A Distributed Wireless MAC Scheme for Service Differentiation in WLANs

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Engineering in Electrical and Electronic Engineering
At the
University of Canterbury,
Christchurch, New Zealand.

Abdulla Firag

Department of Electrical and Computer Engineering
University of Canterbury
Christchurch, New Zealand

May 2003

Abstract

Mobile communications is evolving due to the recent technological achievements in wireless networking. Today, wireless networks exist in many forms, providing different types of services in a range of local, wide area and global coverage. The most widely used WLAN standard today is IEEE 802.11. However, it still has problems with providing the QoS required for multimedia services using distributed methods. In this thesis, a new distributed MAC scheme is proposed to support QoS in wireless LANs. In the scheme, stations use CSMA for channel access, with collisions between stations being resolved by sending a set of beacons in a predefined manner, and *virtual collisions* being resolved by schedulers at the stations.

The proposed MAC scheme is analyzed mathematically, for two-priority case, and the results obtained are validated by simulation. The mathematical model estimates the average delay experienced by data packets of priority one and two under different conditions. A performance evaluation study of the proposed MAC scheme as well as the IEEE 802.11 DCF, and IEEE 802.11e EDCF MAC schemes is also done by means of stochastic simulation. It is found that the results obtained by simulation are in very good agreement with the analytical results, thereby validating them. Moreover, the simulation study evaluated different performance measures of these MAC schemes. The results showed that the IEEE 802.11 DCF scheme does not support QoS, but the proposed MAC scheme and the upcoming IEEE 802.11 EDCF both do.

In general, the results show that the proposed MAC scheme performs equally or better than the current IEEE 802.11 DCF scheme in every case considered. It is also found that the proposed MAC scheme performs equally well as the upcoming IEEE 802.11e EDCF scheme, in every case considered in this thesis.

Acknowledgements

I greatly appreciate the help given by my supervisor, Associate Professor Harsha Sirisena. I would like to thank him for his constant guidance, support and encouragement throughout this thesis. I would also like to thank my co-supervisor, Associate Professor Krzysztof Pawlikowski for his guidance, especially towards the completion of this thesis. I am grateful for my colleagues at the Network Research Group at the University of Canterbury for the updated information that we shared from different networking areas. In particular, I would like to thank my wife for her support throughout this thesis and also checking the contents of this thesis. I would not have this opportunity without the scholarship provided by NZAID (New Zealand Agency for International Development). Therefore I would like to specially thank them.

Finally, I would like to thank the most important people in my life, my family. You have always given me the strength to overcome all the difficulties in life and I know you will be there wherever life takes me.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACF	Authentication Confirmation Frame
ACK	Acknowledgement
AIF	Authentication initiation frame
$AIFSC_A$	Arbitrary Interframe Space for authentication collision resolution
$AIFSC_i$	Arbitrary Interframe Space for Collided data of priority i
$AIFSN_i$	Arbitrary Interframe Space for New data of priority i
AIIFS	Authentication initiation interframe space
AP	Access point
BCCH	Broadcast control channel
BER	Bit error rate
BSA	Basic service area
BSS	Basic service set
CAC	Channel access control
CBR	Continuous bit rate
CDMA	Code division multiple access
CFP	Contention free period
CP	Contention period
CR	Collision resolution
CRB	Collision resolution beacon
CRC	Cyclic redundancy code
CRIFS	Collision Resolution IFS
CSMA	Carrier sense multiple access
CSMA-CA	CSMA collision avoidance
CSMA-CD	CSMA collision detection
CTS	Clear to send
CW	Contention window
DBPSK	Differential binary phase shift keying
DCF	Distributed coordination function
DIFS	DCF interframe space
DiL	Direct Link
DL	Downlink phase
DLC	Data link control

DQPSK	Differential quadrature phase shift keying
DS	Distributed system
DS-CDMA	Direct sequence CDMA
DSSS	Direct sequence spread spectrum
EDCF	Enhanced distributed coordination function
EIFS	Extended interframe space
ESS	Extended service set
ESV	Elimination survival verification
EY-NPMA	Elimination yield non pre-emptive priority multiple access
FCCH	Frame control channel
FCS	Frame check sequence
FDD	Frequency division duplex
FDM	Frequency division multiplexing
FDMA	Frequency division multiple access
FH-CDMA	Frequency hopping CDMA
FHSS	Frequency hopping spread spectrum
GFSK	Gaussian shaped frequency shift keying
GSM	Global System for Mobile
HC	Hybrid coordinator
HIPERLAN	High performance local area network
IBSS	Independent BSS
IEEE	Institute of Electrical and Electronics Engineers
IFS	Interframe space
IR	Infrared
ISM	Industry, science and medicine
LAN	Local Area Network
LLC	Logical link control
MAC	Medium access control
MAN	Metropolitan area network
MPDU	MAC protocol data unit
MSAP	MAC service access point
NAV	Network allocation vector
NPB	No packet beacon
PCF	Point coordination function
PDU	Protocol data unit
PF	Persistence factor
PHY	Physical layer
PIFS	PCF interframe space

PLCP	Physical layer convergence procedure
PMD	Physical medium dependent
PPB	Packet present beacon
PPM	Pulse position modulation
QBSS	Quality of Service supporting BSS
QoS	Quality of Service
RA	Radio access
RCH	Random access channels
RLC	Radio link control
RTS	Request to send
SDIFS	Scheduled Data IFS
SIFS	Short interframe space
SLRC	Station long retry count
SSRC	Station short retry count
STAID	Station identification
TC	Traffic category
TDD	Time division duplex
TDM	Time division multiplexing
TDMA	Time division multiple access
TP	Token Pass
TR	Token Receive
TXOP	Transmission opportunity
UL	Uplink phase
WEP	Wired equipment privacy
WLAN	Wireless Local Area Network

1.0 Introduction

This Chapter provides a general background and overview of the thesis. The goals of the thesis are presented and a brief summary describing the succeeding chapters is included. This chapter concludes with a brief discussion of the results of the thesis.

1.1 Background

Mobile communication is fast evolving outcome of recent technological achievements in wireless networking and portable devices, like notebook computers. Today, wireless networks exist in many forms, providing different types of services in a range of local, wide area and global coverage. Improvements in both wireless and wired network technologies like emergence of Voice-over-IP (VoIP) have resulted in providing multimedia services to mobile users.

Wireless Local Area Networks (WLANs) are deployed almost everywhere today. The WLAN market is growing at annual rate of 40 to 60 %, according to the Wireless LAN Alliance (WLANA) [SpectraLink 2000]. The international IEEE 802.11 standard for WLANs is the driving force behind this growth. The aim of the IEEE 802.11 committee is to define a common air interface (through standardizing the data link, Medium Access Control (MAC) and physical layer functions and specifications) which is independent of the underlying wireless transmission technologies [Toh 1997].

Multimedia applications impose requirements on communication parameters, such as data rate, drop rate, delay and jitter. Combining delay sensitive multimedia services, like voice, with streams of delay non-sensitive data traffic over the same the backbone creates lots of problems. Today, network managers are looking for a practical solution to provide the Quality of Service (QoS) required for such applications.

1.2 Goals of the Thesis

Although, IEEE 802.11 standard is the most widely used WLAN standard today, it has problems with providing QoS required for multimedia services. The IEEE 802.11 standard proposes to use two different methods to access the channel. The distributed access method, Distributed Coordination Function (DCF), is used normally for non real-time services. The centralized access mode, Point Coordination Function (PCF), is used for real-time services, to provide higher priority than that of non-real time services, as explained in [Cocker et al. 2001], [Chen and Yeh 2002], and [Petrick 1997]. However, the centralized scheme requires the existence of access points with specialized functions, and moreover, as it is concluded in [Visser and Zarki 1995], the centralized mode performs poorly. Recently, distributed methods for implementing QoS in IEEE 802.11 standard via service differentiation have been proposed; see for example [Krishnakumar and Sobrinho 1996], [Aad and Castelluccia 2001] and [Barry et al. 2001].

On the basis of these results, IEEE 802.11 Task Group E currently defines enhancements to IEEE 802.11 standard, called IEEE 802.11e, which introduces Enhanced DCF (EDCF). Like the DCF, the EDCF is a distributed access method in which no station has more responsibility than others to maintain the functionality of the network. The main objective of the EDCF is to support service differentiation among different priority traffic.

As the most widely used WLAN standard, IEEE 802.11 standard has problems with providing the QoS required for multimedia services, the main goal of this thesis is to develop an alternative fully distributed Medium Access Control (MAC) scheme that supports service differentiation in WLANs. In addition, this thesis also presents:

- performance evaluation of the service differentiation achieved by the IEEE 802.11 protocol,
- performance evaluation of the service differentiation achieved by the upcoming IEEE 802.11e protocol,
- an alternative fully distributed Medium Access Control (MAC) scheme that supports service differentiation in WLANs,
- an analytical model for the proposed MAC scheme, and
- performance comparison of the proposed MAC scheme with IEEE 802.11 and IEEE 802.11e protocols.

In order to cover the above topics the thesis is divided into seven chapters. The outline of the chapters is given in the next section.

1.3 Outline of Thesis

This section provides a brief overview of the chapters in this thesis. The thesis is organized as follows:

Chapter 1: This Chapter provides a general background and overview of the thesis. The goals of the thesis are given and a brief summary describing the overview for succeeding chapters are presented. This chapter concludes with a brief discussion of the results of this thesis.

Chapter 2: In the past decade, a great deal of research has been done in wireless communication. There have also been numerous new developments in this area. This chapter gives an overview of the basic results about wireless networks. This chapter briefly explains the different types of wireless networks, wireless LAN access techniques, and major issues regarding wireless LANs such as Quality of Service and hidden and exposed terminal problems in wireless networks.

Chapter 3: The most widely used wireless LAN standards today are IEEE 802.11 WLAN and HIPERLAN. This chapter gives detailed descriptions of the most widely used WLAN standards; the IEEE 802.11 WLAN and HIPERLAN. The chapter also gives a brief introduction to the upcoming WLAN standard IEEE 802.11e. The IEEE 802.11e is able to support Quality of Service in wireless environments by method of service differentiation.

Chapter 4: This chapter is one of the main chapters of the thesis. The chapter proposes a fully distributed MAC scheme that supports service differentiation in a wireless LAN environment. In this scheme, stations use CSMA for channel access, with collisions between stations being resolved by sending a set of beacons in a predefined manner. *Virtual collisions*, i.e. clashes between data packets of different priority within a station, are resolved by a scheduler inside the station. This chapter also presents other relevant information required for the proposed MAC scheme.

Chapter 5: This chapter presents a detailed mathematical model of the proposed MAC scheme. In order to simplify the model, a maximum of two priority classes are considered. The model is derived to estimate the average delay experienced by priority one and two packets under different conditions. The assumptions made to simplify the model are given in the chapter with brief discussions. The chapter also presents an expression to estimate the saturation throughput of the priority one traffic. This expression is used to estimate the saturation throughput of the proposed MAC scheme when the packet payload and the number of stations in a given BSS are given.

Chapter 6: This chapter discusses a performance evaluation study of the proposed MAC scheme as well as the IEEE 802.11 DCF, and IEEE 802.11e EDCF MAC schemes by means of stochastic simulation. Simulation models were developed using MATLAB. The chapter presents the simulation models used and the results obtained. The chapter also compares results obtained analytically and by simulation for the proposed MAC scheme. Furthermore, this chapter also presents a comparison study of the proposed MAC scheme with IEEE 802.11 DCF and IEEE 802.11e EDCF.

Chapter 7: This chapter provides conclusions drawn from the work done in this thesis. This chapter also suggests possible future work that can be done to expand the work presented in this thesis.

Appendix A: The appendix contains MATLAB scripts used for the simulation of the current IEEE 802.11 DCF, the upcoming IEEE 802.11e EDCF, and the proposed MAC schemes. The scripts contain extra information about the models.

1.4 Results

Although, the IEEE 802.11 standard is the most widely used WLAN standard today, it still has problems with providing QoS required for multimedia services using distributed methods. As a result, distributed methods for implementing QoS in IEEE 802.11 standard via service differentiation have been proposed by several researchers. On the basis of these researchers, IEEE 802.11 Task Group E currently defines enhancements to IEEE 802.11 standard, called IEEE 802.11e, which introduces new methods to support service differentiation among different priority traffic by using a distributed method. As IEEE 802.11e is still not standardized, the main aim of this thesis is to develop a MAC scheme that supports QoS in wireless LANs. In the proposed scheme, stations use CSMA for channel access, with collisions between stations being resolved by sending a set of beacons in a predefined manner, and *virtual collisions* resolved by a scheduler inside the stations.

The proposed MAC scheme was analyzed mathematically and the results obtained were validated by simulation. It was seen that the results obtained by the mathematical model were in very good agreement with the simulation results.

Both the simulation and analytical results of the proposed MAC scheme show that the delay experienced by the higher priority traffic has less average delay compared to the lower priority traffic. This means that the proposed MAC scheme is capable to support QoS in wireless LANs, by means of service differentiation. The results also showed that the upcoming IEEE 802.11e EDCF supports service differentiation but the current IEEE 802.11 DCF does not.

In general, the results show that the proposed MAC scheme performs equally or better compared to the current IEEE 802.11 DCF scheme in every single case considered. It is also found that the proposed MAC scheme performs as well as the upcoming IEEE 802.11e EDCF scheme, in every single case considered in this thesis.

2.0 Wireless Local Area Networks

The number of computer-based systems have grown because of recent advances in technology and as a result many different types of computer networks have been introduced and implemented. One of them is called the wireless Local Area Networks (WLANs). As its name implies, a local area network covers a limited geographical area, unlike wide area network [Hopper et al 1986]. In the past decade, a great deal of research has been done in wireless communication. There have also been numerous new developments in this area. This chapter gives an overview of the basic results about wireless networks. This chapter briefly explains different types of wireless networks, WLAN access techniques, and major issues regarding WLANs such as Quality of Service (QoS) and hidden and exposed terminal problems in wireless networks.

2.1 Introduction

In typical LANs the network elements such as servers, terminals, printers and other peripheral are connected by a system of copper or optical fiber wires. These are wired LANs. In the wired LANs the cabling cost can be as high as 40 percent of the whole installation [Santamaria and Lopez 1993]. Each network extension or reconfiguration would require additional wiring which is both time consuming and costly.

Traditional wireless networks, such as cordless cellular telephones, paging systems, mobile data networks, and mobile satellite systems, have experienced enormous growth over the last two decades. With this growth in wireless communication industry, a new concept of personal wireless communication system, WLANs has appeared in the industry due to the above mentioned economic factors. Unlike in wired LANs, the WLANs can offer portability and lower cost of installation. In general WLANs have these advantages over the wired LANs [Wesolowski 2002].

- *Flexibility* – WLAN nodes can communicate with each other within the network coverage area without major limitations in the terminal locations.
- *Simplified Planning* – The network planning is related to radio part. However configuration of network is not necessary.
- *Possibility of a temporary network configuration* – Wireless communications open a possibility to construct a local network which is only needed temporarily.

Like every other achievements, WLANs also have some disadvantages as a result of using a radio channel as the signal propagation medium. The main disadvantages include:

- *Lower transmission quality* – Wired LANs which uses optical fiber as a transmission medium has channel error rate at most 10^{-10}. However in WLANs the error rate in the radio channel can be as high as 10^{-3} and even worse.
- *Cost of wireless equipment* – The components of the WLANs is still much higher than the equivalent components of wired LANs.
- *Lower level of data security* – The information transmitted on the radio channel can be intercepted much easier than in wired LANs and also inappropriately used WLAN can be a source of interference for other sensitive devices such as medical equipment.

Although there are disadvantages like these in the WLANs, the WLAN industry is growing rapidly due to their benefits. Today, there are two widely accepted WLAN standards: IEEE 802.11 and HIPERLAN. In each of these WLAN, the network can operate either as ad hoc network or as an infrastructure network. These WLAN configurations and more details about the WLANs are given in the following sections.

2.2 Types of WLANs

WLANs may be divided into groups according to their network configurations. On this basis, the WLANs can either be infrastructure-base wireless networks or *ad hoc* wireless networks. These two types of WLANs are shown in Figure 2.1.

An ad hoc network is a collection of wireless mobile hosts forming a temporary network without the aid of any established infrastructure or centralized administration. This is the simplest configuration of the WLAN. Two or more mobile hosts if at any time are within range of each other, can set up an ad hoc independent network. In ad hoc networks, the mobile stations communicate directly with each other. Every mobile station may not be able to communicate with every other station due to range limitations. Mobile ad hoc networks do not rely on any fixed infrastructure but communicate in a self-organized way. The main advantages of the ad hoc network are that it has no single point of failure. A major disadvantage of the ad hoc network is that when operating across large networks, performance will degrade unless the transmitter power is enhanced.

