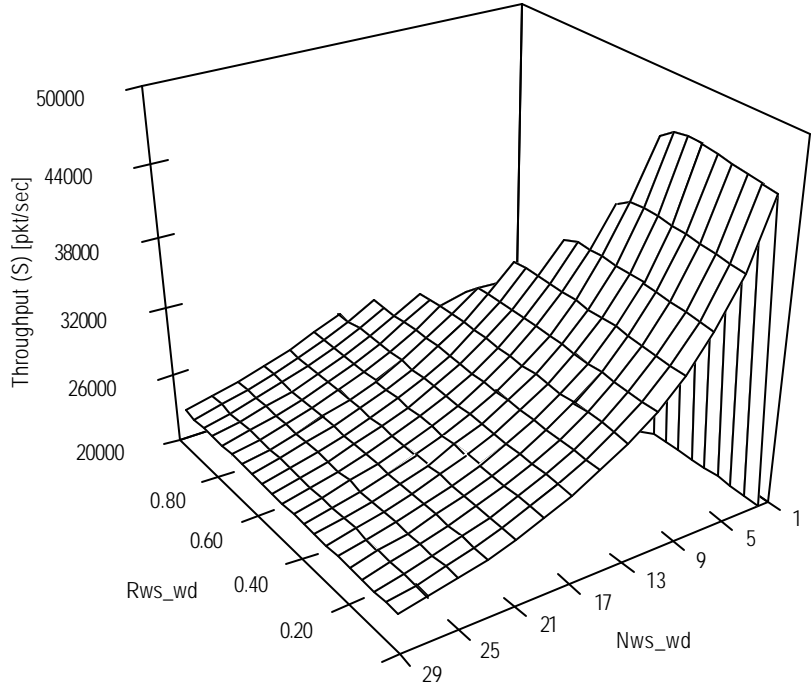
Fig 5-22 Throughput *S* in packets per second as a function of R_{ws_wd} and N_{ws_wd} for

R_{wd_ws}=0.3

5.8.   Comments on the Numerical Results

From the numerical results plotted in the previous section the following remarks can be deduced:

From Fig 5-12 and Fig 5-13 it is noticed that increasing R_{ws_wd} increases the throughput, but, as can be seen from Fig 5-12, this increase reaches a saturation value at a certain value of R_{ws_wd} if instability did not occur first. The value of R_{ws_wd}, at which saturation occurs, changes with changing N_{ws_wd}. The increase in throughput is due to the use of more slots of the wireless to wired channel slots N_{ws_wd}. This means that the wireless offered traffic is making more efficient use of the available bandwidth. On the other hand, this enhancement of performance is for lower loads on the wireless to wired inbound sub-channel. When the load on this sub-channel reaches its maximum value i.e. uses all available slots no further enhancement of the throughput is possible. Hence, the throughput saturates. On the other hand, the saturation value depends on N_{ws_wd} as this value affects the traffic that goes from wireless network to the wired part causing it to decrease in terms of packets per second with increasing N_{ws_wd}.

From Fig 5-14 one can also observe that increasing R_{ws_wd} increases the delay. This means that it makes the system performance worse. This degradation in performance continues on till the system reaches the unstable region where the throughput goes to zero and delay goes unbounded to infinity.

As expected, increasing the value of N_{ws_wd} increases the system stability. This is because it increases the frame size, which means it decreases the rate of packets coming from wireless nodes to wired nodes and hence decreases the load on the CSMA/CD protocol.

From Fig 5-15, and Fig 5-16 it is noticed that changing R_{ws_wd} does not affect the stability threshold of the system.   This is because mostly the system becomes unstable from the wireless to wired traffic rather than from the wired to wireless traffic for the given value of N_{ws_wd}. On the other hand, changing the values of R_{ws_wd}, N_{ws_wd}, and I_{ws} can get us to a point where R_{wd_ws} do affect the system stability by overloading the M/M/n queue of the wired to wireless traffic.

Also from the same figure one can notice that increasing R_{wd_ws} does not affect the stable region throughput.   This is because both CSMA/CD and M/M/n queue of the wired to

150

wireless traffic are having a throughput that is equal to the input traffic as long as they are in the stable region. Accordingly, studying the system throughput for only R_{ws_wd} is enough.

It is also noticed from Fig 5-15 that increasing either R_{ws_wd} or R_{wd_ws} increases the average value of the delay. This is because the most delayed messages are those transferred through the two sub-networks. This is expected as those messages suffer from both sub-networks delays.

Looking at Fig 5-17 one can notice the two stability controlling variables (r and r_{MMN}) and their maximum possible values (r_{MAX} and r_{MMN_MAX}) before the system becomes unstable. From this figure it is noticed that increasing N_{ws_wd} decreases both variables and hence, makes the system more stable. It is further noticed that the value of r_{MMN} decreases faster with increasing N_{ws_wd}. This means that stability depending on this condition is faster to reach with increasing N_{ws_wd}. Instability of the system starts whenever any of these variables crosses its upper limiting maximum value.

On the other hand, Fig 5-18 shows that increasing N_{ws_wd} decreases the delay up to a certain minimum value after which it increases again. It is this value that can be adaptively chosen by the base station depending on the system traffic to guarantee least possible delay in the stable operation region.

The 3-D diagrams in Fig 5-19, Fig 5-20, Fig 5-21, and Fig 5-22 show that the best operating point for the system is for the value of N_{ws_wd} that just satisfies the stability conditions. This is because increasing N_{ws_wd} further than this value increases the system delay. On the other hand, it decreases the throughput. This fact is valid for different values of R_{ws_wd} and R_{wd_ws}.

Again the throughput 3-D diagram does not change with changing R_{wd_ws} only, which assures our conclusion about the effect of this parameter on the system throughput.

5.9.   Conclusions

From the above, we conclude that the protocol proposed in this chapter, which integrates two already known protocols with some modifications can be used to interconnect wireless and wired LANs with an acceptable performance under certain traffic limitation and constraints.

Analysis of this protocol, as well as any other proposed hybrid wireless-wired LAN, can be done using already established analysis techniques of each sub-network of the hybrid LAN

Analysis of this protocol, as well as any other proposed hybrid wireless-wired LAN, can be done using already established analysis techniques of each sub-network of the hybrid LAN under certain assumptions. This type of analysis gives a good idea for system engineers on the expected worst case performance in such hybrid connection.

It must also be noticed that although the system is assumed to be loss-less, the throughput changes with changing N_{ws_wd} and R_{ws_wd}. This is happening because of the fact that the PRMA technique is analyzed assuming that the arrival rate l_{ws} is a conditional arrival conditioned on station being idle. This means that is if a station is idle, it will wait on the average $1/l_{ws}$ before an arrival occurs to it. On the other hand, if a station is active it blocks all new arrivals to it till it becomes active again. This means that at any point in time the average non-conditional arrival rate per station is dependent on the service time of this station. This is why changing the service time changes the system throughput by changing the input traffic to the system, but the system stays to be a loss-less system.

From the analysis we also found that this type of network can have better performance in terms of decreasing the delay to a minimum value by using adaptive calculation algorithms of the system parameters such as N_{ws_wd}. These adaptive algorithms can run on the base station (access point) and their results transformed to the other stations if necessary, as the base station is the traffic controller in this system anyway.

Chapter 6

Summary and Future Work

Chapter 6
Summary and Future Work

6.1. <u>Summary</u>

In this thesis, we have given an introduction to the concept and needs for wireless LANs in Chapter 1. Then, in Chapter 2 we have given an overview of the different technologies involved in this area. In this overview we explained the different solutions used in the physical layer and the medium access control techniques used for wireless LANs with a list of the advantages and disadvantages for each alternative. It was found from Chapter 2 that the big challenges facing wireless LANs are the need for standardization and the lack of wireless frequency band allocation, especially for RF WLANs.

In Chapter 3, we moved to explain the standardization effort done to achieve interoperability between different wireless LAN products from different vendors. Two main standardization organizations are working in the field; the IEEE in the US and ETSI in Europe. They have produced in the late 90's stable draft of their standard specifications. These standards are called IEEE 802.11 and ETSI HIPERLAN, respectively.

In Chapter 4 we proposed an analysis for the CSMA/CA protocol and numerical results were obtained from this analysis. From the analysis done on CSMA/CA we conclude that :

The protocol gives good performance in terms of throughput and delay for low load conditions.

The protocol is not suitable for time critical applications such as audio and video applications as there are no guarantee for an upper bound of the delay.

The analysis using SSS gave results that are near to those obtained via simulation which indicates that the technique is accurate, yet easier mathematically than traditional techniques and can be used to analyze other protocols.

In Chapter 5 we consider a composite network consisting of a wireless LAN using the PRMA access protocol, which is interconnected with a wired LAN using the CSMA/CD protocol. For the composite network to operate, we modified the model for the access point. We then

used established results in the literature for each individual network, to evaluate the performance of the combined network. From the analysis done on Hybrid PRMA-CSMA/CD we concluded that:

Interconnection between wireless and wired LANs using the Hybrid PRMA-CSMA/CD protocol proposed in this thesis is possible and gives adequate performance results.

Analysis of hybrid wireless/wired interconnected LANs can be obtained approximately from the analysis of each sub-network in isolation if suitable assumptions are made.

This type of analysis is an approximate one and gives an idea of the expected worst case performance of the system under different traffic conditions.

The analysis showed that some parameters of the system can be chosen to maximize the system performance while assuring stability. This suggests that these parameters can be calculated adaptively during operation to assure best performance under different load conditions.

6.2. <u>Future Work</u>

Some of the work that can be done in the future include:

1. The effect of using handshaking techniques (RTS/CTS) in CSMA/CA could be studied using Single Station Superposition (SSS) technique.

2. The performance of CSMA/CA could be studied with the effect of channel parameter, hidden terminal problem, and capture effect using the same SSS technique. The effect of these parameters in this case will appear in the probability of collision through a different channel state model.

3. The performance of interconnecting PRMA with CSMA/CD could be studied assuming adaptive calculations of the inbound wireless-to-wired sub-channel number of slots (N_{ws_wd}) in terms of the traffic ratios R_{ws_wd} and R_{wd_ws} to give the best performance.