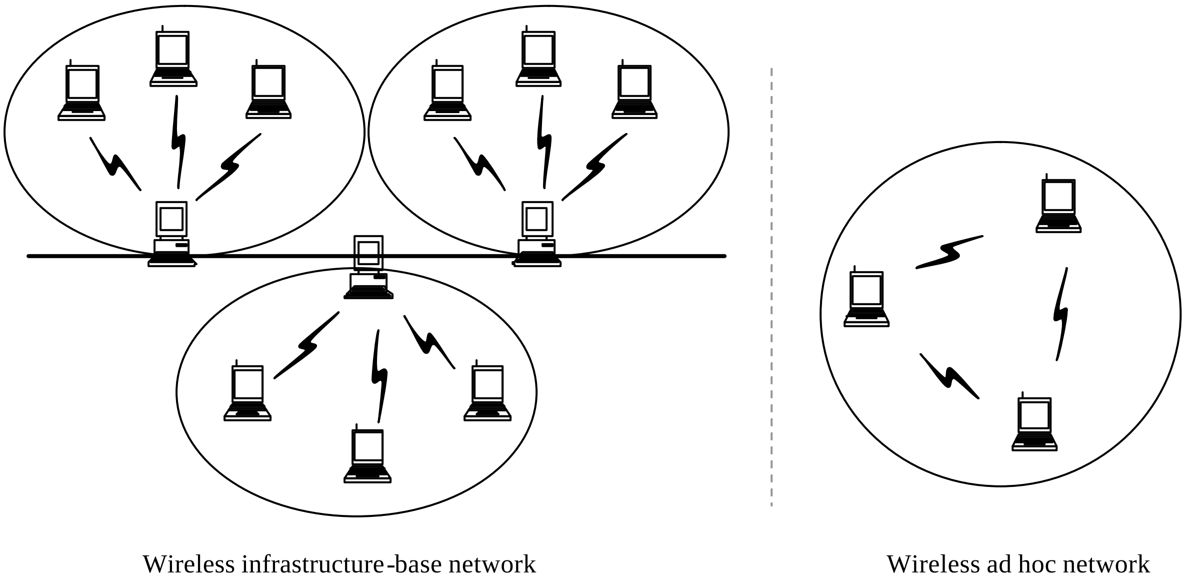


Figure 2.1 Different configurations of WLANs

An integrated wireless and wired LAN is called an infrastructure configuration. In infrastructure WLANs, the important elements are network Access Points (APs) which interface the wireless terminals with the wire-line network infrastructure. In most cases transmission is performed only between access points and wireless terminals, so two network terminals communicate via the appropriate access points. The main advantage of the network is that the network can be designed to operate relatively efficiently in the use of the signal transmission power. So compared to the ad hoc configuration, user stations in the infrastructure network can reach stations at twice the distance with the same signal power. The major disadvantage of the network is the presence of a single failure point. If the AP fails then the whole network will fail.

2.3 Technologies used in WLANs

Three distinct technologies are currently in use for WLAN systems: infrared light systems, cellular systems operating at 18-19 GHz [Buchholz et al 1991], and unlicensed spread-spectrum systems operating in the ISM[1] bands (902-908 MHz, 2.4-2.5 GHz, and 5.8-5.9 GHz) [Marcus 1991].

2.3.1 Infrared

The transmission technology used in infrared (IR) LANs is the same as that found in familiar consumer electronic devices used for remote control of television sets, VCRs, and other entertainment products, as well as in wireless keyboards and wireless printer connections for laptop computers. Transmitters used in these applications are simple and inexpensive LEDs, and receivers are similarly simple avalanche photodiodes or *p*-intrinsic-*n* photodiodes.

IR systems operate at very low power levels [Pahlavan and Levesque 1995]. As result, separate IR LANs are not likely to interfere with each other, unless placed very close together. Unlike radio transmissions, IR signals will not penetrate walls, and this is advantageous in applications where signal confinement within a walled office is desirable. Disadvantages for IR LANs are that some applications have limited range relative to radio LANs, interference from the sun and other light sources, and sensitivity to shadowing.

An Example of IR LANs is the InfraLAN Token-Ring System, manufactured by InfraLAN Technologies, Inc. The system uses the token-ring multiple access protocol and is in fact fully compliant with the IEEE 802.5 token ring standard.

2.3.2 Microwave LANs

WLANs operating at tens of gigahertz offer high data rates than which are currently achievable with most IR LANs. The signal power levels for these systems are fractions of a Watt [Pahlavan and Levesque 1995], and the signals are confined within a general area in which the network is

[1] Industry, Science and Medicine (ISM)

to be used. The nature of propagation at these frequencies is such that signals are significantly constrained by dense structures like concrete floors or many building walls. As a result, signal transmission can be very satisfactory throughout an open office area but confined enough by floors and walls to permit reuse of the same frequencies in other portions of the same building, without interference between systems. The main disadvantage of the 18-19 GHz technology is that the radio transceivers for this band are rather specialized and expensive.

An example of microwave LAN product is Altair and Altair Plus systems, which are deployed in microcells, much like cellular telephone systems [Freeburg 1991]. These products operate at 18 GHz band and provide compatibility with the IEEE 802.3 Ethernet protocol.

2.3.3 Spread-Spectrum LANs

Spread-Spectrum WLANs are designed to use in the 900 MHz, 2 GHz, and 5 GHz ISM bands and can be operated without the need for FCC[2] licensing. These systems are limited to power levels less than 1 W and is typically intended to provide signal coverage up to about 150-240 m. This coverage area is generally larger than that provided by either IR or microwave LANs. The use of spread-spectrum and transmission, combined with effective multi-user access protocols such as CSMA[3], make it feasible to deploy multiple systems in the same general area, even though signal coverage is overlapping. The drawback of unlicensed operation is that one must be able to operate in the presence of interference from other users of the same bands.

The IEEE 802.11 committee has developed a WLAN standard, IEEE 802.11 WLAN, for operation at transmission rates of 1 Mbits/sec and 2 Mbits/sec in the 2.4-2.485 GHz band. The standard specifies the two data rates for both frequency-hopped and direct-sequence spread-spectrum (FHSS and DSSS) radio transmission. The bit rate have been since increased to 11 Mbps, 22 Mbps, and 54 Mbps in the new standards; IEEE 802.11a, IEEE 802.11b, and IEEE 802.11g.

---

[2] The Federal Communications Commission (FCC) is an independent United States government agency, directly responsible to Congress. The FCC was established by the Communications Act of 1934 and is charged with regulating interstate and international communications by radio, television, wire, satellite and cable.

[3] Carrier Sense Multiple Access (CSMA)

2.4 WLAN Access Techniques

Regulating the access to the shared link is an important task in the transmission of packets on the WLANs to achieve efficient operation and good performance. Users in the wireless network seldom have need to access a channel for a long period of time. The schemes are needed for providing multi-user access to the frequency and time resources of the network in an orderly manner and in a way that minimizes transmission overhead while maximizing overall network capacity. These schemes are called Medium Access Control (MAC) protocols. The MAC protocol regulates the access to the channel by giving each node a chance to transmit its packet. This control is necessary since there is a common medium which is shared by many nodes. Without this control several nodes could be transmitting simultaneously producing corrupted messages.

The two major categories of media access control protocols for WLANs are fixed assignment and random assignment. These two major categories are briefly explained in the following subsections.

2.4.1 Fixed Assignment

In fixed assignment techniques, access to the common channel is independent of user demands and some portion of the resources of the channel is assigned to each user in a static, predetermined manner. In most fixed assignment methods there is no overhead, in the form of control messages. However, due to fixed assignment, parts of the channel might be idle even though some users have data to transmit [Rom and Sidi 1990].

The most well known fixed assignment techniques are the Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), and Code Division Multiple Access (CDMA). These techniques are briefly described below. A good discussion of these techniques can be found, for example, in [Rappaport 1996].

2.4.1.1 FDMA

FDMA is the most widespread system, and the most comprehensible. This technique is built upon the well known frequency-division multiplexing (FDM) scheme for combining non overlapping user channels for transmission as a wider-bandwidth signal. FDMA assigns different individual channels to individual users. As seen from Figure 2.2, each user is assigned a unique frequency band or channel. These channels are assigned on demand to users who requests the service. During data transmission, other users cannot share the same frequency band. FDMA is used in the cellular mobile telephone systems and in VHF and UHF land mobile radio systems. FDMA is also the most common form of multiple access in satellite networks [Bhargava et al 1981].

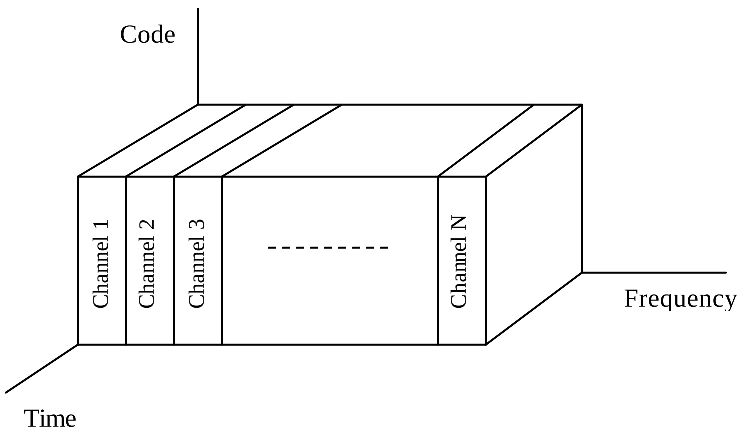


Figure 2.2 FDMA where different channels are assigned different frequency bands

2.4.1.2 TDMA

TDMA system divides the radio spectrum into time slots as shown in Figure 2.3, and in each time slot only one user is allowed to either transmit or receive. TDMA scheme is build upon time-division multiplexing (TDM). In fixed assignment TDMA operation, a transmit controller serves to assign users to time slots, and an assigned time slot is held by a user until the user releases it. At the receiving end, a user station synchronizes to the TDMA signal frame, and extracts the time slot designated for that user.

Two different forms of TDMA are being used in wireless systems; time-division duplex (TDD), and frequency-division duplex (FDD). In a TDMA/TDD system, alternating time slots on the same carrier frequency are assigned to the forward and reverse directions of the communication. However, in TDMA/FDD systems, similar frame structure would be used solely for either forward or reverse transmission, but carrier frequency would be different for the forward and reverse link.

TDMA is the access method used in current European cellular systems, GSM[4], and the United States cellular system, IS-54.

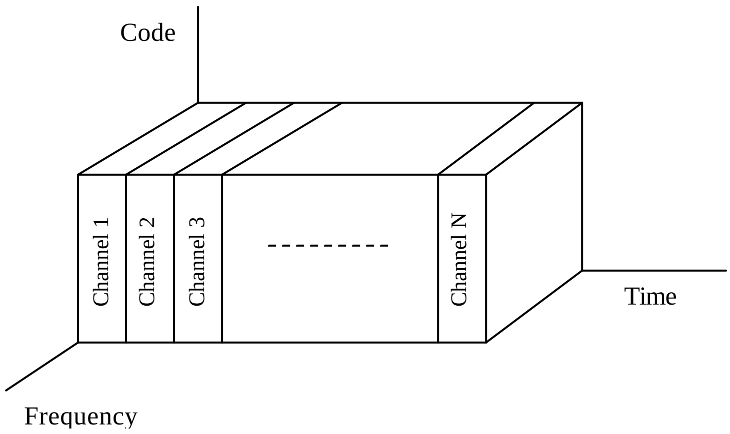


Figure 2.3 TDMA where each channel occupies cyclically repeating time slots

2.4.1.3 CDMA

The third access method is called code division multiple access (CDMA). This can be viewed as a hybrid combination of FDMA and TDMA. In CDMA, multiple users operate simultaneously over the entire bandwidth of the time-frequency signaling as shown in Figure 2.4. In CDMA systems, the narrowband message signal is multiplied by a very large bandwidth signal called the spreading signal. The spreading signal is a pseudo-random noise code sequence that has a chip rate which is orders of magnitudes greater than the data rate of the message. Each user has its own pseudorandom codeword which is approximately orthogonal to all other codewords. The spreading signal is used to provide sufficient degrees of freedom to be able to separate the user signals in the time-frequency domain. The two common CDMA techniques used are direct-sequence CDMA (DS-CDMA) and frequency hopping CDMA (FH-CDMA).

[4] Global System for Mobile

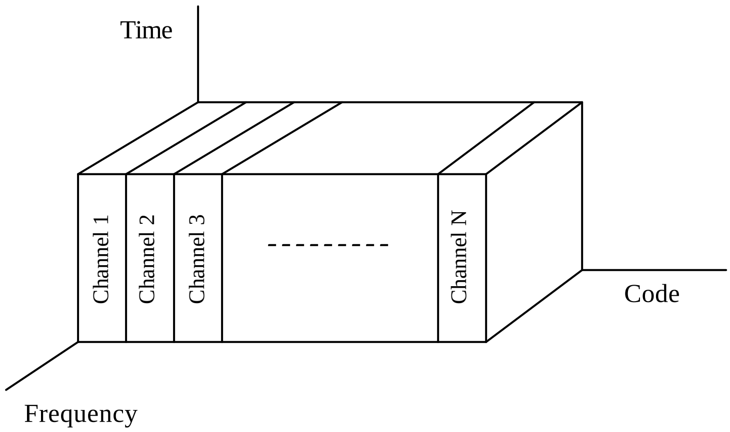


Figure 2.4 CDMA in which each channel is assigned a unique pseudo-random code which is orthogonal to pseudo-random codes used by other users

Examples of a WLAN using spread-spectrum technique is the NCR's WaveLAN [Tuc 1991], and the IEEE 802.11. The IEEE 802.11 uses DSSS and FHSS technologies operating in 2.4 GHz ISM bands.

2.4.2 Random Assignment

The fixed assignment method is suitable when each user has a steady flow of data to be transmitted. However, if the information to be transmitted is intermitted or bursty in nature, fixed assignment access methods can result in communication resources being wasted most of the time. On this contrast, random assignment access methods provide a more flexible and efficient way of managing channel access for communication of short messages. In random assignment techniques, the entire bandwidth is provided to the users as a single channel to be accessed randomly. In this method, the success of a transmission is not guaranteed in advance. The reason is that whenever two or more users are transmitting on the shared channel simultaneously, a collision occurs and the data cannot be received correctly. This being the case in this method, packets may have to be transmitted and retransmitted until eventually they are correctly received. In designing this type of access method, the main concern is transmission scheduling. The random access is probably the richest family of medium access protocols. Many local area networks today use random assignment methods due to its simplicity so that their implementation is straightforward.

Briefly discussed below are the most commonly used random access methods: (1) the basic ALOHA scheme; (2) enhanced version of the basic ALOHA, slotted-ALOHA; (3) Carrier Sense Multiple Access with Collision Detection (CSMA-CD); (4) Carrier Sense Multiple Access with Collision Avoidance (CSMA-CA). A good discussion of CSMA, and other multiple access techniques can be found, for example, in [Hammond and O'Reilly 1988].

2.4.2.1 Pure ALOHA

This protocol derives its name from the ALOHA systems, a communication network developed by Abramson and his colleagues at the University of Hawaii [Abramson 1970]. The basic idea of pure ALOHA is simple. Users transmit immediately whenever they have data to send. If the transmission is successful the receiver sends an acknowledgment to the sender within a time period equal to one propagation time. Propagation time is the time it takes a packet to travel from the sender to the receiver and back again. If no acknowledgment is received, that means the packet is lost due to a collision in the channel and the packet will be sent again with a randomly selected delay to avoid repeated collisions. The operation of the pure ALOHA is shown in Figure 2.5.

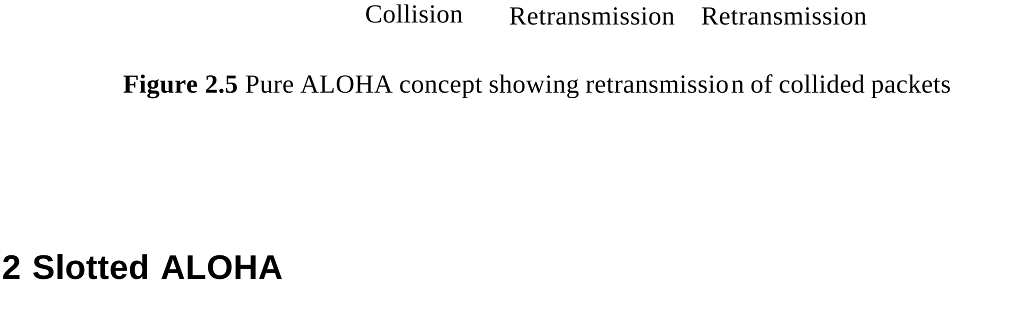


Figure 2.5 Pure ALOHA concept showing retransmission of collided packets

2.4.2.2 Slotted ALOHA

It can be shown that the maximum throughput achieved by the pure ALOHA is 0.184; see for example [Walrand 1991]. Hence, the efficiency achieved by the pure ALOHA is very low. The slotted ALOHA scheme was proposed to increase the efficiency of the ALOHA method [Roberts 1975]. In the scheme, the time axis is divided into time slots with durations equal to the time taken to transmit a packet on the channel. All the users are then synchronized to these time slots, so that when a user terminal generates a data packet, the packet is held and transmitted in the next time slot. In the slotted ALOHA, the interval of vulnerability to collision for any packet is reduced to one packet time from two packet times in the pure ALOHA, as shown in Figure 2.6.

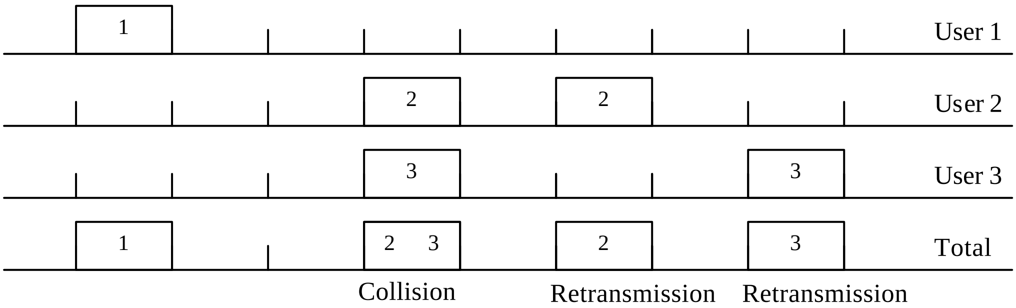


Figure 2.6 Slotted ALOHA concept, showing retransmission of collided packets

As the interval of vulnerability is reduced by a factor of two, the maximum throughput of the slotted ALOHA reaches to 0.368; see for example [Pahlavan and Levesque 1995]. So the efficiency of the slotted ALOHA is increased compare to the pure ALOHA.

Slotted ALOHA is used in wireless data communication applications where long transmission delays are encountered. The protocol is used in Satellite communication networks and other data communication systems as well.

2.4.2.3 CSMA

The efficiency of the ALOHA schemes are low due to the fact that users take no account of what other users are doing when they attempt to transmit data packets, and this leads to a high rate of packet collisions. A better result can be obtained if a user station listens to the channel before attempting to transmit a packet. This technique is used as the basis of several protocols termed carrier sense multiple access (CSMA). There are several common protocols of this type. These protocols are briefly described below.

The simplest form of CSMA is one in which each user terminal with data to transmit first listens to the channel to determine if other users are transmitting. If the channel is busy, the user terminal listens continuously, waiting until the channel become idle, and then sends a data packet immediately. This protocol is called *1-persistent CSMA*, due to its transmission strategy, which is to transmit with probability 1 as soon as the channel is available. The objective of the scheme is to avoid collisions with packets of other users.

Nonpersistent CSMA is another type of CSMA. In this type, the user station does not sense the channel continuously while the channel is busy. Instead, after sensing the busy condition, it waits a randomly selected interval of time before sensing again. As with *1-persistent CSMA*, a user with data to send begins transmitting immediately when the channel is sensed to be idle. The randomized waiting times between channel sensing eliminate most of the collisions that would result from multiple users transmitting simultaneously upon sensing the transition from busy to idle condition. This leads to much higher throughput values compared to *1-persistent CSMA*.

Another variation of the CSMA is *p-persistent CSMA*. The scheme is only usable in slotted channels. In the scheme, when a user station has data to send, it senses the channel. If the channel is sensed idle, it transmits with probability p. With probability $1-p$ the station defers action to the next time slot, where it senses the channel again. If that slot is idle, the station transmits with probability p or defers again with probability $1-p$. This procedure is repeated until the packet is transmitted or the channel is sensed to be busy. When the channel is sensed to be busy, the station then senses the channel continuously, and when the channel becomes idle, the station starts the above procedure again.

Beside CSMA, some networks have additional features, namely, Collision Detection (CD) and Collision Avoidance (CA). CSMA-CD is used only in the wired networks. The collision detection method is not suitable for wireless networks because a station is unable to listen to the channel for collisions while transmitting. The main reason is the large difference between transmitted and received signal powers at the station. As a result some WLANs use CSMA-CA instead of CSMA-CD.

CSMA-CD - In CSMA-CD, the user can detect interference among several transmissions while transmission is in progress and abort transmission of their collided packets. If this can be done sufficiently fast then the duration of an unsuccessful transmission would be shorter than that of a successful one. The operation of all CSMA-CD protocols are identical to the operation of the corresponding CSMA protocols, except that if a collision is detected during transmission, the transmission is aborted and the packet is scheduled for transmission at some later time.

CSMA-CA - As mentioned before, CSMA-CA is used in wireless networks instead of CSMA-CD. One example of a wireless network which uses this procedure is IEEE 802.11 WLANs. In CSMA/CA, the collision avoidance part is done by using the random backoff procedure. In this procedure, if a station wants to send a data frame, the station initially senses the status of the channel. If the station senses that the channel is busy then the station waits until the channel becomes idle. When the channel is sensed to be idle, the station calculates a random backoff time. After waiting this amount of backoff time, the station transmits the packet. If the packet transmission is unsuccessful the station repeats the above procedure choosing the random backoff time for a larger interval, until the packet is successfully transmitted.

2.5 Issues in WLANs

As the WLAN market grows, new features are required to meet user demands. With increasing technologies, people demand to use data-intensive, time-sensitive data traffic like audio and video around wireless networks. However, today, these traffic are combined and serviced as a single data stream with relatively simple types of data like text. So the required features of these traffic is not met. As a result of this, Quality of Service (QoS) has become a major issue in the wireless networks. Furthermore, new problems like hidden and exposed station problems are found with the use of WLANs. These two issues of the WLANs are discussed in the following sections.

2.5.1 Quality of Service

In the last few years, the IEEE 802.11 and HIPERLAN wireless LANs have been increasing in popularity. Today these WLANs, specifically IEEE 802.11 WLANs have been implemented in many different environments such as offices, homes, and other places. However these WLANs still need to be improved in the sense of QoS supported by them.

QoS is a networking term that is a bit more complex than it might sound. QoS refers to the concept of being able to control and measure data transmission rates, or throughput, and error rates. Fundamentally, QoS enables you to provide better service to certain flows. This is done by either raising the priority of a flow or limiting the priority of another flow. Delivering text and other relatively simple types of data around a network does not necessarily require complex QoS mechanisms. Hence, in the past, strict measures for QoS did not matter because the data was not multimedia and the end-user would not notice or be materially affected by latencies.

Today, the most widely used WLAN standard is IEEE 802.11. IEEE 802.11 uses two different methods to access the channel as will be explained in chapter 3. Its distributed access method, Distributed Coordination Function (DCF), is used normally for non real-time services and the centralized access mode, Point Coordination Function (PCF), is for real-time services to provide higher priority than for non-real time services. However, the centralized scheme requires the existence of access points with specialized functions. The centralized scheme has the drawback in that sense that failure of the access point leads to failure of the whole network. So in centrally controlled networks, the performance of the network very much depends on the correct operation of the access point. Due to this reason, users prefer a distributed control scheme.

The basic access method of IEEE 802.11 MAC is DCF. The frames in DCF do not have priorities, and there is no other mechanism to guarantee an access delay bound to the stations. In other words, real-time applications like voice or live video traffic have to compete with non real-time data traffic to access the channel. As a result, real-time traffic suffers in the sense of latency. To overcome this problem, the IEEE 802.11 standard needs improvement so that different priority traffic can be supported in DCF frames.

Today, researchers have proposed several different approaches to support QoS within the framework of the IEEE 802.11 WLAN standard. Sobrinho and Krishnakumar proposed a scheme called Blackburst, with the main goal of minimizing delay for real time traffic [Krishnakumar and Sobrinho 1996]. Blackburst requires that all high priority stations try to access the medium with constant intervals and also requires the ability to jam the wireless medium for a period of time. Low priority stations use the ordinary DCF access mechanism of IEEE 802.11. One of the drawbacks of this scheme is that it supports only two priority traffic classes. Other methods to support QoS in IEEE 802.11 standard by means of service differentiation have been proposed. These methods include:

- *Backoff increasing function*: Each priority class traffic has different backoff increment function.
- *DCF Inter Frame Space (DIFS)*: Each priority class traffic is assigned a different DIFS.
- *Maximum frame length*: Each priority class traffic has a maximum frame length allowed to be transmitted at once.

By using these differentiation mechanisms, Task Group E of the IEEE 802.11 working group are currently working on an extension to the IEEE 802.11 standard called IEEE 802.11e. The goal of this extension is to enhance the access mechanisms of IEEE 802.11 and provide a distributed access mechanism that can provide service differentiation. All the details have not yet been finalized, but a new access mechanism called Enhanced DCF (EDCF), which is an extension of the basic DCF mechanism, is used to achieve QoS. The detail about the IEEE 802.11e standard and other widely used wireless LAN standards, IEEE 802.11 and HIPERLAN are given in chapter 3.

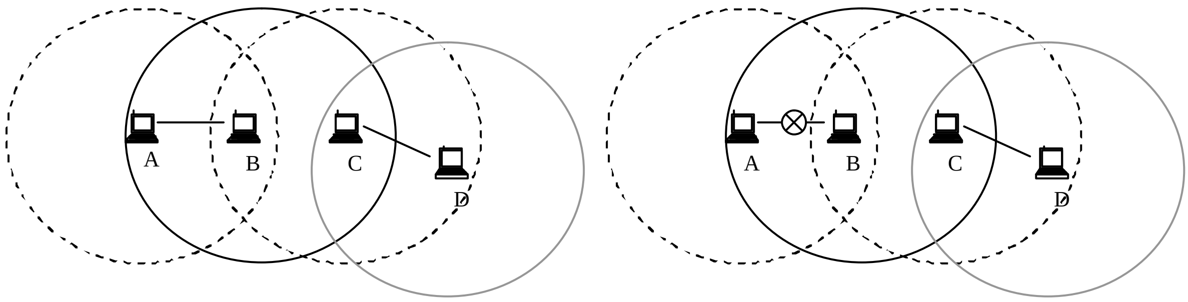
2.5.2 Hidden and Exposed Terminal Problems

Collision detection has caused many problems in networking and this is particularly the case with wireless networks. Collisions occur when two or more nodes sharing a communication medium transmit data together, the two signals corrupt each other, and the result is garbage. This has always been a problem for computer networks and the simplest protocols often do not overcome this problem. More complex protocols check the channel before transmitting data. This is very simple with Ethernet as it merely involves checking the voltage on the wire before transmitting. Coupled with this, the hidden and exposed terminal problem has been identified in wireless systems and has caused problems in collision detection. The Figure 2.7 illustrates these two situations.

In Figure 2.7a, the transmission ranges of mobile terminal A and C do not allow them to hear each other, but can both be heard by terminal B in between. The terminals A and C are hidden terminals with regards to each other. As a result, terminal A might sense the channel to be idle although terminal C is transmitting. So terminal A might start to transmit even if terminal C is transmitting. If this happens, the massage received at the terminal B will be corrupted because transmission from both terminal A and C will interfere at that point. Hence the message is unusable by terminal B. This problem is called the hidden terminal problem in wireless networks.

In order to remove the hidden terminal problem, some wireless LAN standards have introduced the four-way handshake procedure based on RTS and CTS. In this procedure, a source terminal, a terminal which wants to send a packet, first sends a RTS message to the destination terminal. The destination terminal responds to this by sending a CTS message to the source terminal. The

RTS and CTS messages contain information about how long the channel will be busy for the data transmission. So all the terminals which are in range of the source terminal will know how long the channel will be busy from RTS message and all the terminals which are in range of the destination terminal will know how long the channel will be busy from CTS message. Hence, if the CTS message is received correctly by the source terminal, all the terminals in range of both source and destination terminals will refrain from transmitting. Then the source terminal can send the data packet without a hidden terminal problem. On the other hand, some other WLANs partially avoid the hidden terminal problem by ensuring the channel sensing range to be much higher than the receiving range.



a. Hidden station problem

b. Exposed station problem

Figure 2.7 Examples of hidden and exposed station problem

In Figure 2.7b the exposed terminal problem is described. A node is exposed when it is within the sensing range of the sender but out of the interfering range of the receiver. In the case of the figure, terminal C is transmitting a packet to the terminal D. At the same time, terminal B wants to send a packet to terminal A. But terminal B is waiting until the transmission of terminal C is complete because terminal B is in range of terminal C. However, the terminal B really does not have to wait until the transmission of terminal C is complete because at terminal A, the transmission strength of station C will be very weak compared to transmission strength of station B and so the terminal A can receive correctly the transmission of the terminal B although the terminal C is transmitting at the same time. This situation is called the exposed station problem because some stations refrain from transmitting although they really do not have to.

The use of RTS and CTS access method minimizes the exposed terminal problem but can not remove completely. Hence it can deeply affect the performances of multi hop wireless ad hoc networks.

2.6 Summary

As discussed above, QoS is an important element in wireless networks, and most widely used wireless LAN standards currently do not support QoS required for emerging applications. Hence, the main aim of this thesis is to design a new distributed medium access control (MAC) scheme for WLANs to support QoS. A new MAC scheme to support QoS in WLAN environment is proposed in chapter 4 and the MAC scheme is analyzed and compared with current most widely used WLAN standard, IEEE 802.11, in chapters 5 and 6.

3.0 WLAN Standards

In the past decade, a great deal of research has been done in wireless communication. There have also been numerous new developments in this area that resulted in several different types of WLAN standards being proposed. The most widely used WLAN standards today are IEEE 802.11 WLAN and HIPERLAN. This chapter gives detailed description of the most widely accepted WLAN standards; the IEEE 802.11 WLAN and HIPERLAN. The chapter also gives a brief introduction to the upcoming WLAN standard IEEE 802.11e. The IEEE 802.11e is able to support Quality of Service (QoS) in a wireless environment by method of service differentiation.

3.1 Introduction to IEEE 802.11

The IEEE 802.11 is a part of a family of standards for local and metropolitan area networks. This family is called the IEEE 802 standard family. The relationship between the IEEE 802.11 standard and other members of the family is shown in Figure 3.1. The IEEE 802 standards define different media-access technologies and the associated physical media. The component parts of the IEEE 802 standards are as follows:

802.1 Overview, Internetworking, and Systems Management
802.2 Logical link control
802.3 Carrier sense multiple access – collision detection (CSMA-CD) bus
802.4 Token bus
802.5 Token ring
802.6 Metropolitan area networks (MANs)
802.9 Integrated voice and data LANs
802.10 Interoperable LAN/MAN Security
802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer Specifications
802.12 Demand Priority Access Method, Physical Layer and Repeater Specifications

The scope of the IEEE 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," [IEEE 802.11 1999]. There are two main purposes of this standard. First it is to provide wireless connectivity to automatic machinery, equipment, or stations that require rapid deployment, which may be portable or hand-held, or which may be mounted on moving vehicles within a local area. Secondly the standard offers regulatory bodies a means of standardizing access to one or more frequency bands for the purpose of local area communication.

The IEEE 802.11 standard describes mandatory support for 1 Mbps WLAN with optional support for a 2Mbps data transmission rate [Crow et al 1997]. In this standard, support for asynchronous data transfer is specified as well as optional support for distributed time-bounded services like packetized voice and video.

The basic building block of the IEEE 802.11 architecture is the *basic service set* (BSS). A BSS is defined as a group of stations that are under the direct control of a single coordination function. The geographical area covered by the BSS is known as *basic service area* (BSA). All the stations in the same BSS can communicate with each other directly. Hence, if a station moves out of its BSS, it can no longer directly communicate with other members of the BSS.

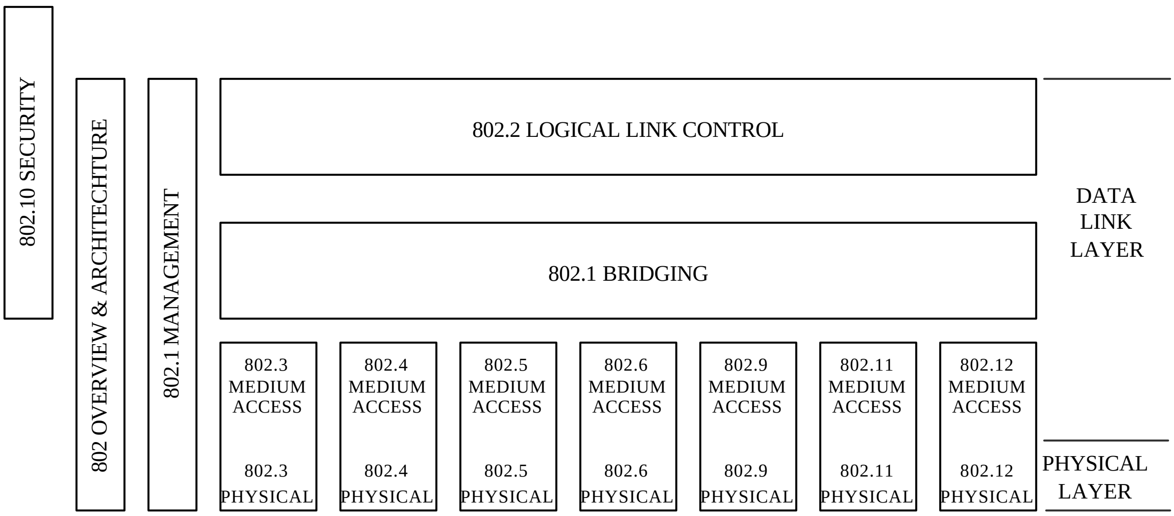


Figure 3.1 IEEE family of protocols

The most basic type of IEEE 802.11 WLAN is the independent BSS (IBSS). This mode of operation is only possible if all the stations in the WLAN are able to communicate directly and is called *an ad hoc network*. The smallest IEEE 802.11 *ad hoc network* may consist of only two stations.