4. The performance of interconnecting CSMA/CA with CSMA/CD could be studied using the same approach used in Chapter 5 and reusing the results of CSMA/CA given in Chapter 4.

5. The performance of using PCF and DCF, together, in IEEE 802.11 could be analyzed using SSS again and some of the techniques used in Chapter 5 for hybrid protocols.

6. The performance of NPMA used in HIPERLAN can also be analyzed using SSS since only simulation results exist for this protocol.

Appendix A

Mathematical Analysis of CSMA/CA Protocol

Appendix A
Mathematical Analysis of CSMA/CA Protocol

Calculation of Normalized Throughput (*S*) and Delay (*D*)

The equilibrium equations for CSMA/CA model given in Chapter 4 are as follows:

$$B_{ij}\left(1-r_{ij}\right)=B_{i\,j+1} \qquad , \qquad 1\leq i\leq m, \qquad 0\leq j\leq V_i-2 \tag{A-1}$$

$$\left(1-\boldsymbol{s}\right)K_{m3}+\boldsymbol{1}.\boldsymbol{I}+\left(1-\boldsymbol{s}\right)K_3=A \tag{A-2}$$

$$\boldsymbol{s}.K_3+\boldsymbol{s}.K_{m3}=\boldsymbol{1}.\boldsymbol{I} \tag{A-3}$$

$$K_{i3}=\left(1-b\right)B_{i+1\,0} \qquad , \qquad 1\leq i\leq m-1 \tag{A-4}$$

$$\left(1-b\right)\left(1-f\right)\left[A+\sum_{i=1}^{m}\sum_{j=0}^{V_i-1}r_{ij}.B_{ij}\right]=T_1 \tag{A-5}$$

$$\left(1-b\right)f\left[A+\sum_{j=0}^{V_1-1}r_{1j}.B_{1j}\right]=C_{11} \tag{A-6}$$

$$\left(1-b\right)f\sum_{j=0}^{V_i-1}r_{ij}.B_{ij}=C_{i1} \qquad , \qquad 2\leq i\leq m \tag{A-7}$$

$$T_1=T_2=...T_i=...T_{L-1}=K_1=K_2=K_3 \tag{A-8}$$

$$C_{i1}=C_{i2}=...C_{ij}=...C_{iL-1}=K_{i1}=K_{i2}=K_{i3} \qquad , \qquad 1\leq i\leq m \tag{A-9}$$

$$I+A+\sum_{i=1}^{m}\sum_{j=0}^{V_i-1}B_{ij}+\sum_{i=1}^{L}T_i+\sum_{i=1}^{3}K_i+\sum_{i=1}^{m}\left[\sum_{j=1}^{L}C_{ij}+\sum_{j=1}^{3}K_{ij}\right]=1 \tag{A-10}$$

A first step in the solution is to give the expression for r_{ij}. This is given by the following expression:

$$r_{ij}=\frac{1}{V_i-j} \qquad , \qquad 1\leq i\leq m, \qquad 0\leq j\leq V_i-1 \tag{A-11}$$

Then from (A-1) one can write:

$$B_{i\,j+1}=B_{i0}\prod_{l=0}^{j}\left(1-r_{il}\right) \qquad , \qquad 1\leq i\leq m, \qquad 0\leq j\leq V_i-2 \tag{A-12}$$

Substituting (A-11) into (A-12) one obtains:

$$B_{i\,j+1} = B_{i0} \frac{V_i - (j+1)}{V_i}$$

Accordingly:

$$B_{ij} = B_{i0} \frac{V_i - j}{V_i} \quad , \qquad 1 \le i \le m, \qquad 0 \le j \le V_i - 1 \qquad \text{(A-13)}$$

Also, equation (A-9) implies that:

$$K_{i3} = C_{i1}$$

Then by substituting into (A-4) the following is obtained:

$$C_{i1} = (1-b)B_{i+1\,0} \quad , \qquad 1 \le i \le m-1 \qquad \text{(A-14)}$$

Substituting (A-13) into (A-7) it is found that:

$$C_{i1} = (1-b)f \sum_{j=0}^{V_i-1} \left(\frac{1}{V_i - j} \right) B_{i0} \left(\frac{V_i - j}{V_i} \right) \quad , \qquad 2 \le i \le m$$

which yields:

$$C_{i1} = (1-b)f \cdot B_{i0} \quad , \qquad 2 \le i \le m \qquad \text{(A-15)}$$

Substituting from (A-14) into (A-15) it can be shown that:

$$C_{i1} = f \cdot C_{i-1\,1} \quad , \qquad 2 \le i \le m$$

Hence:

$$C_{i1} = f^{i-1} \cdot C_{11} \quad , \qquad 2 \le i \le m \qquad \text{(A-16)}$$

Also, by substituting (A-13) and (A-11) into (A-6) it can be shown that:

$$C_{11} = (1-b)f\left(A + B_{10}\right) \qquad \text{(A-17)}$$

Substituting (A-17) into (A-16) yields:

$$C_{i1} = f^i (1-b)\left(A + B_{10}\right) \quad , \qquad 1 \le i \le m \qquad \text{(A-18)}$$

Also, substituting (A-6) and (A-7) into (A-5) yields:

$$T_1 = \left(\frac{1-f}{f} \right) \sum_{i=1}^{m} C_{i1} \qquad \text{(A-19)}$$

Using (A-18), equation (A-19) can be written as:

$$T_1 = \left(\frac{1-f}{f} \right)\left(A + B_{10}\right) f \left(\frac{1-f^m}{1-f} \right)(1-b)$$

Thus:

$$T_1 = \left(1 - f^m\right)\left(A + B_{10}\right)\left(1 - b\right) \tag{A-20}$$

From equations (A-8) and (A-9) it is obvious that:

$$K_3 = T_1 \qquad \& \qquad K_{m3} = C_{m1}$$

Substituting these values and equation (A-18) into (A-2) and (A-3) it is found that:

$$\left(1 - s\right)f^m\left(1 - b\right)\left(A + B_{10}\right) + \mathbf{1}.I + \left(1 - s\right)T_1 = A \tag{A-21}$$

$$s.T_1 + s.f^m\left(1 - b\right)\left(A + B_{10}\right) = \mathbf{1}.I \tag{A-22}$$

But using (A-13) one can obtain:

$$\sum_{i=1}^{m}\sum_{j=0}^{V_i-1} B_{ij} = \sum_{i=1}^{m} B_{i0} \sum_{j=0}^{V_i-1} \frac{V_i - j}{V_i} = \sum_{i=1}^{m} B_{i0}.\frac{V_i + 1}{2}$$

Also from (A-15) the above expression can be rewritten to give:

$$\sum_{i=1}^{m}\sum_{j=0}^{V_i-1} B_{ij} = \frac{V_1 + 1}{2}B_{10} + \sum_{i=2}^{m} \frac{C_{i1}}{f\left(1 - b\right)}.\frac{V_i + 1}{2}$$

Substituting from (A-18) into the above expression yields:

$$\sum_{i=1}^{m}\sum_{j=0}^{V_i-1} B_{ij} = \frac{V_1 + 1}{2}B_{10} + \sum_{i=2}^{m} \frac{f^i\left(1 - b\right)\left(A + B_{10}\right)}{f\left(1 - b\right)}.\frac{V_i + 1}{2}$$

$$= \frac{V_1 + 1}{2}B_{10} + \sum_{i=2}^{m} f^{i-1}\left(A + B_{10}\right).\frac{V_i + 1}{2}$$

Replacing each i by $i+1$ in the summation at the right hand side yields:

$$\sum_{i=1}^{m}\sum_{j=0}^{V_i-1} B_{ij} = \frac{V_1 + 1}{2}B_{10} + \sum_{i=1}^{m-1} f^i\left(A + B_{10}\right).\frac{V_{i+1} + 1}{2}$$

For a binary exponential back-off V_i takes the expression:

$$V_i = 2^i.CW_{\min}$$

Substituting this expression into the summation above yields:

$$\sum_{i=1}^{m}\sum_{j=0}^{V_i-1} B_{ij} = \frac{2.CW_{\min} + 1}{2}B_{10} + f\left(A + B_{10}\right)\left(\frac{2.CW_{\min} - 2.CW_{\min}\left(2f\right)^{m-1}}{1 - 2f} + \frac{1 - f^{m-1}}{2 - 2f}\right)$$

$$\tag{A-23}$$

In the same manner it can be also deduced that:

$$\sum_{i=1}^{m}\left[\sum_{j=1}^{L} C_{ij} + \sum_{j=1}^{3} K_{ij}\right] = \left(L + 3\right)\sum_{i=1}^{m} C_{i1} = f\left(L + 3\right)\left(1 - b\right)\left(A + B_{10}\right)\left(\frac{1 - f^m}{1 - f}\right) \tag{A-24}$$

Substituting (A-22), (A-23) and (A-24) into (A-10), and performing some mathematical manipulations, the normalizing equation can be rewritten to give:

$$
\begin{aligned}
&\frac{s}{l}T_1 + \frac{s}{l}f^m(1-b)(A+B_{10}) \\
&+\frac{2CW_{min}+1}{2}B_{10} + f(A+B_{10})\left[\frac{2CW_{min}-2CW_{min}(2f)^{m-1}}{1-2f} + \frac{1-f^{m-1}}{2-2f}\right] \\
&+(L+3)T_1 \\
&+f(L+3)(1-b)(A+B_{10})\left(\frac{1-f^m}{1-f}\right) = 1
\end{aligned}
\tag{A-25}
$$

Now equations (A-20), (A-21), and (A-25) form three linear equations in three state probability variables. Hence, solving equations (A-21) and (A-22) are together we get:

$$
T_1 + B_{10} + \left[f^m(1-b)-1\right](A+B_{10}) = 0
\tag{A-26}
$$

Also, from (A-20) it can be shown that:

$$
(A+B_{10}) = \frac{T_1}{(1-f^m)(1-b)}
\tag{A-27}
$$

Substituting (A-27) into (A-26) and simplifying, the following formula is obtained:

$$
B_{10} = \frac{b}{(1-b)(1-f^m)}T_1
\tag{A-28}
$$

Accordingly, by substituting (A-27), and (A-28) into (A-25), and solving for T_1 one obtains:

$$
T_1 = \frac{1}{\frac{s}{l}+(L+3)+\frac{1}{(1-b)(1-f^m)}\left[1+\frac{2CW_{min}-1}{2}.b+\frac{s}{l}f^m(1-b)+f\left(\frac{2CW_{min}-2CW_{min}(2f)^{m-1}}{1-2f}+\frac{1-f^{m-1}}{2-2f}\right)+f(L+3)(1-b)\left(\frac{1-f^m}{1-f}\right)\right]}
\tag{A-29}
$$