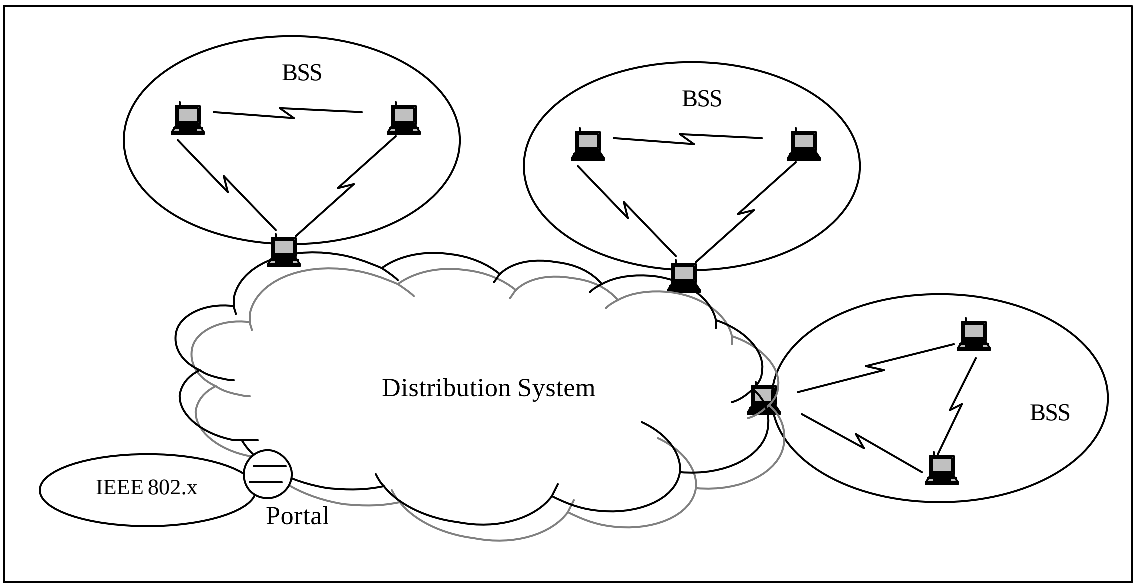


Figure 3.2 IEEE 802.11 infrastructure network

There are physical layer limitations that determine the distance between stations in the WLAN so that they can directly communicate with each other. When the stations are further away than this limited distance the stations cannot form an IBSS. Hence, to overcome these kinds of problems, the IEEE 802.11 standard also supports infrastructure networks. Infrastructure networks consist of more than one BSS. These BSSs can communicate through a common *distribution system* (DS) by using access points (APs), as shown in Figure 3.2. An AP is a station that provides access to DS by providing DS services in addition to acting as a station. Each DS is implementation independent and could be a wired IEEE 802.3 Ethernet LAN, IEEE 802.4 token bus LAN, IEEE 802.5 token ring LAN, or any another IEEE 802.11 wireless medium. The DS and BSSs allow IEEE 802.11 to create a wireless network of arbitrary size and complexity. This type of network is referred to as the *extended service set* (ESS) network.

The IEEE 802.11 architecture can be integrated with a traditional wired LAN by using a *portal*. A portal is the logical point at which packets from an integrated non-IEEE 802.11 LAN enter the IEEE 802.11 DS. It is possible for one device to offer both the functions of an AP and a portal. This is also shown in Figure 3.2. The portal incorporates functions, which are the same as a bridge so that it provides range extension and translation between different frame formats.

3.1.1 IEEE 802.11 Physical layer

Figure 3.3 shows elements of the protocol stack of IEEE 802.11. The IEEE 802.11 physical layer is divided into two sublayers. The lower one is called the *Physical Medium Dependent* (PMD) sublayer and the other one is called the *Physical Layer Convergence Procedure* (PLCP) sublayer. The PMD sublayer defines the characteristics of a wireless medium between two or more stations, and the method of transmitting and receiving data through the wireless medium. The main function PLCP is to map the MAC sublayer protocol data units (MPDUs) into a framing format suitable for sending and receiving user data and management information between two or more stations using the associated PMD sublayer. Other task of the PLCP sublayer is carrier sensing, whose result is transferred to the MAC sublayer [Wesolowski 2002].

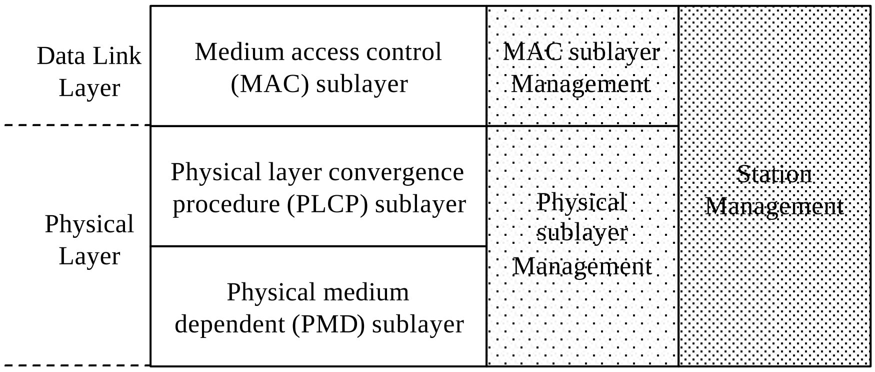


Figure 3.3 IEEE 802.11 OSI Reference Model

IEEE 802.11 have standardized three different types of physical layer implementations: direct sequence spread spectrum (DSSS), frequency hoping spread spectrum (FHSS) and Infrared (IR). The frames sent over the channel generally consist of three parts. The physical layer adds the PLCP preamble and the PLCP header to each MAC packet data unit that arrives from the MAC sublayer. The contents of the PLCP preamble and header depend on the type of physical layer implementation used. A brief description of each type of physical layer implementation and frame details are given in the following.

3.1.1.1 DSSS Physical Layer

In the DSSS technique, the user data is represented by a sequence of pulses (chips) of much higher rate than the original data bits. This implementation uses 2.4 GHz Industrial, Scientific and Medical (ISM) frequency band. The DSSS system provides a wireless LAN with both a 1 Mbps and a 2 Mbps data payload communication. The 1 Mbps and 2 Mbps basic data rates are encoded using differential binary phase shift keying (DBPSK) and differential quadrature phase shift keying (DQPSK), respectively.

A preamble and a header is added to each MPDU arrived from the MAC sublayer as shown in Figure 3.4. The PLCP Preamble contains the following fields: Synchronization (Sync) and Start Frame Delimiter (SFD). The PLCP Header contains the following fields: IEEE 802.11 Signaling (Signal), IEEE 802.11 Service (Service), LENGTH (Length), and CCITT CRC-16. The PLCP part of each frame is always transmitted at the rate of 1 Mbps and the payload part can be transmitted either at 1 Mbps or 2 Mbps. The function of the PLCP preamble and header fields are as follows:

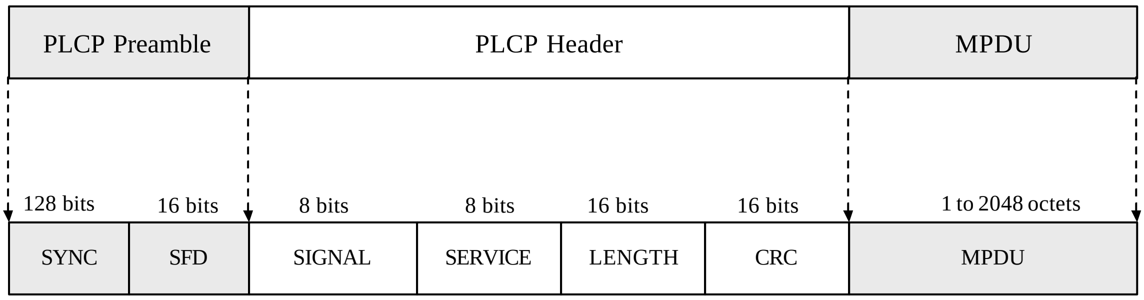


Figure 3.4 DSSS PLCP frame format

- *Synchronization* (SYNC): used for synchronization, gain setting, frequency offset compensation and energy detection.
- *Start Frame Delimiter* (SFD): used for frame synchronization.
- *IEEE 802.11 Signaling* (SIGNAL): used to indicate the data rate used for payload transmission.
- *IEEE 802.11 Service* (SERVICE): reserved for future use.
- *Length* (LENGTH): used to indicate the length of the MPDU field.
- *CCITT CRC-16* (CRC): used for header error control.

DSSS implementation of the physical layer ensures high data rates and a wide range of coverage area. However, the DSSS products cost more and use more power than the FHSS techniques used in IEEE 802.11 WLANs [Wesolowski 2002].

3.1.1.2 FHSS Physical Layer

The FHSS physical layer implementation has properties such as high distortion immunity, high system capacity, low power consumption, medium range and low cost of RF parts [Wesolowski 2002]. The FHSS also uses the 2.4 GHz ISM frequency band. In the United States and Europe, a maximum of 79 channels are specified in the hopping set. Three different hopping sequence sets are established with 26 hopping sequence per set. This has an advantage that multiple BSSs can coexist in the same geographical area allowing to maximize throughput and minimize congestion. The data are transmitted using Gaussian-shaped Frequency Shift Keying (GFSK) modulation. In 1 Mbps data rate, two level modulation is used, and in 2 Mbps it is four levels GFSK.

The FHSS physical layer frame format is shown in Figure 3.5. As in the DSSS physical layer frame, the FHSS physical layer frame has three parts, the PLCP Preamble, the PLCP Header and MPDU. However the FHSS frame contents of each of these parts differ from that of the DSSS frame. The PLCP Preamble of FHSS provides a period of time for several receiver functions. These functions include antenna diversity, clock and data recovery, and field delineation of the PLCP Header and the MPDU [IEEE 802.11 1999]. The PLCP Header is used to specify the length of the whitened MPDU field and supports any PLCP management information. The functions of the PLCP part are as follows:

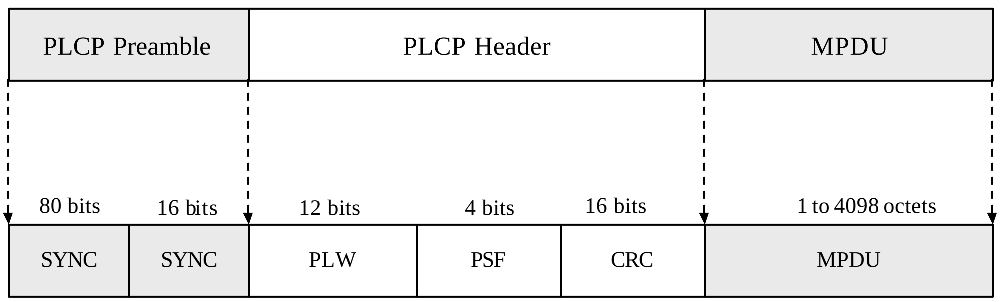


Figure 3.5 FHSS PLCP frame format

- *Synchronization* (SYNC): used for synchronization of the receivers and signal detection by the Clear Channel Assessment (CCA) process.
- *Start Frame Delimiter* (SFD): indicates the start of the frame.
- *PLCP-PDU Length Word* (PLW): used to indicate the length of the payload field.

- *PLCP Signal Field* (PSF): indicates the data rate of the payload.
- *CCITT CRC-16* (CRC): used for header parity control.

As in the DSSS, the PLCP part of the frame is always sent with a data rate of 1 Mbps but the payload data can be sent either at 1 Mbps or at 2 Mbps.

3.1.1.3 IR Physical Layer

In this implementation, signals are sent using infrared rays of wavelength in the range 850 to 950 nm. As in other physical layer implementations, the IR physical layer also supports 1 and 2 Mbps data rate. Encoding of 1 Mbps data rate is performed using 16-pulse position modulation (PPM), where 4 data bits are mapped to 16 coded bits for transmission. On the other hand, 2 Mbps data rate is encoded using 4-PPM modulation, where 2 data bits are mapped to 4 coded bits for transmission.

The PLCP frame format for IR physical layer implementation is shown in Figure 3.6. The frame also contains three parts: the PLCP Preamble, the PLCP Header and MPDU. The functions of the fields of PLCP part of the frame are given below.

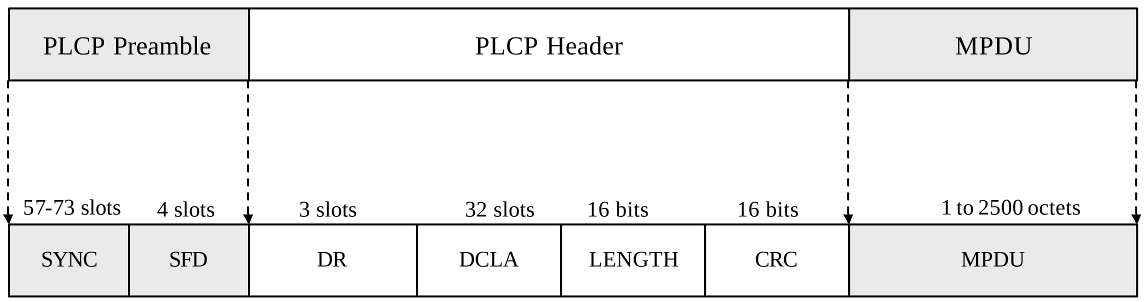


Figure 3.6 IR PLCP frame format

- *Synchronization* (SYNC): This field is provided so that the receiver can perform clock recovery (slot synchronization), automatic gain control, signal-to-noise ratio estimation, and diversity selection.
- *Start Frame Delimiter* (SFD): indicates the start of the frame.
- *Data Rate* (DR): indicates the data rate of the payload

- *PLCP DC Level Adjustment* (DCLA): used for the receiver to stabilize the DC level after the SYNC, SFD, and DR fields.
- *Length* (LENGTH): used to indicate the length of the MPDU field.
- *CCITT CRC-16* (CRC): used for header parity control.

The IR implementation is the cheapest of all IEEE 802.11 physical layer implementations and does not need any frequency regulation. But, it has the lowest range among all other implementations and has to operate indoors as ceilings, which could reflect the infrared signals, are required.

3.1.2 IEEE 802.11 Medium Access Control Sublayer

The main responsibilities of the MAC sublayer of IEEE 802.11 include: channel allocation, protocol data unit (PDU) addressing, frame formatting, error checking, and fragmentation and reassembly of data blocks. The WLAN can operate in two basic modes.

- *Contention mode:* All the stations that need to transmit have to compete for the channel access for each packet transmitted. Networks operating only in this mode are known as *ad hoc* networks. Contention services imply that each station with an MPDU queued for transmission must contend for access to the channel and, once the MPDU is transmitted, must recontend for access to the channel for all subsequent frames. Contention services promote fair access to the channel for all stations.
- *Mixed mode:* The medium will alternate between contention mode, known as *contention period* (CP), and *contention free period* (CFP). During the CFP the medium is controlled by the access point (AP), hence stations do not need to compete for the channel access during the CFP. The AP informs other stations the start of the CFP by sending a beacon.

The stations in the IEEE 802.11 WLANs use three different types of frames to support functions of the MAC sublayer. These are management frames, control frames, and data frames. The management frames are used for association, disassociation with the AP, and timing and synchronization. The uses of the Control frames are handshaking during the CP, positive acknowledgments during the CP, and to end the CFP. Data frames are used for the transmission of data during the CP and CFP, and can be combined with polling and acknowledgments during the CFP.