What equation (A-29) gives is the value of T_1 in terms of the other system parameters and f and b. This is what needed to find throughput and delay values. The throughput for the whole system is given by the following expression:

$$
S = U.P_s
$$

where P_s is the throughput for one station, which is equal to the sum of the probabilities of being in a successful transmission state. Accordingly, the above expression can be written as follows:

$$
S = U.\sum_{i=1}^{L}T_i
$$

Substituting from (A-8) into the above expression one obtains:

$$S = U.L.T_1 \tag{A-30}$$

To complete the performance equatio ns an expression for the system delay in terms of $T1$, f, b and system parameters is needed. This expression will be derived in the following:

Substituting equation (A-20) into (A-22) one can write:

$$I = \frac{s}{I} T_1 + \frac{s}{I} f^m (1-b) \frac{T_1}{(1-b)(1-f^m)}$$

$$= \frac{s}{I} T_1 \left[1 + \frac{f^m}{1-f^m} \right] \tag{A-31}$$

From Little's formula the normalized delay for one station in terms of unit packet transmission time is given by:

$$D = \frac{1-I}{P_s}$$

Hence, the delay in slots is given by multiplying the above expression by the packet length L which yields:

$$D = \frac{1-I}{P_s}.L$$

Substituting (A-31) and the value of P_s into the above expression one can write:

$$D = \frac{1 - \frac{s}{I} T_1 \left(1 + \frac{f^m}{1-f^m} \right)}{T_1} \tag{A-32}$$

Calculating the Probabilities of Busy (b) and Collision (f)

Assuming that the total number of stations in the network is U, the probability of collision f is the probability that one or more of the other U-1 nodes transmits at the same time the current station is transmitting. This happens only if one of the other U-1 stations is ending its back-off or transmitting for the first time at the same time the current station is doing so. This is defined as follows:

f = Pr{One or more of the U-1 stations is deciding to transmit}

= Pr{One or more station is in A state $\cup$ (one or more station is in back-off

$\cap$ such station(s) end back-off)}

= 1-Pr{All U-1 stations are not deciding to transmit}

= 1-(Pr{Not deciding to transmit for one station})$^{U-1}$

= 1-(1-Pr{Deciding to transmit for one station})$^{U-1}$

= 1-(1-Pr{Station is in A $\cup$ (it is in $B_{ij} \cap$ it decides to end back-off)})$^{U-1}$

Accordingly, f is written as follows:

$$f = 1 - \left[1 - \left(A + \sum_{i=1}^{m} \sum_{j=0}^{V_i - 1} B_{ij} r_{ij} \right) \right]^{U-1}$$

Substituting from (A-5) into the above expression yields:

$$f = 1 - \left(1 - \frac{T_1}{(1-b)(1-f)} \right)^{U-1} \tag{A-33}$$

Next we find the probability b that the medium is busy. To do so we construct a Markov queuing model that represents the channel different states (Fig A-1).

O = Occupied with successful transmission

N = Occupied with collision

E = Empty (no transmission)
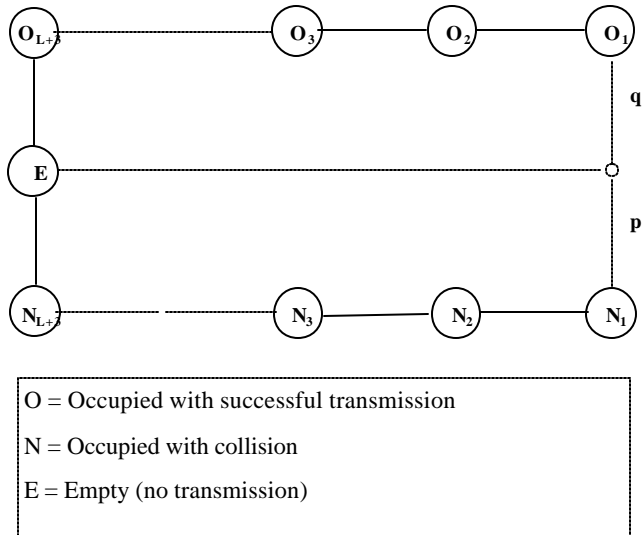
Fig A-1 Markov state model of the channel

The model shown in Fig A-1 represents the states of the channel as seen by an observing station. These states are empty E, occupied with successful transmission O_i, or occupied with collision N_i. The channel is empty when there is no node transmitting at all. It becomes occupied when one or more of the U-1 remaining nodes is transmitting. If only one node is transmitting, the channel goes to states O_i with probability q. In this case it is seen by the observing station as being busy and carrying a successful transmission. This continues for the L slots period of the packet plus the 3 slots period of the acknowledgment. On the other hand, if more than one node is transmitting, the channel goes to states N_i with probability p, in which case it is seen as being in collision. It must be noted however that an observing station does not take any actions towards this collision, and it is detected by the transmitting station by means of NACK or no ACK at all.

What is required here is to find the probability of finding the medium free at the sensing time. This probability is denoted by 1-b and is equal to the steady state probability of having the channel at state E of the Markov model given in Fig A-1. To find this probability, the steady state equations of the above queuing models must be solved. The equilibrium equations for the above model are (again we use the label of the state to represent the probability of being in the state):

$$O_1 = O_2 =...O_i...= O_{L+3} \tag{A-34}$$

$$N_1 = N_2 =...N_i...= N_{L+3} \tag{A-35}$$

$$O_1 = q.E \tag{A-36}$$

$$N_1 = p.E \tag{A-37}$$

and the normalization equation which will replace the equilibrium equation at state E is as follows:

$$E+\sum_{i=1}^{L+3} O_i + \sum_{i=1}^{L+3} N_i =1 \tag{A-38}$$

By substituting equations (A-34), (A-35), (A-36), and (A-37) into equation (A-38) it is found that:

$$E+(L+3).q.E +(L+3).p.E =1$$

This yields:

$$E = \frac{1}{1+(L+3)(q+p)} \tag{A-39}$$

Since the probability of finding the channel free (1-b) is the probability of being at state E, then:

$$1-b = \frac{1}{1+(L+3)(q+p)} \tag{A-40}$$

From the previous description q is the probability of only one station of the U-1 stations is transmitting. This is given by the following equation:

$$q = (U-1)\left(\sum_{i=1}^{L} T_i + \sum_{i=1}^{3} K_i \right)$$

This yields:

$$q = (U-1)(L+3)T_1 \tag{A-41}$$

On the other hand, p is the probability of two or more stations transmitting at the same time on the channel. This is given by:

$$p = 1 - (U-1)(L+3)T_1 - \left(I + A + \sum_{i=1}^{m}\sum_{j=0}^{V_i-1} B_{ij} \right)^{U-1}$$

Substituting equation (A-10) into the above expression yields:

$$p = 1 - (U-1)(L+3)T_1 - \left(1 - \left(\sum_{i=1}^{L} T_i + \sum_{i=1}^{3} K_i + \sum_{i=1}^{m}\left[\sum_{j=1}^{L} C_{ij} + \sum_{j=1}^{3} K_{ij} \right] \right) \right)^{U-1}$$

Using (A-20), (A-24) and the above expression one can write:

$$p = 1 - (U-1)(L+3)T_1 - \left[1 - (L+3)T_1 - (L+3)\left(\frac{f}{1-f} \right)T_1 \right]^{U-1} \tag{A-42}$$

Substituting (A-41) and (A-42) into (A-40) yields:

$$1-b = \frac{1}{1+(L+3)\left(1 - \left[1 - (L+3)T_1 - (L+3)\left(\frac{f}{1-f} \right)T_1 \right]^{U-1} \right)}$$

The above expression can be rewritten as follows:

$$b = 1 - \cfrac{1}{1 + (L+3)\left(1 - \left[1 - (L+3)T_1 - (L+3)\left(\dfrac{f}{1-f}\right)T_1\right]^{U-1}\right)} \tag{A-43}$$

Equations (A-29), (A-33), and (A-43) are three non linear equations in three variables $T_1, f,$ b. Assuming that all the system parameters like arrival rate l and number of users U take arbitrary values, these equations can be solved together numerically to get numerical values for T_1, f, b. These values can be used in the throughput equation (A-30) and delay equation (A-32) to obtain their values.

Calculation of Normalized Fresh Offered Load (G)

The fresh offered load presented to the medium by one station G_1^m in terms of messages per

G_1^m = average number of new messages arriving on station per second
= average number of new messages arriving per slot / slot duration
= probability of new arrival per slot / slot duration

second is defined by

This is given mathematically by:

$$G_1^m = \frac{l}{t} \qquad \text{messages/sec} \tag{A-44}$$

where t is the time slot duration in seconds.

To obtain the offered load for one station in terms of packets per second G_1 we need to multiply the above expression by the average number of packets per message which is equal to $1/s$. This results in:

$$G_1 = \frac{l}{t} \cdot \frac{1}{s} \qquad \text{packets/sec} \tag{A-45}$$

Finally, to normalize this expression against the channel capacity C given in bits per second we need to multiply G_1 by the message length in bits and then divide it over C. This gives the normalized fresh offered load per station. To get the total normalized fresh offered load G we multiply by the number of stations U to get G as follows:

$$G = \frac{\left(\dfrac{1}{t.s}\right)\left(\dfrac{L.t}{C}\right)}{C} U$$

which can be reduced to:

$$G = \frac{U.L.1}{s} \tag{A-46}$$

Appendix B

Packet Reservation Multiple Access (PRMA) Protocol

Appendix B
Packet Reservation Multiple Access (PRMA)
Protocol

Introduction

This protocol is a hybrid protocol, i.e. it uses both contention based, and contention free access methods. This technique is already adopted in the IBM wireless RF LAN [BAUCHOT95], and has proven to give a satisfactory performance compared to other techniques. Moreover, it has several advantages over techniques using only contention based MAC protocols. Some of these advantages will be stated after the description of the technique itself. In this appendix, we will repeat the analysis done for this protocol in [MITROU90]. In order to explain the system and the analysis in more details, a queuing model that describes the system operation is proposed. Also, based on the understanding of this model explanations that go beyond those found in [MITROU90] will be given to allow the reader to further understand the analysis equations. This is necessary as the work in Chapter 5 is based on the analysis given in this appendix.