3.1.2.1 General MAC frame format

The general MAC frame format of IEEE 802.11 is given in Figure 3.7. Each frame consists of three basic components. A MAC Header, a frame body, and a frame check sequence (FCS). The Mac Header has a fixed length component and contains fields for frame control, duration, address, and sequence control information. The variable length frame body contains information specific to frame type. The FCS field contains IEEE 32 bit cyclic redundancy code (CRC). The fields Address 2, Address 3, Sequence Control, Address 4, and Frame Body are only present in certain frame types. The functions of MAC frame fields are:

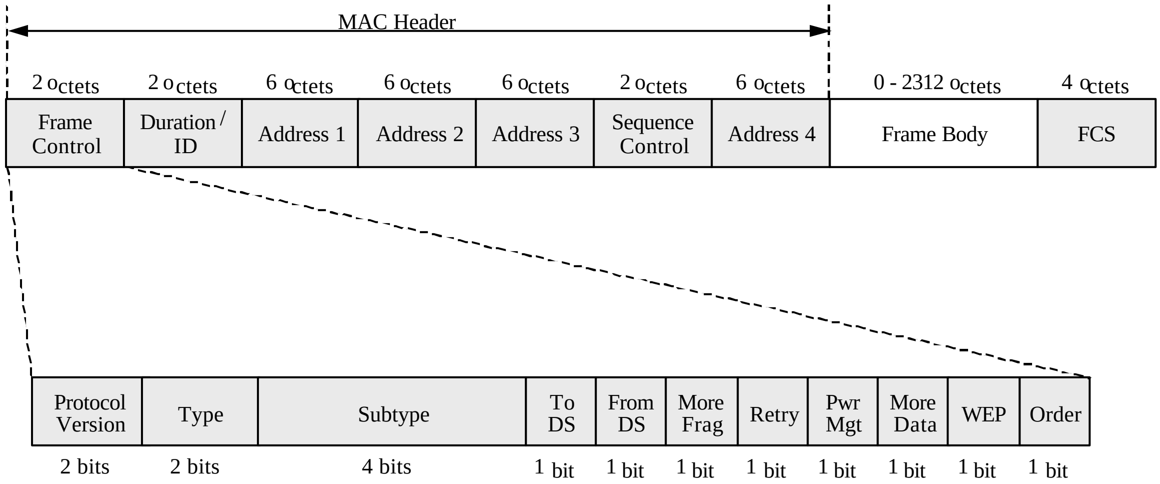


Figure 3.7 General MAC frame format

- *Frame Control*: As shown in Figure 3.7, this field contains several other subfields. There are subfields to determine the protocol version and type of frame. This field also indicates whether the frame is fragmented and whether the frame is directed to the distributed system (DS), arrives from it or whether the source and destination are mobile terminals or APs. The More Fragments field is used to indicate another fragment of the current MPDU. The Retry field is set to 1 in any data or management type frame that is a retransmission of an earlier frame. The Power Management field is used to indicate the power management mode of a station. The Wired Equivalent Privacy (WEP) is set to 1 if the Frame Body field contains information that has been processed by the WEP algorithm.
- *Duration ID*: The field determines the period of time during which the channel will be busy.

- *Address fields*: These fields indicate the source and destination of the transmitted frame and are interpreted depending on the frame control bits determining the address meaning.
- *Sequence Control*: This field contains the frame sequence number and is used to avoid frame duplication which would occur due to the acknowledgement mechanism.
- *Frame body*: This field carries the payload of length 2312 octets.
- *FCS*: This field carries IEEE 32 bit cyclic redundancy code (CRC) and is used for acknowledgement procedure for error detection.

3.1.2.2 Distributed Coordination Function (DCF)

The fundamental access method of the IEEE 802.11 MAC is Distributed Coordination Function (DCF). The DCF access mechanism only supports asynchronous data transfer on the best effort basis. The MAC architecture of IEEE 802.11 is shown in Figure 3.8. As seen from the figure, all the stations in the BSS must support the DCF channel access method. The DCF operates solely in ad hoc networks, and either operates solely or coexists with the Point Coordination Function (PCF) in infrastructure networks. In DCF access method, each station with an MPDU queued for transmission must contend for access to the channel and, once the MPDU is transmitted, it must recontend for access to the channel for all subsequent frames. The DCF mechanism promotes fair access to the channel for all stations.

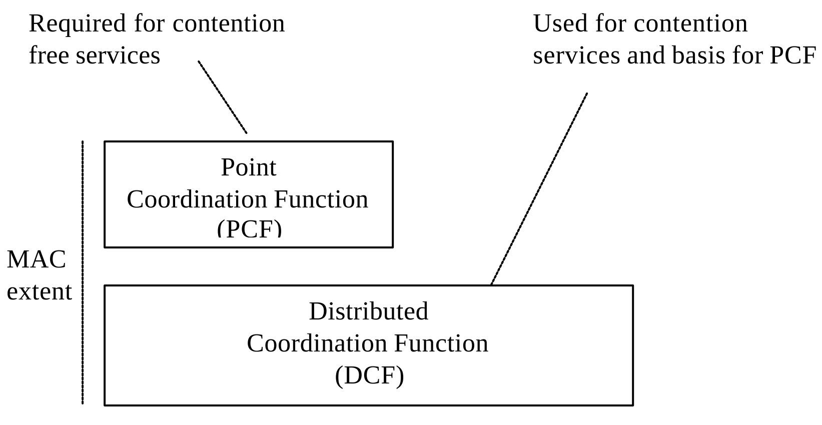


Figure 3.8 General MAC frame format

The DCF mechanism is totally dependent on carrier sense multiple access with collision avoidance (CSMA/CA). As in wired networks, CSMA/CD (collision detection) method is not suitable for the wireless networks because a station is unable to listen to the channel for detecting collisions while transmitting. The main reason being the large difference between transmitted and received signal power at the station.

In IEEE 802.11, the state of the medium is determined by using carrier-sensing functions. Carrier sensing is done at both the air interface known as *physical carrier sensing* and also at the MAC sublayer referred as *virtual carrier sensing*. A physical-carrier sense mechanism is performed by the physical layer and this information is passed to the MAC sublayer. In physical-carrier sensing, stations detect the presence of carrier waves in the medium via relative signal strength from other sources.

The MAC provides a virtual carrier-sense mechanism. This mechanism is referred to as the network allocation vector (NAV). The NAV indicates the amount of time that must elapse until the current transmission session is complete and the channel can be sampled again for idle channel. The NAV predicts this time period based on duration information that is announced in Request to Send (RTS) or Clear to Send (CTS) frames prior to the actual exchange of data. The duration information is also available in the MAC headers of almost all frames sent during the CP. A station updates the NAV with the information received in the Duration/ID field when the station receives a valid frame. But the station updates the NAV, when receiving a correct frame, only when the new NAV value is greater than the current NAV value and only when the frame is not addressed to the station.

The physical carrier-sensing mechanism combines the NAV state and the transmitter status of the station to determine the busy or idle state of the medium. The NAV may be thought of as a counter, which counts down to zero at a uniform rate. When the counter is zero, the virtual carrier-sense indication is that the medium is idle and when nonzero, the indication is busy. The medium is determined to be busy whenever the STA is transmitting.

The time interval between frames is called interframe spaces (IFS). Stations in the BSSs access the medium by using IFS. Four different IFSs are defined to provide priority levels for access to the wireless media: short IFS (SIFS), point coordination function IFS (PIFS), distributed coordination function IFS (DIFS), and extended IFS (EIFS). The SIFS is the shortest IFS interval, followed by PIFS, followed by DIFS and then EIFS. The different IFSs are independent of the station bit rate. The IFS timings are defined as time gaps on the medium, and are fixed for each type of physical layer implementation. As SIFS is the shortest of the IFSs, those stations that has to wait a SIFS to access the channel has the highest priority compared to those stations that have to wait a PIFS, DIFS, or EIFS.

Figure 3.9 shows a timing diagram illustrating the successful transmission of data frames using the basic access method. A station that is using the basic access method to send a data frame first has to sense that the channel is idle. The station uses physical and virtual carrier sensing mechanisms to find the status of the channel. When the station finds the channel is idle then it has to wait a DIFS time and sense the channel again. If now the channel is idle, the station starts to transmit its data frame. At the destination station, when the frame is received, the station checks the checksum of the frame to determine whether the frame is received correctly. If the frame is received correctly then the destination station sends an acknowledgement frame (ACK) back to the source station after waiting a SIFS time. The duration field in the frame transmitted is used to let other stations know how long the medium will be busy. So, by using the duration field, all the other stations in the BSS update their NAV. This duration includes the SIFS interval and the ACK following the data frame.

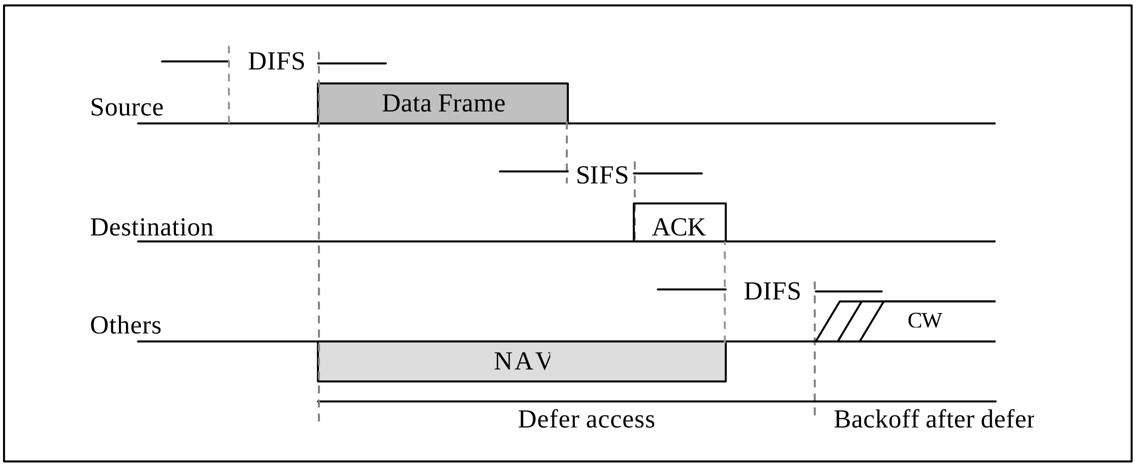


Figure 3.9 Successful data transmission in the basic access method

Collisions are possible to occur due to simultaneous transmissions of frames from two different stations in the BSS. When a collision has occurred, the MPDU transmitted may be corrupted and it may not be possible to receive the MPDU frame correctly by the destination station. In wireless environment, the source is unable to listen for collisions while transmitting so that it will continue to transmit until the frame transmission is complete. If the MPDU is large, a lot of channel bandwidth will be wasted due to corrupted MPDU. RTS and CTS control frames can be used by a station to reserve channel bandwidth prior to the transmission of an MPDU and to minimize the amount of bandwidth wasted when collisions occur. This method is also advantageous to prevent a hidden station problem within the medium. RTS and CTS control frames are relatively small (RTS is 20 octets and CTS is 14 octets) when compared to the maximum data frame size (2346 octets). In this method of data transmission, the source station first sends a RTS frame to the destination station. If destination station receives the RTS frame correctly then the destination station sends back a CTS frame to the source station after waiting SIFS interval as shown in Figure 3.10. The two frames (RTS and CTS) also contain a duration field to let other stations know when the data transmission will be completed. Hence, all the other stations which receive either RTS or CTS or both correctly will set their NAV to an appropriate value given in the duration field of the frames. When the CTS frame is received correctly by the source station, it knows that the medium is free. Then it waits for SIFS interval and sends the data frame. But if the CTS is not received by the source station then it has to try again to send the data frame after backoff. If the data transmission is successful the source station will receive an ACK frame form the destination station.

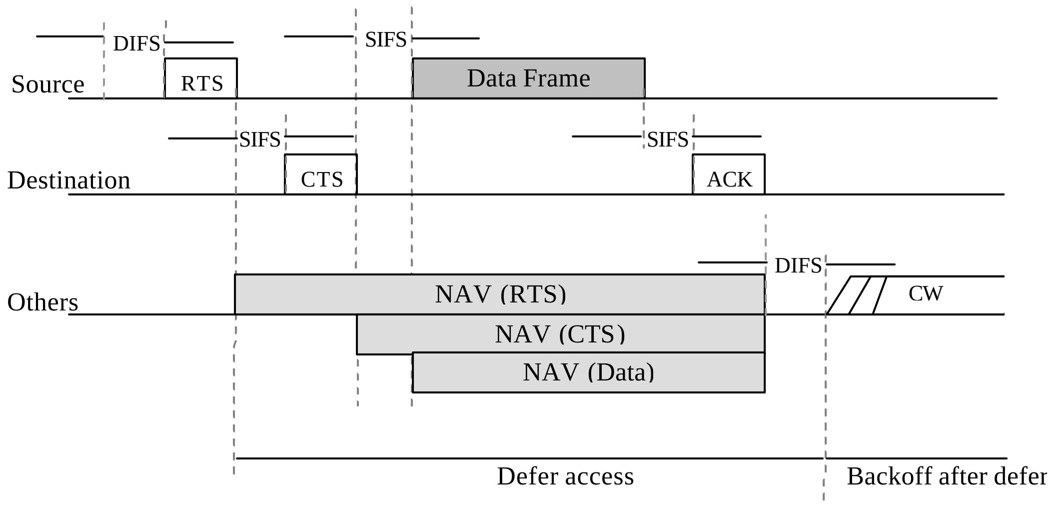


Figure 3.10 Successful data transmission by using RTS/CTS method

Stations can choose to never use RTS/CTS, use RTS/CTS whenever the MSDU exceeds the value of RTS_Threshold (manageable parameter), or always use RTS/CTS. If a collision occurs with an RTS or CTS MPDU, far less bandwidth is wasted when compared to a large data MPDU. However, for a lightly loaded medium, additional delay is imposed by the overhead of the RTS or CTS frames.

If a frame body (MSDU) that comes from the logical link control (LLC) to the MAC is large then it may require fragmentation to increase the reliability. In the MAC sublayer there is a manageable parameter called Fragmentation_threshold to decide whether the MSDU needs fragmentation. If the fragmentation is done for an MSDU then all the fragments can be sent one after another when the station has got access to the channel.

In CSMA/CA, the collision avoidance part is done by using random backoff procedure. In this procedure, if a station wants to send a data frame, the station initially senses the status of the channel. If the station found that the channel is busy then the station waits until the channel becomes idle for DIFS interval and then calculates a random backoff time. In IEEE 802.11 the time is slotted in time periods that correspond to a Slot_Time. In this protocol the Slot_Time is used to define the IFS intervals and backoff time of a station in the contention period. The random backoff time is an integer value that corresponds to a number of time slots.

Initially, the station computes a backoff time in the range 0–7. After the medium becomes idle after a DIFS period, stations decrement their backoff timer until the medium becomes busy again or the timer reaches zero. If the timer has not reached zero and the medium becomes busy, the station freezes its timer. When the timer is finally decremented to zero, the station transmits its frame. If two or more stations decrement to zero at the same time, a collision will occur, and each station will have to generate a new backoff time in the range 0–15. For each retransmission attempt, the backoff time grows as $\lfloor 2^{2+i}.randf\,() \rfloor$ Slot_Time, where i is the number of consecutive times a station attempts to send an MPDU, *randf*() is a uniform random variable in (0,1), and $\lfloor x \rfloor$ represents the largest integer less than or equal to x. The idle period after a DIFS period is referred to as the *contention window* (CW).

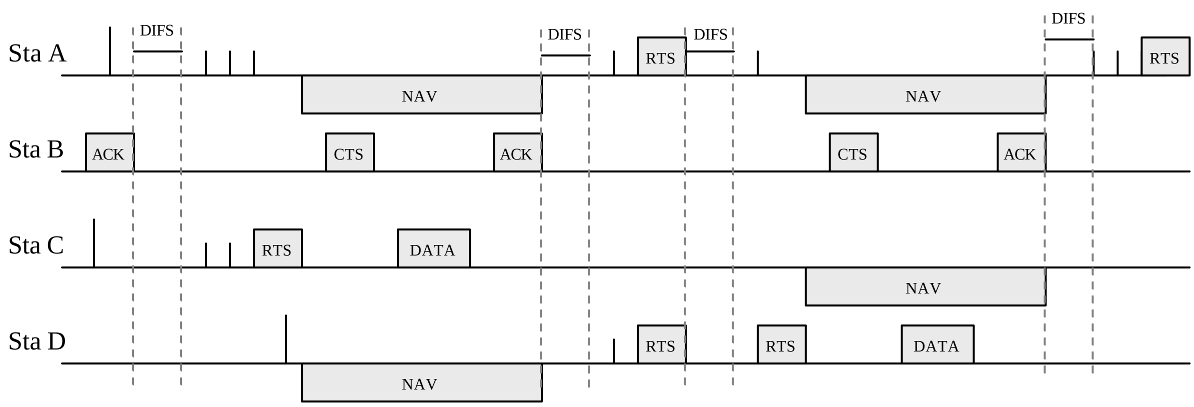


Figure 3.11 An example of collision avoidance by using the backoff procedure

Figure 3.11 shows an example of how the backoff procedure works to avoid collisions in the channel. The figure shows that there are four stations in the channel and at the start the channel is busy due to ACK frame transmission of station B. During that time stations A and C got ready to transmit a frame but sensed the channel was busy. As a result these stations wait DIFS interval of channel free time and calculate a backoff value. Station A got a backoff value of 5 and C got a value of 3. Then station C waits for 3 slot time and started its data transmission. During the data transmission of station C, stations A and D set their NAV to appropriate values and also station D got ready to transmit a frame. Station D waits for DIFS time and calculated a backoff value. Now it is found that the remaining backoff value of station A is the same as the backoff value obtained by station D. As a result, a collision has occurred and these two stations calculated a new backoff value in between 0-15. This time A got backoff value of 2 and D got a value higher than 2. Hence station A sends its frame successfully.