Network Architecture

The main feature of a reservation multiple access technique is that it is a centralized technique i.e. all the traffic of the network goes through a central node; called the *base station* which will be denoted here as *BS*. The other stations will be named *workstations,* which will be denoted *WS*. Hence, the network architecture here is an infrastructure architecture with centralized data flow (Fig B-1).

Fig B-1 Network architecture of PRMA

BS=Base Station

WS=Workstation

Protocol Description

The protocol uses two different frequencies; one is for the *Inbound* (WS to BS) connection, and the other for the *Outbound* (BS to WS) connection. Both of the two channels are divided into frames, and each frame is divided into time slots. A time slot that is repeated periodically in all the frames constitutes a *logical channel*. Thus, for any session two logical channels are required; one for the inbound, and one for the outbound.

Each data source when active is assumed to use one time slot per frames i.e. one logical channel. The group of logical channels assigned for the actual information transfer is called information channels. On the other hand those channels used for slot-reservation, and other connection-control information are called control channels.

In the inbound link a number of uniformly distributed logical channels are marked as permanent reservation channels (R slots) as shown in Fig B-2. The total capacity of permanent reservation channels is what is called the *reservation capacity threshold*[5]. The other remaining slots are mainly information slots (I slots). In order to make the protocol more flexible, any one of the information slots can be temporarily used as an extra reservation channel (ER slots).
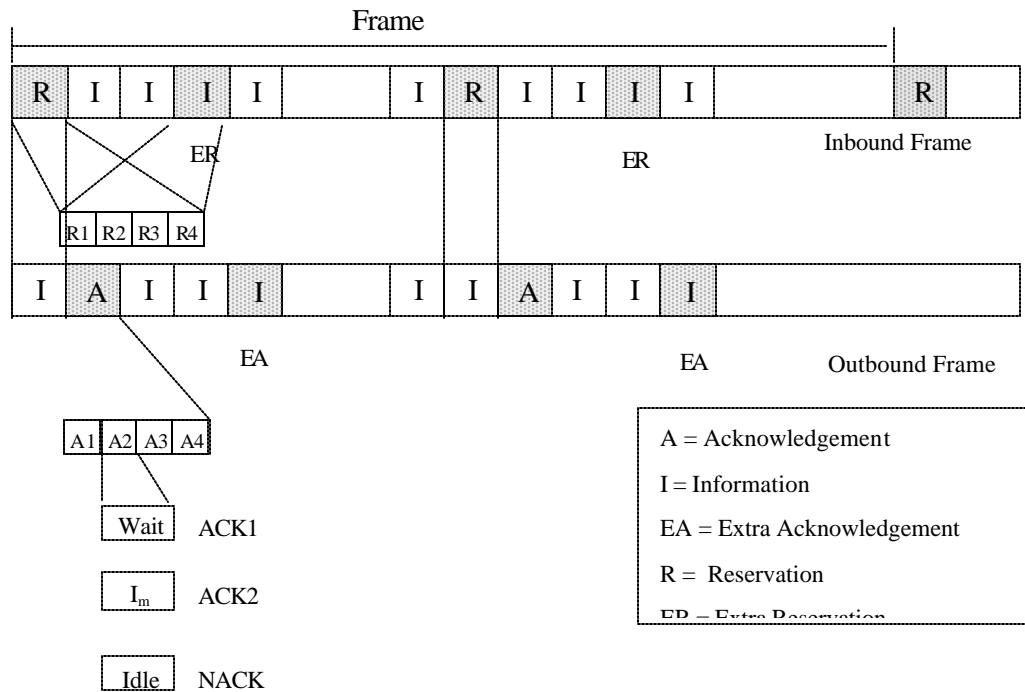
Fig B-2 Inbound and outbound time frames for PRMA

To make more efficient use of the available bandwidth each of the R and ER slots are divided into mini-slots each capable of carrying one reservation packet. In the outbound link the control channels carry out the acknowledgment packets (A slots). These packets are used for backward error control. The A slots follow the R slot in time as shown in Fig B-2. Again any of the I slots can be used as a temporary extra acknowledgment channel (EA slots). Also, both the A and the EA slots are divided into mini-slots. Each mini-slot can convey one ACK or NACK packet.

---

[5] These are the slots that guarantee that the contention-based part never vanishes from the frame time.

In order to describe the protocol a two state periodic source is assumed to be behind each connection. As one specific source changes state from "on" to "off" or vice versa, the corresponding connection passes through four states in a cyclic way. These states are shown in Fig B-3. The MAC procedure proceeds as follows (Fig B-4):

Each connected WS contends for an information slot every time the source behind it changes state from idle to active, by transmitting a short reservation packet on the first available reservation mini-slot with a certain transmission probability P_T. Upon successful transmission the WS is assigned an I slot that is reserved for it till the end of the current packet. If there are no information slot available, the connection is queued, till a slot becomes free which is represented by S3 in Fig B-3(b). If the reservation process is unsuccessful due to collision or due to transmission errors, the WS is backlogged, and it tries again with a certain retransmission probability P_R on the next reservation mini-slot. P_R is calculated for each system state to guarantee stable operation.
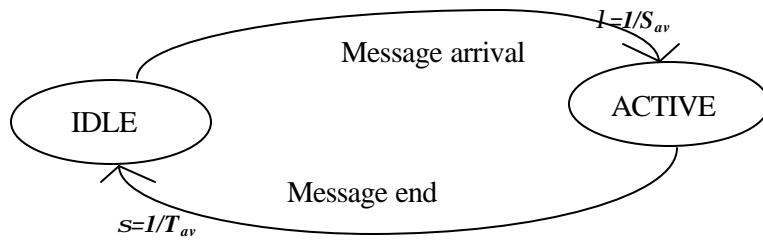
The BS acknowledges the outcome of preceding reservation-packet transmission (success or not) and allocates I slots by means of ACK and NACK packets transmitted through the acknowledgment channel.

The BS station is also responsible for notifying the WS of the channel occupancy as well as the appropriate value of P_R. Special control packets do this.

Looking at Fig B-2, it can be seen from it that there are three types of acknowledge packets; ACK1 which contains the WS id and a "wait" pattern in case that there are no free I slots currently; ACK2 which contains also the WS id and the number of the I slot allocated "I_m" in case a reservation is made successfully; and finally NACK which contains an "Idle" pattern in case of errors in the received reservation packet. Each WS, which has transmitted a reservation packet, waits till it receives its own ACK or any NACK packet before it transmits information. In the case of NACK it assumes that the transmission was unsuccessful and tries again.

$I=1/S_{av}$

Message arrival

IDLE

ACTIVE

Message end

$s=1/T_{av}$

(a)   Source States

S2

(H)

Aloha

Contending for
an R slot

Reservation Success

(S_{aloha})

Message Arrival

(λ)

(K₁)

S3

S1

Idle
(I)

ACTIVE

Waiting for
an I slot

Message End

Start of Transmission

Transmission

(K₂)

S4
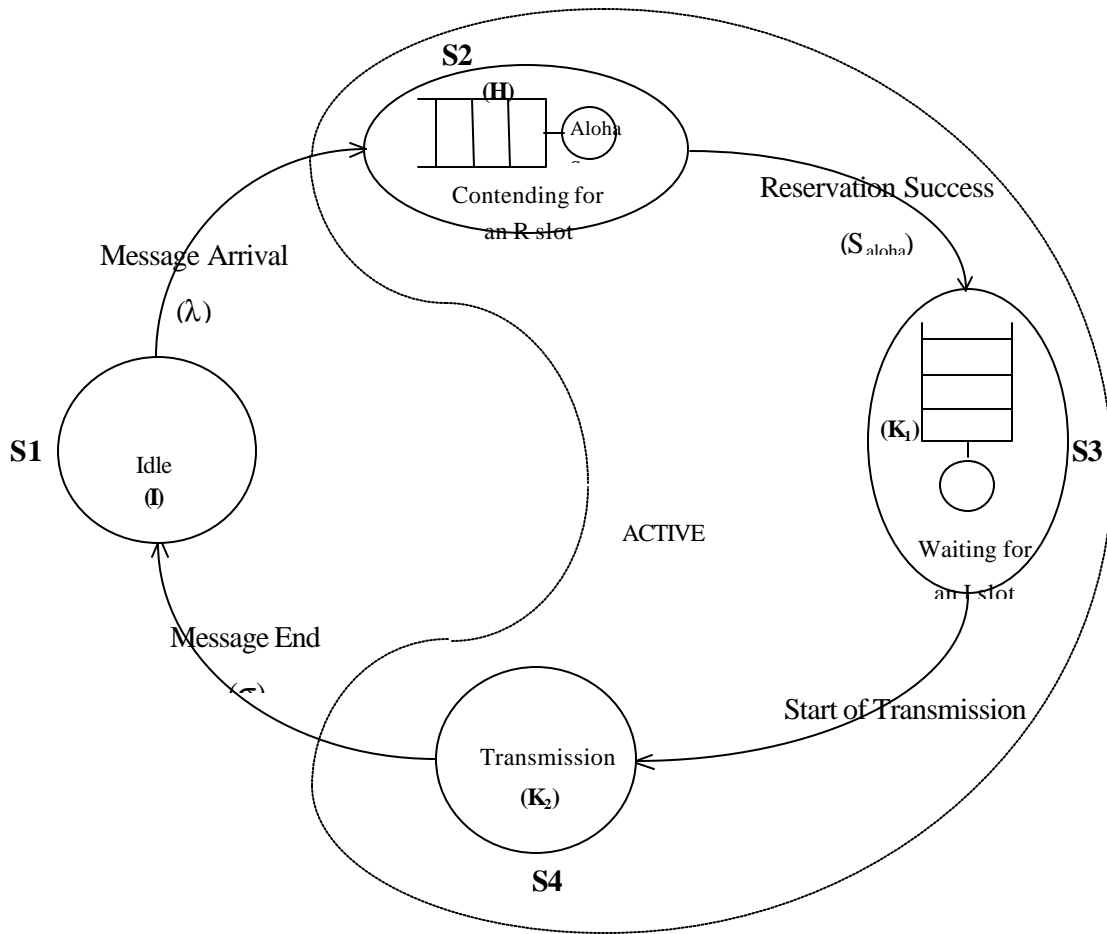
(b)   Connection States

Fig B-3 PRMA state transition diagrams: (a) Source states (b) Connection states

Idle

Message arrives

Ps=Pt

start of R minislot reached ?