In DCF, all stations have equal probability of gaining access to the channel after each DIFS interval. Time-bounded services typically support applications such as packetized voice or video that must be maintained with a specified minimum delay. With DCF, there is no mechanism to guarantee minimum delay to stations supporting time-bounded services. However, the Point Coordination Function (PCF) does support delay sensitive data with help of APs.

3.1.2.3 Point Coordination Function (PCF)

The IEEE 802.11 MAC may also incorporate an optional access method called a PCF, which is only usable in infrastructure network configurations. This access method uses a point coordinator (PC), which operates at the access point of the BSS, to determine which station currently has the right to transmit. The operation is essentially that of polling, with the PC performing the role of the polling master. The access priority provided by a PCF may be utilized to create a *contention-free* (CF) access method. The PC controls the frame transmissions of the stations so as to eliminate contention for a limited period of time. The PC shall reside in the AP. It is an option for an AP to be able to become the PC. All stations inherently obey the medium access rules of the PCF, because these rules are based on the DCF, and they set their NAV at the beginning of each CF period (CFP).

The CFP repetition interval (CFP_Rate) is used to determine the frequency at which the PCF occurs. Within a repetition interval, a portion of the time is allotted to contention-free traffic, and the remainder is provided for contention-based traffic as shown in Figure 3.12. The CFP repetition interval is initiated by a beacon frame, where the beacon frame is transmitted by the AP. The primary functions of the beacon are synchronization and timing. The duration of the CFP repetition interval is a manageable parameter that is always an integral number of beacon frames.

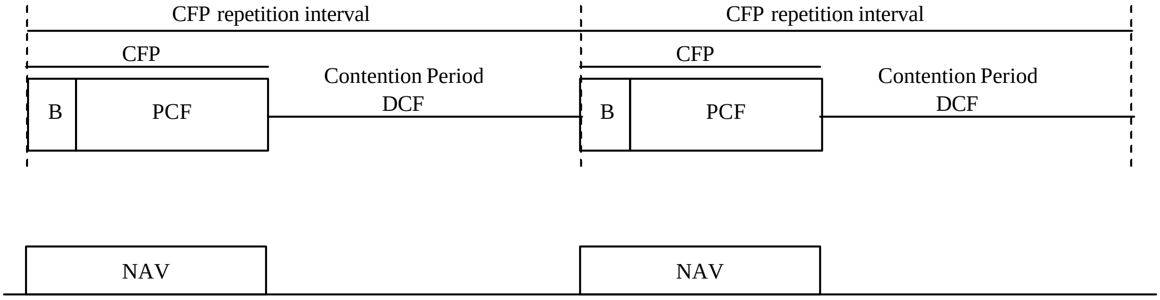


Figure 3.12 PCF and DCF alteration

The CFP is started by the AP by sending a beacon. To send the beacon the AP senses the medium. If the medium remains idle for a PIFS interval, the PC transmits a beacon frame to initiate the CFP. During the contention free period, the AP controls the channel by polling the stations which need data to transmit during CFP. If the stations have no data to transmit during CFP, the AP can terminate the CFP by sending CF-end frame. After CF-end frame, the CP is started again and it will remain so until a CFP starting beacon is transmitted by the AP.

The PCF is used for delay sensitive data to meet their delay requirements. Although the PCF performs its function well in fine conditions, this method is a centrally controlled approach and is not very much dependable. If the AP fails, then the PCF will also fail. As a result, researchers have become interested in distributed controlled methods to also support delay sensitive data.

3.2 IEEE 802.11e protocol

The IEEE 802.11 PCF supports limited Quality of Service (QoS) for the delay sensitive data, but this scheme has problems. These problems include the unpredictable beacon delays and unknown transmission durations of polled stations [Mangold et al 2002]. Furthermore, PCF is only suitable for infrastructure networks, which has an AP to control the channel. The infrastructure networks are centrally controlled networks, and it is concluded in [Visser and Zarki 1995], that the centrally controlled networks perform poorly. Attempts to solve these problems of PCF led to enhancements of the IEEE 802.11 protocol. Several researchers, for eaxmple [Deng and Chang 199] and [Aad and Castelluccia 2001] have proposed methods to support QoS in IEEE 802.11 by using a distributed approach. On their basis, IEEE 802.11 Task Group E currently works on enhancements to IEEE 802.11 MAC, called IEEE 802.11e. It introduces Enhanced DCF (EDCF) and Hybrid Coordination Function (HCF). Stations, which operate under IEEE 802.11e, are called enhanced stations, and an enhanced station, which may optionally work as the centralized controller for all other stations within the same Quality of Service supporting BSS (QBSS), is called the Hybrid Coordinator (*HC*).

The IEEE 802.11e also has two phases of operation, i.e., a contention period (CP) and a contention free period (CFP), which alternate over time continuously. The EDCF is used in the CP only, while the HCF is used in both phases. The EDCF is the basis for HCF. HCF can also extends the EDCF access rules. The HC may allocate transmission opportunity (TXOP) to itself, to initiate MSDU deliveries whenever it wants, only after detecting the channel has been idle for PIFS. Here, only the EDCF access method is discussed because it is a distributed controlled access scheme that supports QoS and it needs to be compared with the proposed distributed controlled access scheme in chapter 4.

3.2.1 Enhanced Distributed Coordination Function (EDCF)

EDCF introduces a prioritization mechanism based on different traffic categories (TCs). Each TC has different queue, different IFS and contention window (CW) parameters as shown in Figure 3.13. The AIFS and CW parameters are set so that higher priority TC has better chance to access the channel. As in DCF, each TC fights for the channel access according to its AIFS and CW parameters. As seen in Figure 3.13, in IEEE 802.11e each station can implement a maximum of eight TCs. The collisions between the TCs are resolved by the scheduler. The scheduler gives the TXOP to the highest priority TC, among the TCs that are collided within the stations. Although the internal collision can be resolved with the help of the scheduler, collisions between stations can still occur. These collisions can be resolved by using the collision avoidance method.

If a TC of a station needs to transmit data, the station contends for TXOP. A TXOP is defined as an interval of time when a station has the right to initiate transmission, defined by starting time and the maximum duration. If the station finds that the channel is busy then it has to wait until the channel is idle for an Arbitration Interframe Space (AIFS) relevant to the TC (*AIFS[TC]*). After waiting *AIFS[TC]* time, backoff counter of the TC is set to a random number drawn from the interval [*1, CW +1*]. The minimum size of the CW (*CWmin[TC]*) and the maximum size of the CW (*CWmax[TC]*) are parameters of the CW and depend on the TC that supports QoS between different TCs. These are shown in Figure 3.14. As shown in the figure, there are three different priority TCs (Low, Medium and High). The lower priority TC packet has to wait longer AIFS time before starting to decrement its backoff counter. Hence as a result, the higher priority TC's backoff counter decrements more faster rate than the lower priority TCs so that the higher priority TC has better chance to access the channel.

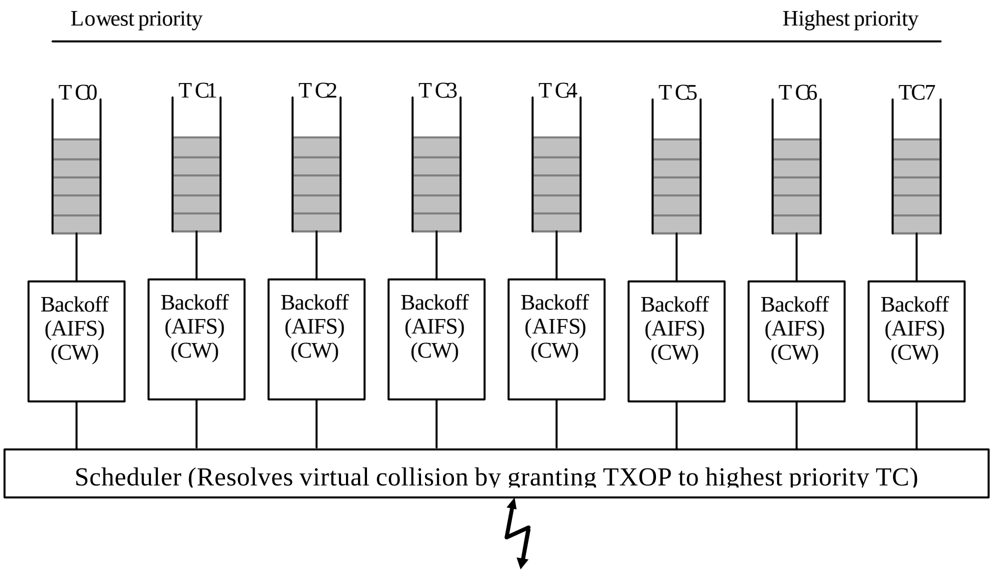


Figure 3.13 Traffic categories (TCs) of a station with the scheduler to resolve internal collisions

As in IEEE 802.11 DCF, when the medium is found busy before the backoff counter of the TC reaches zero, the backoff counter has to wait for the medium to be idle for *AIFS[TC]* again before continuing to count down the backoff counter. A big difference from the DCF to the EDCF is that in DCF, the backoff counter is reduced by one, beginning with the first slot interval after the DIFS period. But in EDCF, the backoff counter is reduced by one after channel has been idle for the period of an AIFS. After any unsuccessful transmission attempt, a new CW is calculated with the help of the persistence factor *PF[TC]* and another uniformly distributed backoff counter out of this new, enlarged CW is drawn, to reduce the probability of a new collision. The PF is also another parameter that helps to support QoS between TCs. In IEEE 802.11 DCF, CW is always doubled after any unsuccessful transmission (equivalent to *PF=2*), 802.11e uses the PF to increase the CW differently for each TC.

After an unsuccessful transmission attempt, the new CW for the TC can be calculated by using the equation given below:

$$newCW[TC] \geq ((oldCW[TC]+1)PF)-1$$

When the CW reaches the maximum value, it stays the same until the transmission is successful or until the maximum retry limit is reached. When the maximum retry limit is reached, the backoff counter resets.

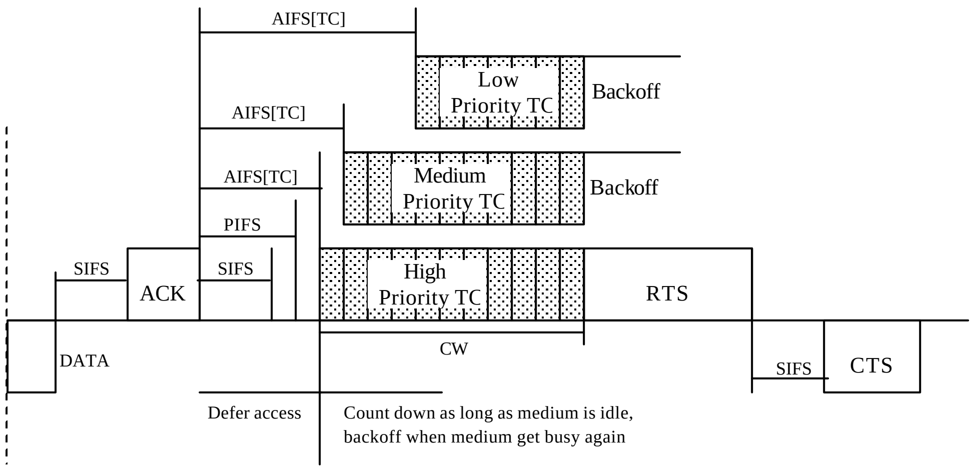


Figure 3.14 Backoff procedure of TCs of different priority

The IEEE 802.11e EDCF supports QoS by using the parameters of the TC: *AIFS[TC]*, *CWmin[TC]*, *CWmax[TC]*, and *PF[TC]*. These parameters can be set to an appropriate value to achieve a desirable QoS for a given TC.

3.3 HIPERLAN Type 1

The *High Performance Local Area Network* Type 1 (HIPERLAN/1) standard was established by ETSI in 1996 [Wesolowski 2002]. HIPERLAN/1 network operates in the 5.15-5.3 GHz band, divided into five frequency channels. The maximum data rate is about 23.5 Mbps. The reference model for HIPERLAN/1 networks is shown in Figure 3.15.

The MAC sublayer of HIPERLAN/1 performs several functions which help to organize the operation of the network. The channel access control (CAC) sublayer contains a protocol which determines which nodes are allowed to transmit and specifies the access priorities.

Medium Access Control (MAC) sublayer	Data link layer
Channel Access Control (CAC) sublayer	
Physical (PHY) layer	Physical layer

Figure 3.15 Reference Model of HIPERLAN/1

The functions of the MAC layer include:

- MAC address mapping. This allows differentiation of the terminal's own HIPERLAN network in the area where more networks can operate.
- The MAC sublayer also ensures communication security by defining the encryption-decryption algorithm.
- Addressing of MAC service access points (MSAP). The MSAPs are addressed using a 48-bit LAN-MAC address which is compatible with ISO MAC service definition.
- Manage data forwarding. Some HIPERLAN/1 terminals can operate as a relay for the packets sent between terminals which are beyond their common range.

CAC sublayer

The CAC sublayer uses a protocol called *Elimination Yield Non Pre-emptive Priority Multiple Access* (EY-NPMA). EY-NPMA provides hierarchical independence of performance by means of channel access priority such that the performance of data transmission attempts with a given channel access priority is not degraded by those with lower channel access priorities. In the EY-NPMA protocol the process of channel access is divided into three phases: Prioritization phase, Contention phase and Transmission phase, and time is divided into channel access cycles as shown in Figure 3.16.

Each channel access cycle is initiated with channel access synchronization. Immediately after channel access synchronization, the prioritization phase starts. The time of this phase is divided into five slots of equal length. The terminal with priority *p,* senses the channel for *p-1* slots. If in these *p-1* slots the channel stays idle the terminal sends an access pattern, else it does not send its access pattern and waits for the beginning of the next access cycle.

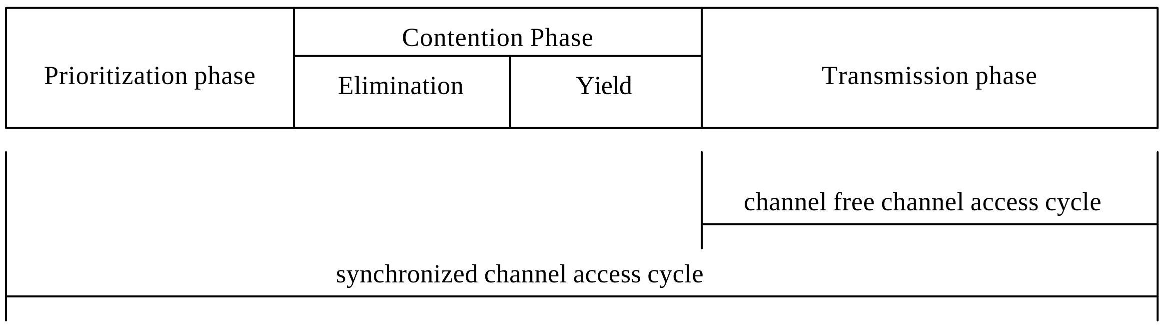


Figure 3.16 The EY-NPMA protocol activities

The next phase of the protocol is the contention phase. This phase is further divided into two phases called Elimination and Yield phases. The time duration of the elimination phase is divided into 0 to 12 slots. Each station which has not been eliminated in the prioritization phase sends an elimination burst of random duration between 0 to 12 slots. After sending its elimination burst, each station senses the channel during a long interval called the *elimination survival verification* (ESV). Those stations which send the longest elimination burst succeed in this phase.

After ESV, the yield phase starts. Again the time duration of the yield phase is divided into 0 to 9 slots. The stations which proceed to the yield phase listen to the channel for a random number of time between 0 to 9 slots. If the station has not detected any activity in the channel during the listening, it immediately starts to transmit its burst, so the transmission phase begins. If the station has detected the signal of another station during channel sensing, it is eliminated from the channel access competition and has to try again in the next access cycle.

Although this is a complex competition for the channel access, collision is still not completely eliminated in the HIPERLAN/1 protocol. A collision can occur if two or more stations happen to start transmitting in the same slot in the transmission phase. If a collision occurs, the collided stations have to try again to contend for the channel in the next access cycle.

3.4 HIPERLAN Type 2

The *High Performance Local Area Network* Type 2 (HIPERLAN/2) has been developed to support higher data rates to WLAN users. The HIPERLAN/2 was also developed partly due to the fact that new bands in the 5 GHz range were assigned to WLAN applications.

HIPERLAN/2 is designed to work in two configurations. The business environment and the home environment. The business environment configuration is an access network which consists typically of several access points connected by a core network. The home environment configuration actually creates an ad hoc network. This type of network can consist of a few subnetworks equivalent to cells in a cellular access system. Each subnetwork operates at a different frequency and has a central controller which is dynamically selected from HIPERLAN/2 terminals operating in this subnetwork [Wesolowski 2002].

DLC Layer

The *Data Link Control* (DLC) layer of the HIPERLAN/2 is situated on top of the physical layer as given in the OSI Reference Model. The DLC layer is further divided into two layers: *Medium Access Control* (MAC) and *Radio Link Control* (RLC).

The MAC sublayer of the HIPERLAN/2 is based on the TDMA/TDD principle. The basic structure of the air interface generated by the MAC is shown in Figure 3.17. It consists of a sequence of MAC frames of equal length, of 2 ms duration. Each MAC frame consists of several phases. Their names and tasks are as follows [BRAN 2000]:

- Broadcast (BC) phase: The BC phase carries the BCCH (broadcast control channel) and the FCCH (frame control channel). The BCCH contains general announcements and some status bits announcing the appearance of more detailed broadcast information in the downlink phase (DL). The FCCH carries the information about the structure of the ongoing frame, containing the exact position of all following emissions, their usage and content type. The messages in the FCCH are called resource grants (RG).
- Downlink (DL) phase: The DL phase carries user specific control information and user data, transmitted from AP or central controller (CC) to the mobile terminal. Additionally,

the DL phase may contain further broadcast information which does not fit in the fixed BCCH field.

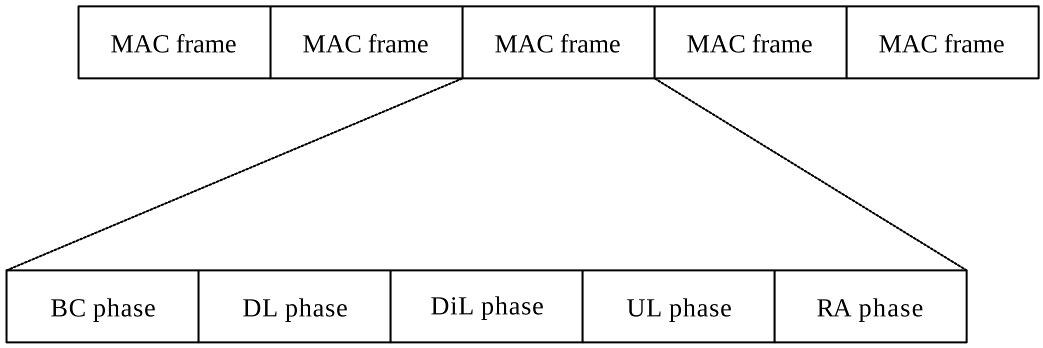


Figure 3.17 HIPERLAN/2 MAC frame format

- Uplink (UL) phase: The UL phase carries control and user data from the mobile terminals to the AP or CC.
- Direct Link (DiL) phase: The DiL phase carries user data traffic between mobile terminals without direct involvement of the AP or CC.
- Random access (RA) phase: The RA phase carriers a number of RCH (random access channels). Mobile terminals to which no capacity has been allocated in the UL phase use this phase for the transmission of control information.

The functions of the RLC include:

- Association Control Function: association of a MAC ID to a terminal and negotiation of link capabilities, encryption key exchange and refresh, authentication, and beacon signaling in the AP or CH and disassociation.
- Radio Resource Control: dynamic frequency selection, measurements performed by mobile terminals, reporting the measurements to the AP, frequency change performed by the AP and its associated mobile terminals.

The DLC layer also performs error control and also deals with the connection setup, release and modification.

3.5 Summary

One of the drawbacks of the current WLAN standards is that these standards have problems with providing QoS required for emerging applications such as voice and video. IEEE 802.11 Task Group E currently works on enhancements to IEEE 802.11 MAC so that it can support QoS by means of service differentiation. As supporting QoS in current WLAN standards is a problem, the main aim of this thesis is to design a new distributed medium access control (MAC) scheme for WLANs to support QoS. The proposed MAC scheme is given in chapter 4. In chapter 6, the proposed MAC scheme is analyzed by means simulation and compared with the most widely used WLAN standard, IEEE 802.11. The proposed MAC scheme is also compared there with the upcoming IEEE 802.11e standard.

The proposed MAC scheme is only compared with IEEE 802.11 standard because IEEE 802.11 WLAN is the most widely used WLAN standard compared to HIPERLAN. Furthermore, the proposed MAC scheme uses random channel accessing method like IEEE 802.11 but HIPERLAN uses fixed channel assignment method (TDMA) for channel access.

4.0 Proposed MAC scheme for service differentiation

This chapter is one of the main chapters of the thesis. The chapter proposes a fully distributed MAC scheme that supports service differentiation in a wireless LAN environment. In the scheme, stations use CSMA for channel access, with collisions between stations being resolved by sending a set of beacons, i.e. pulses of energy of specified duration [Krishnakumar and Sobrinho 1996], in a predefined manner. *Virtual collisions*, i.e. clashes between data packets of different priority within a station, are resolved by a scheduler inside the station. This chapter also presents other relevant information required for the proposed MAC scheme.

4.1 Objectives of the proposed MAC scheme

IEEE 802.11 is the most widely used WLAN standard. Its distributed access method, Distributed Coordination Function (DCF), is used normally for non-real-time services, while its centralized access mode, Point Coordination Function (PCF), is meant for real-time services to provide services of higher priority than non-real-time services. Although the PCF access method supports limited QoS to real-time services, it is a centralized scheme and so requires the existence of access points with specialized functions. A drawback of such a central scheme is that the performance of the station depends on the access point. If the access point fails to function then the PCF access method will not work, and QoS will not be supported. Moreover as concluded in [Visser and Zarki 1995], the centralized access mode performs poorly compared to the distributed access method.

The main objective of this work is to design a fully distributed MAC scheme for the WLAN environment. Furthermore, the proposed MAC scheme has to:

- support QoS over Wireless LANs, and
- achieve better performance than existing Wireless LAN standards

The sections below present a detailed description of the proposed MAC scheme, and the methods used to satisfy high QoS requirements in a WLAN.

4.2 The Proposed MAC Procedure

The MAC scheme allows each station to determine the total number of stations in the same Basic Service Set (BSS) [IEEE 802.11 1999]. The station needs also to know the Identification (ID) numbers of the other stations in the BSS. The ID number is assigned to a station when it joins the BSS. The method used to achieve this is explained in section 4.2.2. For the time being, it is assumed that each station is given an ID number according to the number of stations in the BSS. This is an important property in the proposed MAC scheme because this information is used in collision resolution.

In this MAC scheme, QoS is supported by assigning different priorities to different types of traffic. In each station within a given BSS, traffic can be prioritized to a number of levels, depending on the number of required priority levels. If N different priority levels are required in the network, then each station has N different queues to hold data traffic as seen in Figure 4.1. We assume that priority is given in descending order so that data of priority 1 (P1) has the highest priority and data of priority N (PN) has the lowest priority.

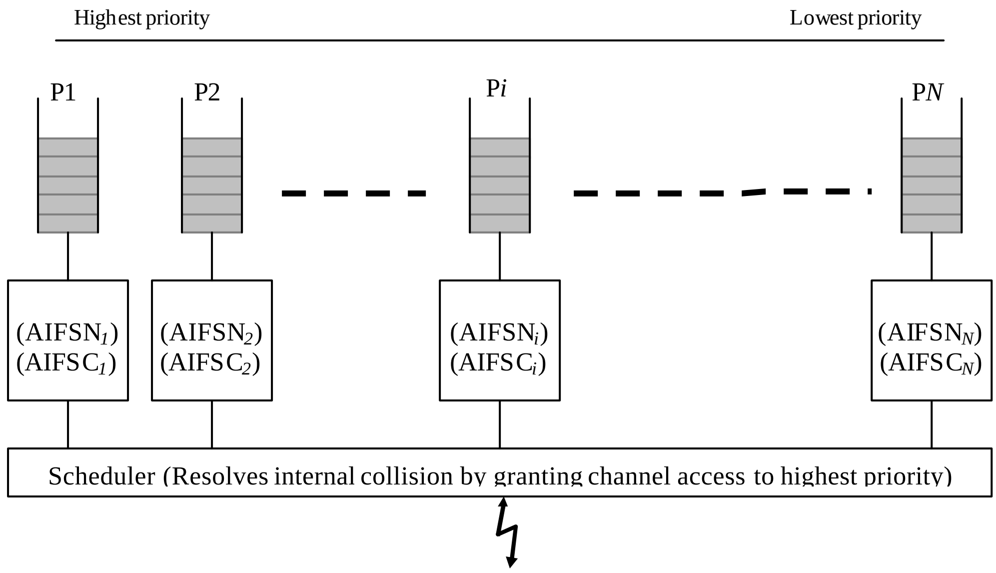


Figure 4.1 Different priority traffic of a station with a scheduler to resolve internal collisions

The channel access prioritization is achieved by assigning a different channel access time, called Interframe Space (IFS), for each different priority class. If a station has data of P*i* (1 = *i* = *N*), the new data of P*i* has to wait until the channel is idle for a time period of $AIFSN_i$ (Arbitrary Interframe Space for New data of priority *i*) before accessing the channel. A collision in the channel can occur if two or more stations start to transmit simultaneously. A station which has a collided data of P*i* has to wait until the channel is idle for a time period of $AIFSC_i$ (Arbitrary Interframe Space for Collided data of priority *i*) before accessing the channel for collision resolution. This is shown in Figure 4.2. The time period $AIFSN_i$ has to be greater than $AIFSC_i$ so that collided data of P*i* has a higher chance to access the channel than new data of P*i*, and the time periods $AIFSN_i$ and $AIFSC_i$ has to be greater than $AIFSN_{i-1}$ so that new data of P_{i-1} has a higher chance to access the channel than new and collided data of P_i. Hence the timings have to satisfy the inequality given below:

$$AIFSC_{i-1} < AIFSN_{i-1} < AIFSC_i < AIFSN_i$$

These IFSs are used to achieve service differentiation among different priority data traffic. The *virtual collision*, i.e. clashes between data packets of different priority within a station, is resolved by a scheduler inside the station. If a *virtual collision* has occurred in a station, the scheduler permits the highest priority data packet to access the channel, first.

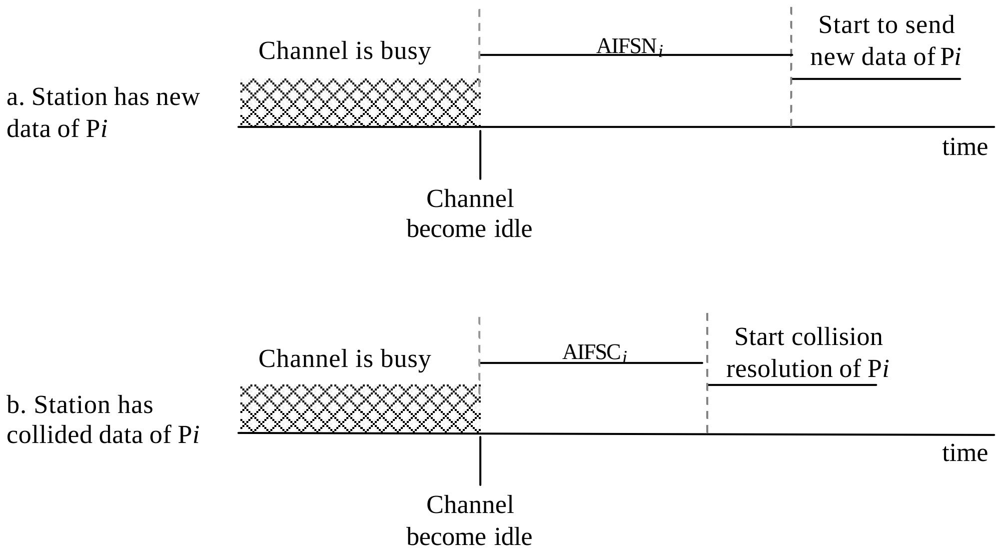


Figure 4.2 Channel accessing procedure for new and collided data of P*i*

As in IEEE 802.11, the state of the medium is determined by carrier-sensing functions. Carrier sensing is done at both the air interface using *physical carrier sensing* and at the MAC sublayer using *virtual carrier sensing*. A physical-carrier sense mechanism is performed by the physical layer and this information is passed to the MAC sublayer. In physical-carrier sensing, stations detect the presence of carrier wave in the medium by relative strength of signals from other sources. The MAC provides a virtual carrier-sense mechanism. This mechanism is based on the network allocation vector (NAV). The NAV indicates the amount of time that must elapse until the current transmission session is complete and the channel can be sampled again for an idle channel. The NAV predicts the length of this time period using the information that is contained in Request to Send (RTS) or Clear to Send (CTS) frames prior to the actual exchange of data. This information is also available in the MAC headers of almost all frames. A given station updates the NAV with the information contained in the Duration/ID field, obtained when the station receives a valid fame. But the station updates the NAV only if the new NAV value is greater than the current NAV value and only if the frame is not addressed to the station.

As stated before, if a station wants to send data of P*i* then the station has to wait until the channel is idle for a time period of $AIFSN_i$. If the station senses that the channel is idle for $AIFSN_i$ time period then it can send data after RTS/CTS exchange, as in [IEEE 802.11 1999]. This is shown in Figure 4.3. In an RTS/CTS exchange, when the source station finds the channel is idle for a time period of $AIFSN_i$, it sends a RTS frame to the destination station. After receiving the RTS frame, the destination station waits for a short IFS (SIFS) time and sends back a CTS frame to the source station. If the source station received the CTS frame correctly then it means the channel is free to send the data frame. So the source station sends a data frame after SIFS channel free time and waits for ACK frame from the destination station. When the ACK frame is received correctly, the data transmission is successfully completed. The RTS, CTS and ACK frames contents are identical to that of IEEE 802.11 [IEEE 802.11 1999].

However, if CTS is not received, the source station will assume that a collision has occurred and collision resolution has to be performed. But if CTS is received but not ACK, then the station has to try to resend the data after again an $AIFSN_i$ channel idle time.

When a collision of data of P*i* occurs and the station senses an idle channel for a time period of $AIFSC_i$, it sends a collision resolution beacon (CRB) to occupy the channel for the collision resolution. The beacon is a pulse of energy used to inform any listening stations that the channel is busy. The duration of the CRB has to be more than an RTS frame duration so that new data of any priority does not occupy the channel when a collision resolution phase has started.

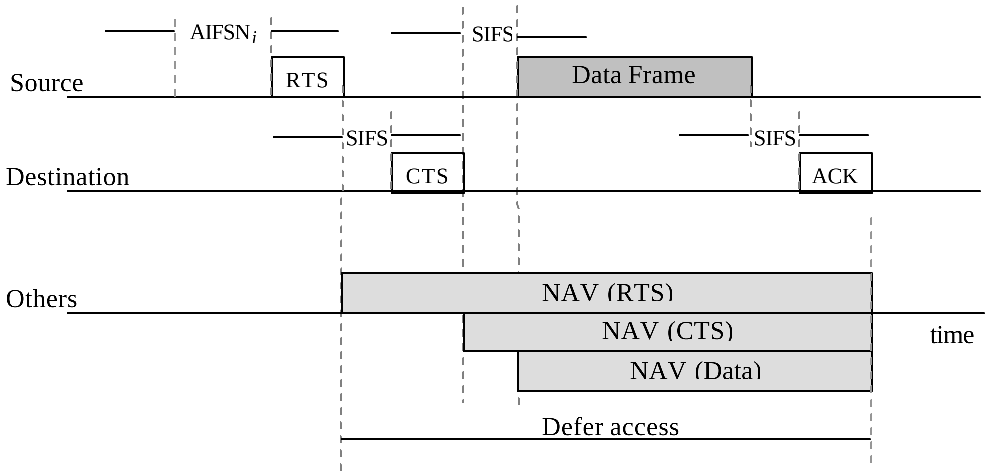


Figure 4.3 Data transmission using RTS/CTS

In the new data transmission, the RTS/CTS exchange method is also used to minimize the hidden and exposed station problems. In the collision resolution phase, when collided data are re-sent, use of RTS/CTS is optional. If the RTS/CTS method is used in collision resolution, it will obviate hidden and exposed station problems as explained in section 2.5. In addition, collision resolution uses Token Pass (TP) and Token Receive (TR) frames, which also help to reduce those problems. The TP and TR frames are just like the ACK and CTS frames. They also contain a duration field to inform other stations how long the channel will be busy by a given data transmission. Figure 4.4 shows data transmission in the collision resolution phase using RTS/CTS. In collision resolution, a collided station has to wait to access the channel until the channel is idle for a time period of SDIFS (Scheduled Data IFS), from the time it receives the TP frame. The TP and TR frames sending procedure, in a collision resolution is explained in section 4.2.1. In general, the use of the TP and TR frames is to give the channel access opportunity to the next known station, which has collided data frame to send. As shown in the figure, all other stations not involved in data transmission set their NAV to the appropriate value transmitted with the frames; RTS, Data, CTS, ACK, and TP. The RTS, CTS, ACK, and TP frames contain information about how long the channel will be busy in a given data transmission from the reception of the specific frames. But Data frame contains information about how long the channel will be busy in a given data transmission from the beginning of the Data frame transmission.

On the other hand, when sending small data packets or when the channel is not so busy then the stations could choose to send data without RTS/CTS method. Figure 4.5 shows collided data transmission without using RTS/CTS. This method of data transmission does not reduce the hidden and exposed station problems as the RTS/CTS access method.