No

Yes

generate uniform number x

x>r ?

No

Yes

Ps=Pr

Transmit reservation request

Ps=Pr

Negative ack. (NAK)

Yes

Ack. received from base station ?

No

Timeout reached

No

Yes

Yes

No

Wait pattern

Yes

My address is in ACK

No

No

Yes

Hold data

My reservation ACK received

Yes

No

Get I slot number from ACK

I slot reached

No

Yes

Transmit packet

No

Message end reached

Yes

Fig B-4 Flow chart of MAC procedure of PRMA

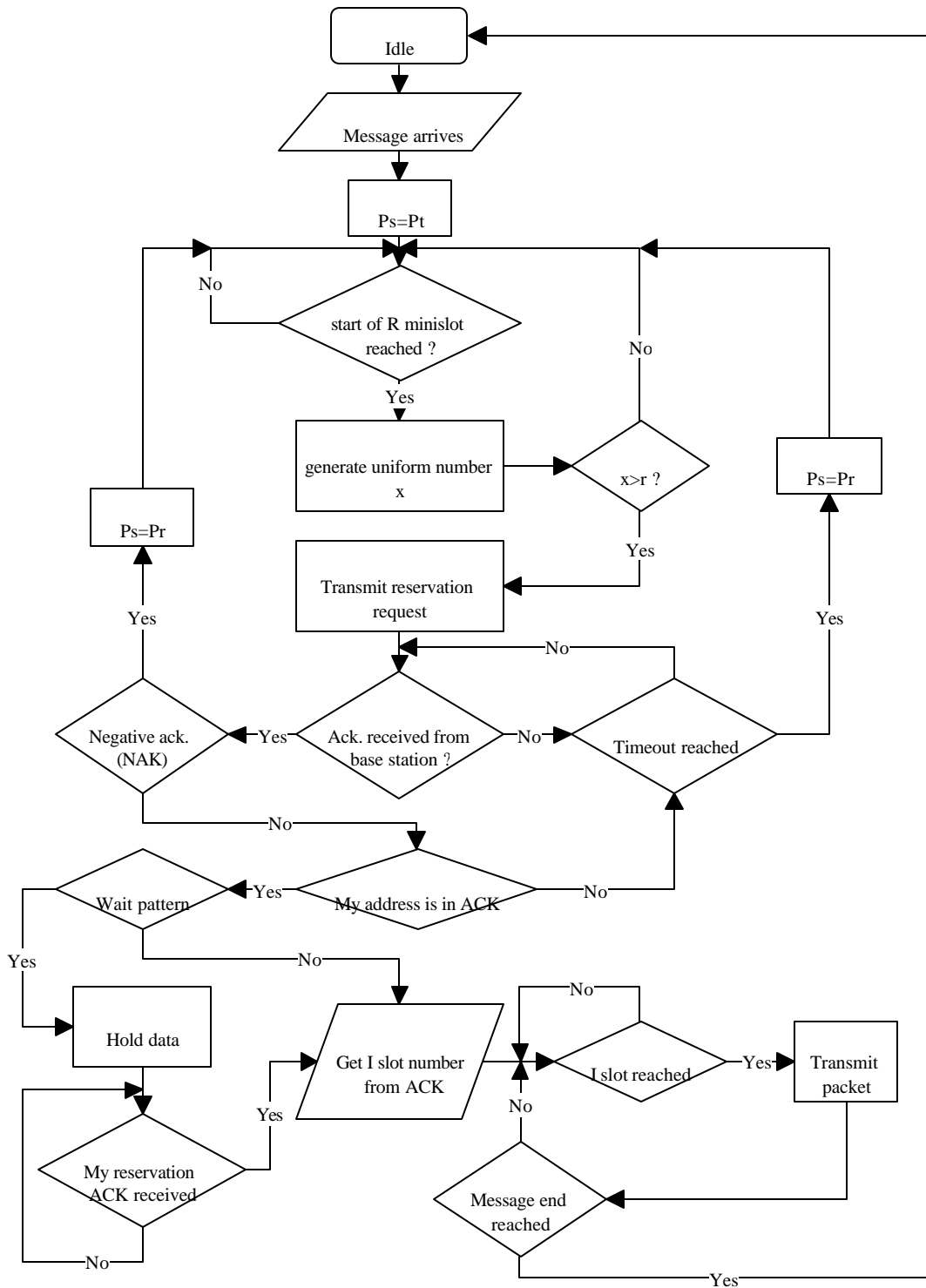The acknowledgment delay depends on the structure, and load of the outbound channel. However, this channel is free from collision, as it's completely a controlled access (no contention) channel. This provides more free BW which can be allocated for the acknowledgment channel.

From the previous descriptions it can be inferred that the main features of this protocol are as follows:

1. The BS has complete control on the information-slot allocation process. The allocation can be done in a way that the remaining free slots are distributed as uniformly as possible over the frame. This increases the probability of collision-free transmission, which enhances the overall performance of the system.

2. The reservation slots are divided to more than one mini-reservation slots, which results in a more efficient usage of the available reservation BW.

3. By assigning some of the frame slots as fixed reservation slots, a minimum bandwidth is reserved for reservations, which results in keeping the access delay under a certain lower limit especially in high load conditions.

4. By allowing for I slots to be temporarily used as reservation slots the protocol has gained an adaptive property that makes its operation more stable under different loading conditions.

5. The outbound channel has only one user, which is the base station. Hence, its only effect on performance is the transmission time of the frame, which is added to the delay.

Performance Evaluation

From the previous description of the protocol, it can be seen that the open queuing model for this protocol is as shown in Fig B-5.
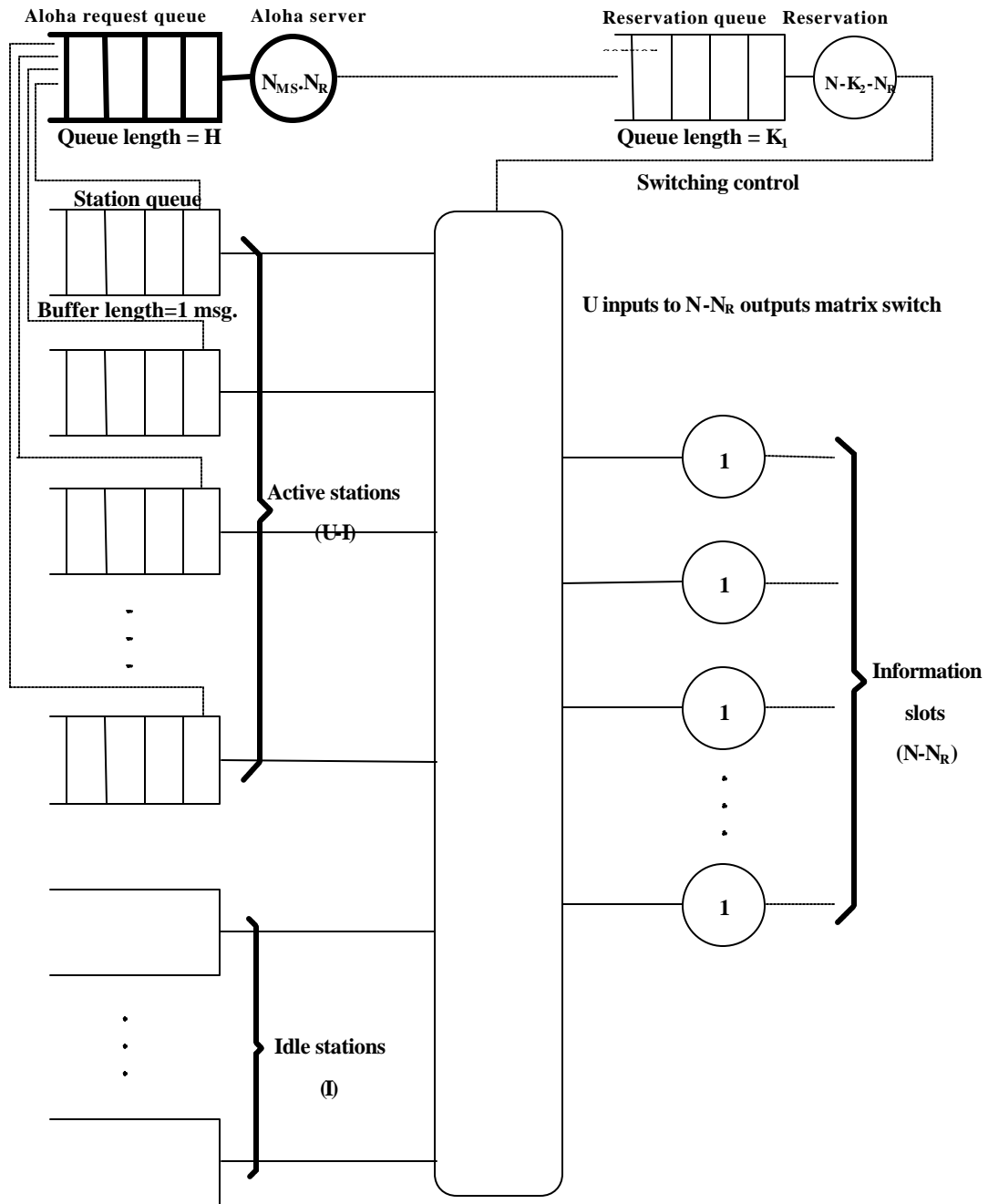
Fig B-5 Queuing model of PRMA

The above queuing model represents a system that consists of U customers. I of these customers are idle, i.e. have no messages to send. The other U-I customers are active and have a buffer for only one message, which may consists of many packets. Active stations are blocked and do not accept any new messages tell they become idle again. Active stations could be in one of three stages. The three stages are respectively; contending to send a reservation for an information slot, waiting to be assigned an information slot after sending reservation, or transmitting after being assigned an information slot.

When a source changes from idle state to active state, it joins the Aloha queue by starting to contend for a reservation mini-slot. As there are (N_R) reservation slots per frame, and each contains (N_{MS}) mini-slots, then the Aloha server has a service rate equal to $N_R.N_{MS}$ packets per second. The length of the Aloha queue is the number of contending stations (H).

After a source succeeds in the Aloha process, its reservation request is queued in the reservation queue. The reservation server services it whenever a free information slot exists. Accordingly, the reservation server has a service rate equal to the number of free information slots per frame. This is equal to subtracting the sum of the total number of busy information slots per frame (K_2) and the total number of reservation slots per frame (N_R), from the total number of slots per frame (N). This means that the service rate is N-N_R-K_2 packets per second. The length of the reservation queue is the total number of stations waiting for an assignment of an information slot (K_1).

Upon service by the reservation server, the node requesting transmission is assigned a fixed information slot that is kept for it till the end of all the packets of the current message. Consequently, the information transmission process is a deterministic process and its server is capable of transmitting only one packet per frame. On the other hand, because there are more than one information slot per frame, the queuing model shows multiple servers. These servers are assigned one for each node queue under control of the reservation server. This control is represented in the queuing model by the matrix switch.

As it can be seen from the above model it is mathematically difficult to analyze the network using this open queuing model. This difficulty can be very obvious by noting that to analyze the system using the above queuing model, an expression for the arrival processes and departure processes to and from each independent queue must be found, which could be a very

exhaustive process specially that the service rates for some servers are dependent on the lengths of some of the queues of the system.

Although the previously given queuing model is not practically useful for mathematical analysis, it is very much useful in describing the system operation. As far as mathematical analysis is considered, another more simple and yet accurate technique, called *closed queuing model*, is used.

The closed queuing model in brief represents the system status by grouping the customers in the same transmission state under one group. This leads to having a number of groups equal to the number of possible states. The number of these groups will constitute the dimensions of the network state vector. The coordinates of the state vector will represent the number of customers in each group. For further explanations, refer to [WOODWARD93].

In the following, an analytical model using closed queuing method will be presented including all assumptions and approximations used, then the final equations will be stated. This model is already derived in [MITROU90], but extra explanations are given here.

The system is modeled as a Markov model with fixed population U. The states are represented by three-dimensional vectors. The components of the vectors are *(H,I,K)*. The closed queuing model for the system is shown in Fig B-3. I is the number of silent stations of all population U. H is the number of stations contending for a slot (i.e. trying in the Aloha reservation). K is the number of stations either transmitting K_2 or waiting for an information slots K_1.

If we can find the probability $P(h,i,k)$, which is the probability of being in state $(H=h,I=i,K=k)$ then all the required statistics like throughput and delay can be calculated easily.

Calculation of Probability of *I*

The aggregate silent process of all U sources, represented by the random variable I is an autonomous process i.e. it is independent of the values of H and K.

In part (a) of Fig B-3 a two-state model of the source is assumed. The two states of the source are Idle and Active. The source is assumed to stay in each state for a random duration that is exponentially distributed with average values equal to S_{av} and T_{av} respectively.

Accordingly, the average rates of entering to and exiting from the Idle state are respectively $s=1/T_{av}$ and $l=1/S_{av}$.

Since the process described by the random variable I is affected by the sum of the arrival processes over all of the nodes regardless of the individual arrival process model of each source, the total aggregate arrival process of them will be approximately Poisson with an arrival rate equal to the sum of the arrival rates of the active sources. This means that the random variable I has an inter-arrival time that is exponentially distributed.

By assuming that the slot size is very small compared to the average time between arrivals, only one arrival can occur during a slot. Also, as the arrival and departure rates are comparable, only one departure can occur during a slot.

It must be noted that service here means that source changes from idle to active state, and arrival means that a source ends its transmission and changes to idle state. Also, any number of sources, up to the maximum number of users, could be active simultaneously, which means that customers can be serviced simultaneously. This means that there are multiple servers in the system, and as the number of these servers is larger than the number of the customers, the queuing system behaves as if it has infinite number of servers.

Finally, as the number of customers in the silent state reaches the maximum number of population U, the system is non-blocking and can be safely assumed to have infinite buffer mathematically.

These assumptions lead to a queuing system that has single arrival and single departure per slot. The population of this system is finite, which is equal to the number of stations. The system can be assumed to have an infinite buffer and infinite number of servers. The inter arrival time is of an exponential distribution as the arrival process is Poisson. Finally, the service process is assumed to have exponential distribution using the same arguments used for the arrival process.

The resulting queue from the above description is an $M/M/\yen/\yen/U$ queue. The solution of this queue is given by the equation [BERTSEKAS87]:

$$\Pr\{I=i\}= P(i)= (s/l)^i \binom{U}{i} \bigg/ (1+s/l)^U \ ,0 \le i \le U$$

(B-1)

where :

s is the rate of entering to the silent state for a single source.

l is the rate of exiting from the silent state for a single source.

U is the total number of stations (Population).

Note that :

$$s = \frac{1}{T_{av}} , l = \frac{1}{S_{av}}$$ (B-2)

where:

T_{av} is the average talk period for a single source.

S_{av} is the average silence period for a single source.

After studying the *Speaking/Silent* process, it is time to look at the *Aloha Reservation* process. This process is described by the random variable H, which is the number of stations contending for a reservation slot.

In contrary to I, the random variables H and K depend on the value of I. This means that to calculate the state probability we need to find $Pr\{H=h,I=i,K=k\}$.

The problem is simplified if it is noticed that $k=U-h-i$. This means that K can be totally determined if H and I are calculated. From this it can be concluded that the states are totally determined if only H and I are determined. This reduces the problem to two-dimensional state variables instead of three-dimensional variables. This reduction means we need to find the probability $Pr\{H=h,I=i\}$ instead of finding $Pr\{H=h,I=i,K=k\}$ which contains three variables. Also from probability theory it is known that:

$Pr\{H=h,I=i\}=Pr\{H=h/I=i\}.Pr\{I=i\}$

An expression for $Pr\{I=i\}$ have already been given. In the following section the expression for the conditional probability $Pr\{H=h/I=i\}$ will be derived.

Calculation of Probability of H|I

Define $H(t)$ as the number of stations contending for an R slot at time t, where t is measured in mini-slot times. It is assumed that with varying H, the process described by $H(t)$ remains stationary. This means that Prob$\{H(t)\}$ is not a function of time t. This is not completely true,

but it is an acceptable approximation if we assume that Prob{$H(t)$} has a settling time constant that is much smaller than the time slot.

It is also assumed that the probability of transmitting in the Aloha scheme is the same for both backlogged stations (stations that are retrying transmission) and stations that are just starting to contend, i.e.,

$$P_T = P_R = p_t \qquad\qquad\qquad (\text{B-3})$$

Here:

P_T is the probability of transmission for a station just starting to transmit.

P_R is the probability of transmission for a backlogged station (i.e. a station that has failed to transmit and trying to transmit again).

This assumption is valid because it is most common that a free I slot that is used as an extra reservation slot (ER slot) follows a long interval of occupied slots. In fact this is the reason why we need ER slots. This means that if we take any of the two probabilities equal to one, the probability of conflict on the ER slot will increase, which will increase the collision probability. So it is reasonable to take both P_T and P_R not equal to one. The reason for taking them equal to each other is to simplify calculations. This assumption means that both newly arriving messages and those backlogged will take the same chance (i.e. priority) for granting access to the channel. In fact it is more reasonable to have backlogged messages having lower priority as they already had a previous chance i.e., $P_T > P_R$.

Moreover the mini-slot size is assumed to be small compared to the mean time between state changes. This assumption means that during the mini-slot time the system can occupy only one state. This results in having $I(t)$, and $H(t)$ constants over the mini-slot period, which means that time can be taken to be of discrete nature with steps equal to the mini-slot size.

It is very essential to notice that the current value of $H(t+1)$ is equal to the previous value of $H(t)$ plus the number of arriving (changing from silent to contending) stations $dH(t)$ minus the number of departing (succeeding in reservation) stations $\tilde{N}H(t)$. Also, the probability density functions of $dH(t)$ and $\tilde{N}Ht)$ depends only on the current value of $H(t)$ at time t and are independent of all previous values of $H(t)$ before time t. Hence, the probability density function of $H(t+1)$ depends only on the values of $H(t+1)$ and $H(t)$. This together with the

assumption of small mini-slot size leads to a system that has discrete time state transitions with the current state of the system having probability distributions that depend only on the current and previous state of the system. This is the so called *Markov Chain* process.

A Markov chain is completely described by its one step transition probabilities. These probabilities for the given process which is $H(t)/I$ will be referenced as P_{hj} where:

$P_{hj}=Pr\{H(t+1)=j|H(t)=h,I=i\}$

They will be stated verbally then quantitatively in the following:

$$P_{hj} = \begin{cases} 0, & j < h-1 \\ \Pr\{oneST(t), noNT(t)| H(t)=h, I=i\}, & j = h-1 \\ \Pr\{[noST(t), noNT(t)] or [oneST(t), oneNT(t)]| H(t)=h, I=i\}, & j = h \\ \Pr\{noST(t), oneNT(t)| H(t)=h, I=i\}, & j = h+1 \\ 0, & j > h+1 \end{cases}$$

(B-4)

where:

$ST(t)$ is the event of successful transmission at the tth mini-slot.

$NT(t)$ is the event of new transmission message at the tth mini-slot.

Now let the probability of being in a reservation mini-slot at time t be denoted by $P_{MS}(t)$. This probability can be given approximately by the following equation:

$$P_{MS}(t) = N_R(t)/N$$

(B-5)

where:

$N_R(t)$ is the number of reservation slots in the frame at time t including extra reservation slots.

N is the total number of slots per frame.

Now the quantitative values of the previously given probabilities will be derived:

$$\Pr\{oneST(t)| H(t)=h, I=i\} = hp_t P_{nerr} \cdot [1-p_t]^{h-1} \cdot P_{MS}(t)$$
$$\Pr\{noST(t)| H(t)=h, I=i\} = 1 - hp_t P_{nerr} \cdot [1-p_t]^{h-1} \cdot P_{MS}(t)$$
$$\Pr\{oneNT(t)| H(t)=h, I=i\} = iIF/(N.N_{MS})$$
$$\Pr\{noNT(t)| H(t)=h, I=i\} = 1 - iIF/(N.N_{MS})$$

(B-6)

where:

F is the frame size in seconds.

N_{MS} is the number of mini-slots per slot.

P_{nerr} is the probability of error free packet delivery, which depends on source coding, channel coding, error control techniques and channel characteristics.

$$p_{hj} = \begin{cases} 0, & j < h-1 \\ hp_t P_{nerr}[1-p_t]^{h-1} P_{MS}(t)(1-i\mathbf{1}F/N.N_{MS}), & j = h-1 \\ [1-hp_t P_{nerr}[1-p_t]^{h-1} P_{MS}(t)][1-i\mathbf{1}F/N.N_{MS}] & j = h \\ +hp_t P_{nerr}[1-p_t]^{h-1} P_{MS}(t).i\mathbf{1}F/N.N_{MS}, & \\ [1-hp_t P_{nerr}[1-p_t]^{h-1} P_{MS}(t)]i\mathbf{1}F/N.N_{MS}, & j = h+1 \\ 0, & j > h+1 \end{cases} \tag{B-7}$$

The value of $P_{MS}(t)$ depends on $N_R(t)$, and $N_R(t)$ is given by the equation:

$$N_R(t) = \begin{cases} N-k, & k < N - N_{RMIN} \\ N_{RMIN}, & k \geq N - N_{RMIN} \end{cases} \tag{B-8}$$

where:

N_{RMIN} is the minimum number of reservation slots per frame i.e. it is the so called reservation threshold.

Note that the above equation states that $N_R(t)$ depends only on k, and since $k=U-h-i$, then $P_{MS}(t)$ also depends only on k by substituting from equation (B-8) into equation (B-5).

Calculation of State Probabilities $Pr\{h,i\}$

The equilibrium state probabilities conditioned on I can be found by solving recursively the following equation:

$$P(h \mid i) = P(h-1 \mid i).b_{h-1}/d_h \tag{B-9}$$

where:

$$b_h = p_{h,h+1}/(F/(N.N_{MS})) = [1 - hp_t P_{nerr}[1-p_t]^{h-1} P_{MS}]i\mathbf{1}$$
$$d_h = p_{h,h-1}/(F/(N.N_{MS})) = hp_t P_{nerr}[1-p_t]^{h-1} P_{MS}[N.N_{MS}/F - i\mathbf{1}] \tag{B-10}$$

and b_h, d_h are the birth and death rates of the birth/death process, respectively.

Denoting the probability of being at state $h=0$ by $P(0/h)$, we can write:

$$P(h\mid i)=\begin{cases}P(0\mid i)\prod_{j=1}^{h}\dfrac{b_{j-1}}{d_{j}}, & 1\le h\le U-h,\quad 0\le i\le U-1,\quad h+i+k=U\\[2mm]1, & h=0,\qquad\quad i=U,\qquad\quad h+i+k=U\\[2mm]P(0\mid i), & h=0,\qquad\quad 0\le i\le U-1,\quad h+i+k=U\\[2mm]0, & otherwise\end{cases}$$

(B-11)

and from the normalization equation of probabilities we can write:

$$\sum_{h=0}^{U-i}P(h\mid i)=1 \tag{B-12}$$

Substituting from (B-11) into (B-12) and with some simple mathematical manipulations it can be shown that:

$$P(0\mid i)=\begin{cases}\dfrac{1}{1+\sum_{h=1}^{U-i}\prod_{j=1}^{i}\dfrac{b_{j-1}}{d_{j}}}, & 0\le i\le U-1\\[4mm]1, & i=U\\[4mm]0, & otherwise\end{cases}$$

(B-13)

Finally, as stated before to find the unconditional state probabilities *Prob{h,i}* from the conditional probability *Prob{h/i}*, the famous relation *Prob{h,i}=Prob{h/i}.Prob{i}* is used, giving:

$$P(h,i)=\begin{cases}P(0\mid i)\cdot\dfrac{\left(\dfrac{s}{l}\right)^{i}\binom{U}{i}}{\left(1+\dfrac{s}{l}\right)^{U}}, & 0\le i\le U,\qquad h=0,\qquad\quad h+i+k=U\\[5mm]P(0\mid i)\prod_{j=1}^{h}\dfrac{b_{j-1}}{d_{j}}\cdot\dfrac{\left(\dfrac{s}{l}\right)^{i}\binom{U}{i}}{\left(1+\dfrac{s}{l}\right)^{U}}, & 0\le i\le U,\qquad 1\le h\le U-i,\quad h+i+k=U\\[5mm]0, & otherwise\end{cases}$$

(B-14)

Performance Parameters

The normalized throughput per frame time S is defined as the total number of successfully transmitted messages during the frame time divided by the channel capacity. As the Aloha process controls the time division multiplexing access, then we will need to find the throughput of the Aloha process first so that we can find the system overall throughput and delay.

The throughput of the Aloha system $S_{aloha}(h,i)$ can be stated verbally to be equal to the number of contending stations h multiplied by the probability that one of them transmits successfully in a slot multiplied by the number of reservation slots per frame divided by the channel capacity per frame.

Hence, we have the Aloha channel normalized throughput $S_{aloha}(h,i)$ given by the following equation:

$$S_{aloha}(h,i) = h p_t P_{nerr} [1-p_t]^{h-1} . N_R(t)/N \tag{B-15}$$

On the other hand, the offered load to the aloha channel at the input is given by the equation:

$$G_{aloha}(i) = i \mathbf{1}\, F/(N.N_{MS}) \tag{B-16}$$

The message access delay $D(h,i)$ consists of two parts; the random access delay of the Aloha process, and the queuing delay in the reservation queue till a slot is reserved.

Then the total message access delay for a certain h and i denoted by $D_{acc}(h,i)$ is given by:

$$D_{acc}(h,i) = D_{aloha}(h,i) + D_{queue}(h,i) \tag{B-17}$$

where:

$$D_{aloha}(h,i) \cong [h/S_{aloha}(h,i)].\,F/(N.N_{MS}) + n_{tr}(h,i).Ackdelay.F/N \tag{B-18}$$

$$D_{queue}(h,i) = \begin{cases} 0.5F & U-h-i-N+N_{RMIN} < 0 \\ F + [T_{av}/(U-i)].(U-h-i-N+N_{RMIN}) & otherwise \end{cases} \tag{B-19}$$

$n_{tr}(h,i)$ denotes the total number of retries before a successful transmission occurs and is given by:

$$n_{tr}(h,i) = 1/([1-p_t]^{h-1}.P_{nerr}) \tag{B-20}$$

Also the Acknowledge delay (*Ack delay*) is small because the return channel is free of collisions. Hence, the *Ack delay* here will be assumed to be constant, and will take the value of three time slots.

Substituting equations (B-11), (B-16), and the value of *Ack delay* in equation (B-14) we get:

$$D_{aloha}(h,i) \cong \left[1/(p_t \cdot N_R(k)N_{MS}) + 3/N \right] . F / \left(P_{nerr} \cdot [1-p_t]^{h-1} \right)$$ (B-21)

This access delay is per message. Another delay that must be added is the average transmission time of the message D_{tr}. This is equal to the number of packets in the message multiplied by the frame size as each frame services only one packet of the message.

$$D_{tr} \cong F . \frac{T_{av}}{\left(F/N \right)}$$ (B-22)

The total message delay $D(h,i)$ is the sum of both.

$$D(h,i) = D_{acc}(h,i) + D_{tr}$$ (B-23)

The throughput of the system $S(h,i)$ can be found from Little's formula as follows:

$$S(h,i) = \frac{U - i}{D(h,i)}$$ (B-24)

Finally, it must be noticed that the values of throughput, offered load, and delay are state-dependent i.e. they are given for a certain value of i and h. To get the average values for throughput, offered load, and delay, the state-dependent values should be averaged over the state probabilities.

In mathematical words this is given by:

$$S = \sum_{h,i} S(h,i).P(h,i)$$
$$G = \sum_i G(i).P(i)$$ (B-25)
$$D = \sum_{h,i} D(h,i).P(h,i)$$

One more thing to note is that an important parameter affecting the performance of this system is the transmission probability p_t. This parameter is actually calculated by the access point according to the state of the system to ensure better adaptive performance. In the following the equilibrium equation will be represented as a function of p_t and k, then it will be shown

how to find a value for p_t. The criterion for choosing p_t is the stability of the Aloha channel. This criterion results in a sufficient but not optimum solution in terms of channel access delay. At the equilibrium point the output throughput of the Aloha system must equal the input offered load to it. Then we get at equilibrium:

$$hp_t P_{nerr}[1-p_t]^{h-1}.N_R(t)/N = i\mathbf{1}\ F/(N.N_{MS}) \tag{B-26}$$

If this equation is plotted on an h-i plane with k as a parameter and a certain value for p_t it gives curves that are the equilibrium contours. This is shown in Fig B-6 below for two values of k.

In this figure there are two equilibrium contours with k taking two values $k=30$ and $k=10$, while p_t is taken to be 0.1, and U is taken to be 50. Also drawn on the figure are the two load lines. These load lines are the plot of equation:

$$k = U - h - i \tag{B-27}$$

The intersection of the equilibrium contour with the load line for a certain value of k gives the values of i and h at equilibrium.

To guarantee that the solution delivers a stable operation of the system then the throughput S must be in the non-saturating region. This region is the region where S is increasing by increasing h. This is the region on the equilibrium contours to the left of the maxima of the curve. Hence for a stable operation the load line must intersect with the equilibrium contour in a single point to the left of the maxima of the contour. For complete stability, this should happen for all possible values of k (i.e. $0<k<U$). In the case of Fig B-6 the system is stable where $k=30$. On the other hand, the case for $k=10$ gives an unstable system for the chosen value of p_t.
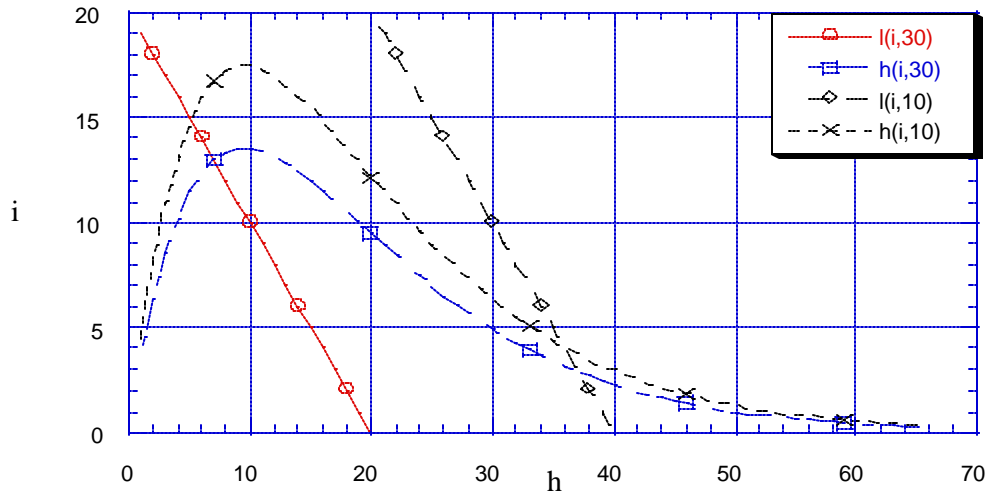
Fig B-6 Choosing p_t for stable operation

REFERENCES

REFERENCES

[AMUNDSEN96]   Keith B. Amundsen, "IEEE 802.11 Wireless LAN: draft standards", The 2nd. IEEE Workshop on Wireless Local Area Networks, Wrocester Ploytechnic Institute, Oct. 1996

[BANTZ94]   David F. Bantz and Frederic J. Bauchot, "Wireless LAN Design Alternatives", IEEE Network Magazine, pp. (43-53), Mar./Apr. 1994

[BATES94]   Regis J. Bates, "Wireless Networked Communications", McGraw-Hill, 1994

[BAUCHOT95]   F.J. Bauchot, and F. Lanne, "IBM Wireless RF LAN Design and Architecture", IBM System Journal, Vol. 34 No.3, pp. (390-408), 1995

[BERTSEKAS87]   Dimitri Bertsekas, and Robert Gallager, "Data Networks", Prentice-Hall, 1987

[BOER96]   Jan Boer, "Direct Sequence Spread Spectrum Physical Layer Specification: IEEE 802.11", 802.11 Tutorial for IEEE P802.11, March, 1996.
http://stdsbbs.ieee.org/groups/802/11/

[BUX81]   W. Bux, "Local Area Subnetworks: A Performance Comparison", IEEE Transactions on Communications, Vol. COM-29 No.10, pp. (1465-1473), Oct. 1981

[CHAYAT96]   Naftali Chayat, "Frequency Hopping Spread Spectrum PHY of the 802.11 Wireless LAN Standard", 802.11 Tutorial for IEEE P802.11, Mar. 1996

http://stdsbbs.ieee.org/groups/802/11/

[CHHAYA96]   Harshal S. Chhaya, "Performance Evaluation of the IEEE 802.11 MAC Protocol for Wireless LANs", M.Sc. Thesis, Illinois Institute of Technology, May 1996

[CROW97]   Brian P. Crow, Indra Widjaja, Jeong Geun Kim, and Prescott T. Sakai, "IEEE 802.11 Wireless Local Area Networks", IEEE Communication Magazine, pp. (116-126), Sep. 1997

[DAVIS95]   Peter T. Davis and Craig R. McGuffin, "Wireless Local Area Networks : Technology Issues and Strategies", McGraw-Hill, 1995

[DIEPSTATEN96]   Wim Diepstraten, and Phil Belanger, "802.11 MAC Entity: MAC Basic Access Mechanisms Privacy and Access Control", 802.11 Tutorial for IEEE P802.11, Mar. 1996

http://stdsbbs.ieee.org/groups/802/11/

[ENNIS96]   Greg Ennis, "802.11 Architecture", 802.11 Tutorial for IEEE P802.11, Mar. 1996

http://stdsbbs.ieee.org/groups/802/11/

[FERNANDES94]   Jose J. G. Fernandes, Peter A. Watson, and Jose C. Neves, "Wireless LANs : Physical Properties of Infra-Red Systems Vs. Mmw Systems", IEEE Communication Magazine, pp. (68-73), Aug. 1994

[FUKUDA83]   Akira Fukuda, and Shuji Tasaka, "The Equilibrium Point Analysis – A Unified Analytic Tool for Packet Broadcast Networks", IEEE Transactions on Communications, Vol. COM-31 No.12, pp. (1133-1140), Dec. 1983

[HAYES96]   Vic Hayes, "Tutorial on 802.11 to 802 Committee", 802.11 Tutorial for IEEE P802.11, Mar. 1996.
                     http://stdsbbs.ieee.org/groups/802/11/

[KAMERMAN97]   Ad Kamerman, and Leo Monteban, "WaveLAN II: A High Performance Wireless LAN for the Unlicensed Band", Bell Labs Technical Journal, pp. (118-133), Summer 1997

[KLEINROCK75]   Leonard Kleinrock, and Fouad A. Tobagi, "Packet Switching in Radio Channels: Part I – Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics", IEEE Transactions on Communications, Vol. COM-23 No.12, pp. (1400-1416), Dec. 1975

[LAMAIRE96]   Richard O. LaMaire, Arvind Krishna, and Pravin Bahagwat, "Wireless LANs and Mobile Networking : Standards and Future

Direction", IEEE Communication Magazine, pp. (86‑94), Aug. 1996

[LANSFORD99]   Jim Lansford, "Bringing Wireless Connectivity Home", IEEE Standards Web Site document, Mar. 1999
http://stdsbbs.ieee.org/groups/802/11

[MITROU90]   M. Mitrou, TH. D. Orinos, and E. N. Protonotarios, "A Reservation Multiple Access Protocol for Microcellular
Mobile Communication Systems", IEEE Transactions on Vehicular Technology, Vol. 39 No.4, pp. (340-351) Nov. 1990

[MULLER95]   Nathan J. Muller, "Wireless Data Networking", Artech House, 1995

[O'DONNELL94]   Bernard A. O'Donnell, "Wireless LANs: Overview", "Data‑pro: Technology Issues & Trends in Managing LANs", part1220, pp. (1-8), Apr. 1994

[PAHLAVAN95]   Kaveh Pahlavan, Thomas H. Probert, and Mitchell E. Chase, "Trends in Local Wireless Networks", IEEE Communication Magazine, pp. (88-95), Mar. 1995

[PAHLAVAN96]   Kaveh Pahlavan, "Trends in Wireless LANs', The 2nd IEEE Workshop on Wireless LAN, Worcester Polytechnic Institute, Oct. 1996

[PAHLAVAN97]   K. Pahlavan, Ali Zahedi, and Prashant Krishnamurthy, "Wideband Local Access: Wireless LAN and Wireless ATM", IEEE

Communication Magazine, pp.(68-76) Nov. 1997

[PROAKIS95]   John G. Proakis, "Digital Communications: third edition", McGraw-Hill, 1995

[ROOSMALEN94]   Huub van Roosmalen, Jos Nijhof, and Ramjee Prasad, "Performance Analysis of a Hybrid CDMA/ISMA Protocol for Indoor Wireless Computer Communications", IEEE Journal on Selected Areas in Communications, Vol. 12 No.5 pp. (909-916), Jun. 1994

[RUNE95]   Torben Rune, "Wireless Local Area Networks", Netplan Aps
                   http://www.netplan.dk/netplan/wireless.html

[SCHWARTZ87]   Mischa Schwartz, "Telecommunication Networks: Protocols, Modeling and Analysis", Addison-Wesley, 1987

[STALLING97A]   William Stalling, "Data and Computer Communications : Fifth Edition", Prentice Hall, 1997

[STALLING97B]   Willaim Stallings, "Local and Metropolitan Area Networks: fifth edition", Prentice Hall, 1997

[TOBAGI76]   Fouad A. Tobagi and Leonard Kleinrock, "Packet Switching in Radio Channels: Part III – Polling and (Dynamic) Split-Channel Reservation Multiple Access", IEEE Transactions on Communications Vol. COM-24 No.8, pp. (832-845), Aug. 1976

[TUCH93]   Bruce Tuch, "Development of WaveLAN: and ISM Band Wireless
                   LAN", AT&T Technical Journal, pp. (27-37), Jul./Aug. 1993

[VITERBI95]   Andrew J. Viterbi, "CDMA Principles of Spread Spectrum
                   Communication", Addison-Wesley, 1995

[WEINMILLER97]   Jost Weinmiller, Morten Schlager, Andreas Festag, and Adam
                   Wolisz, "Performance Study of Access Control in Wireless LANs
                   - IEEE 802.11 DFWMAC and ETSI RES 10 HIPERLAN",
                   Work supported by a grant from the Deutsche Forschungs
                   Gemeinschaft (DFG), for The Priority Program Mobile
                   Communications.
                   http://www.tu-berlin.de/people/~jost/pubs/

[WLANA97]   "Overview of Wireless LANs"
                   http://www.wlana.com/into/intro.html

[WOODWARD91]   M. E. Woodward, "Equivalence of Approximation Techniques for
                   Solution of Multidimensional Markov Chains in Network
                   Modeling", Electronic Letters, Vol. 27 No.12, pp. (1019-1021),
                   6th Jun. 1991

[WOODWARD93]   Michael E. Woodward, "Communication and Computer Networks
                   modeling with discrete time queues", Pentech Press, 1993