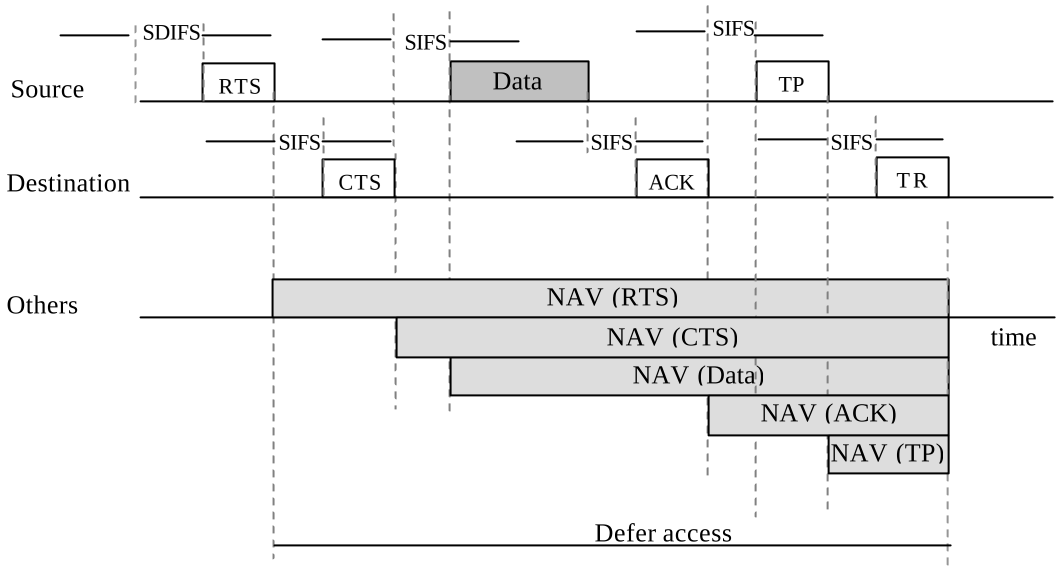


Figure 4.4 Data transmission using RTS/CTS, in collision resolution phase

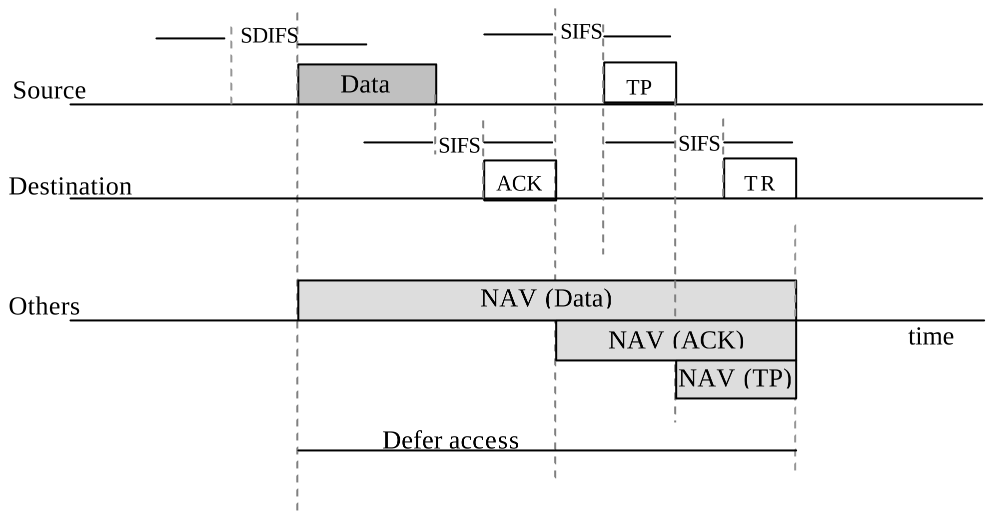


Figure 4.5 Data transmission without using RTS/CTS, in collision resolution phase

4.2.1 Collision Resolution

Each station in the BSS keeps information about every other station in the BSS in the form of a table shown in Figure 4.6. The figure shows that the BSS has M stations. The table contains stations ID numbers and corresponding address of the stations, and also a field called scheduled data (SD). The SD field of the table carries information about which stations have collided data (scheduled data) to send in a given collision resolution period. All the stations involved in the collision resolution use SD field of the table to keep information about the stations which has scheduled data to send. In a *scheduled data transmission phase* (explained further below), if k^{th} ($1 = k = M$) position of the SD field is set to 1, it means that the station with ID number k has scheduled data to send otherwise the station does not have a scheduled data. In addition to these information, each station also keeps a counter called Collision Resolution (CR) counter. The CR counter of a station tells in which position that the station has to send its' scheduled data in a given collision resolution. The CR counter is set to the appropriate value during *collision resolving beacons phase* (explained later) of a given collision resolution.

STA ID	1	2	3	4	- - - - - - - - - -	*M*
STA Address	XXXXX	XXXXX	XXXXX	XXXXX	- - - - - - - - - -	XXXXX
Scheduled Data	0 or 1	0 or 1	0 or 1	0 or 1	- - - - - - - - - -	0 or 1

Figure 4.6 Stations information table

A collision resolution period is divided into two phases as shown in Figure 4.7. If a collision between data of P*i* has occurred, the collided stations start a collision resolution period after the channel is idle for a time period of $AIFSC_i$. The first phase is called *collision resolving beacons phase*. In this phase, a collided station sends a set of beacons to inform other stations that the station has collided data to send. The second phase is called *scheduled data transmission phase*. In this phase, collided stations send collided data according their ID numbers.

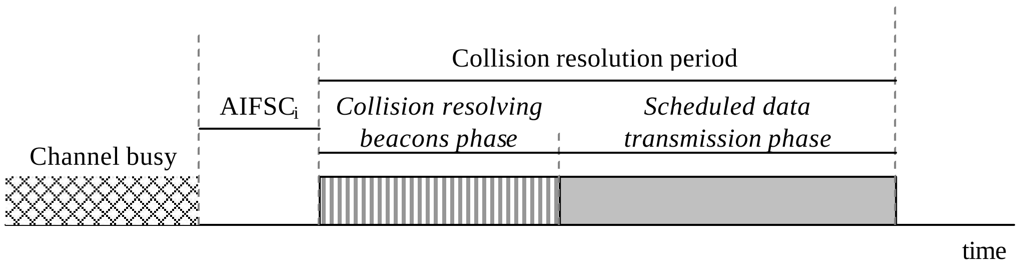


Figure 4.7 Collision resolution period for collided data of P*i*

If a station with ID number l ($1 = l = M$) has a collided data of P*i*, the station first senses the channel. If the channel stays idle for a time period of $AIFSC_i$, the station starts a collision resolution period by sending a CRB beacon. In the first phase, *collision resolving beacons phase*, the station sends sets of beacons. These beacons are send in specific way so that after *collision resolving beacons phase* collided stations will know how many collided stations are there in the channel, and all the collided stations will know in which position that the stations have to send the collided data. The procedure of this phase is shown in Figure 4.8.

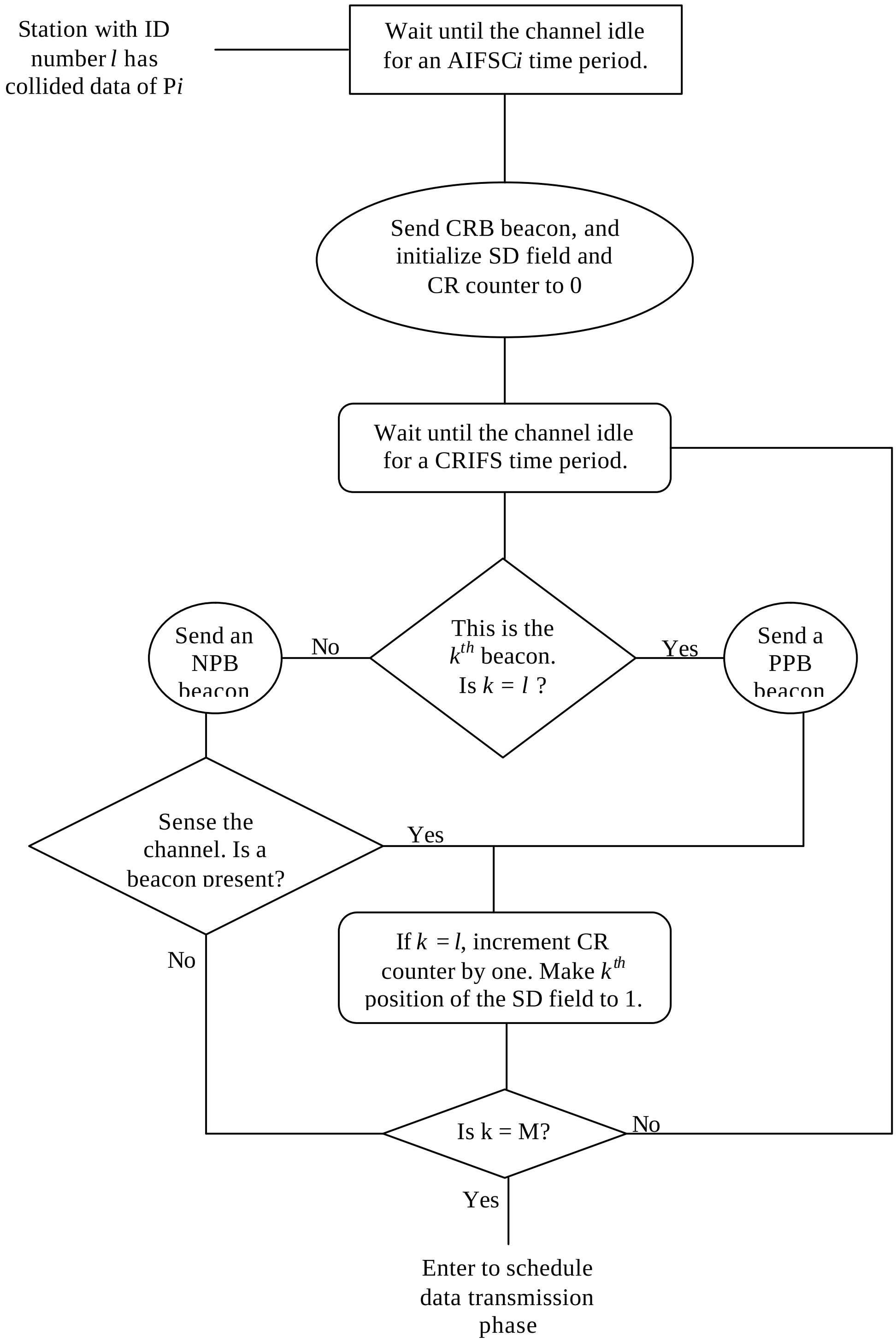


Figure 4.8 Procedure of the *collision resolving beacons phase* of the collision resolution period

After CRB, the station with ID number *l*, initializes the SD field and CR counter of the station to 0. Then the station waits until the channel is idle for a time period of CRIFS (Collision

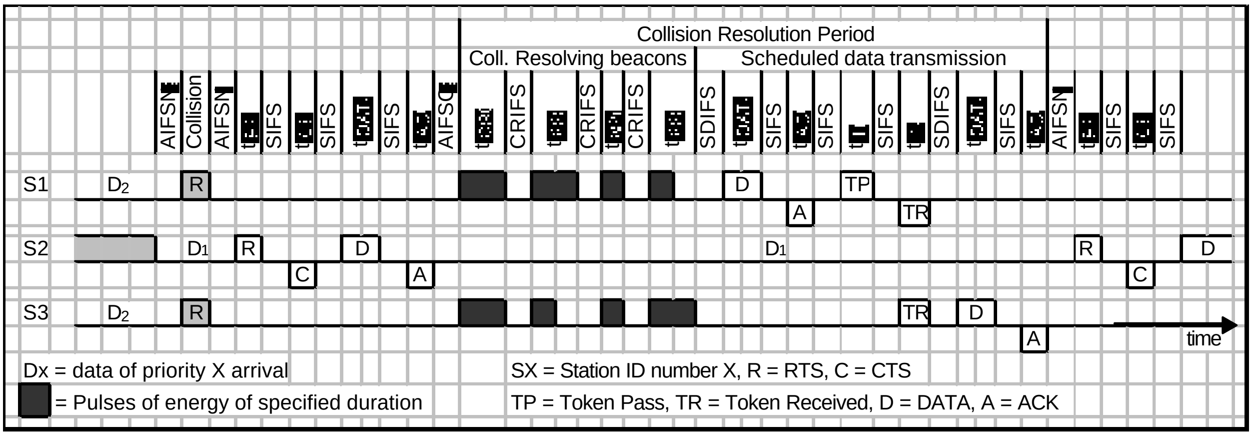
Resolution IFS) and sends a NPB (No Packet Beacon) of length t_{NPB}. CRIFS is the maximum time needed by a station to switch from transmitting to receiving and back again. Likewise, the station sends *l-1* NPBs. In between each of these NPBs, the station waits until the channel is idle for a time period of CRIFS. After the k^{th} (1 = *k*= *l-1*) NPB, the station senses the channel for presence of a beacon. If a beacon is present, this means the station with ID number *k* has sent a PPB (packet present beacon) announcing it has collided data of P*i* to send. Using this information, the station updates the k^{th} position of the SD field to 1, and increments the CR counter by one.

Then after $l\text{-}1^{th}$ NBP, the station again waits until the channel is idle for a time period of CRIFS and sends a PPB of length t_{PPB}, announcing that the station has collided data of P*i* to send. This time the station does not sense the channel for presence of a beacon because it sent the PPB, and the station updates the l^{th} position of the SD field to 1, and increments the CR counter by one. The CR counter now has the value equivalent to the number of stations which had sent a PPB.

After sending the PPB, the station again sends *M-l* NPBs using a similar procedure by waiting until the channel is idle for a time period of CRIFS in between the NPBs. Even in this time, after each NPB the station senses the channel for presence of a beacon. If a beacon is present the station updates the corresponding position of the SD field to 1 but does not change the CR counter.

When all the M beacons (M-1 NPB and 1 PPB) are sent, the station then enters the second phase of the collision resolution period, *scheduled data transmission phase*. Now the station has updated SD field telling which stations have scheduled data to send, and CR counter telling in which position the station has to send the collided data.

An example to show these steps is given in Figure 4.9. As shown in the figure, the BSS has 3 stations. Initially, the station with ID number 1 (S1) and S3 have collided with data of P2. Hence, these stations enter a collision resolution period after waiting until the channel is idle for a time period of $AIFSC_2$. The stations start a collision resolution period by sending a CRB. After CRB, the S1 sends a PPB to announce that it has collided data of P2 to send and likewise the S3 sends a PPB in 3rd position. At the end of the collision resolving beacons phase, the S1 has CR counter value of 1, the S3 has CR counter value of 2, and stations information table for the S1 and S2 are also shown in Figure 4.9.



STA ID	1	2	3
STA Address	XXXXX	XXXXX	XXXXX
Scheduled Data	1	0	1

Stations information table of S1 and S2 after collision resolving beacons phase.

Figure 4.9 Collision Resolution for a BSS that has 3 stations

At the end of the *collision resolving beacons phase*, a token is generated at the station that has CR counter value 1, giving it the right to transmit its collided data. The last station which carries the token in the scheduled data transmission will destroy the token generated in this collision resolution period by not sending the TP message to any other station. A station knows whether it is the last station in the collision resolution by using the SD field. The highest position of the SD field which is set to 1 corresponds to the ID number of the last station in the collision resolution. In each collision resolution period, a new token is generated by the station which has CR counter value 1.

In *scheduled data transmission phase*, the CR counters of the collided stations are decreased by one for each channel idle time of SDIFS (Scheduled Data IFS). If a station does not receive a TP message when its counter reaches 0, it assumes that the TP message is lost and will try to send the data again after the channel is idle for a time period of AIFSN$_i$. CR counter is used to tell a station whether the scheduled data transmission phase is going on or not, and whether the station still has the chance to transmit its' collided data in the collision resolution period.

If the station with ID number *l* has CR counter value 1 after *collision resolving beacons phase*, then the station has the chance to send its' collided data. In this case, the station first waits until the channel is idle for a time period of SDIFS (Scheduled Data IFS). SDIFS has to be greater than SIFS but less than AIFSC$_1$ so that newly arriving data will not collide with scheduled data. After waiting a time period of SDIFS, the station sends its collided data with or without RTS/CTS as explained in section 4.1. When the collided data has been sent, the station sends a TP message to a collided station which has the next lowest ID number, from which it must receive a TR message. The station finds which station to send the TP message by using the SD field of the stations information table. If a TR is not received and a timeout (specified channel idle time) occurs then the station sends a TP to the next STA in line. The station tries to send the TP message like this until it is successful or until there is no more collided station to send. The timeout duration has to be greater than SIFS but less than SDIFS so that the CR counters of the station will not decrement. This is illustrated in Figure 4.10. In the figure, S1 did not get a TR message from S2 and Timeout has occurred. Hence, S1 sends a TP message to the next station in line, S3 and S2 tried to send their collided data of P2 after waiting until the channel is idle for a time period of AIFSN$_2$.

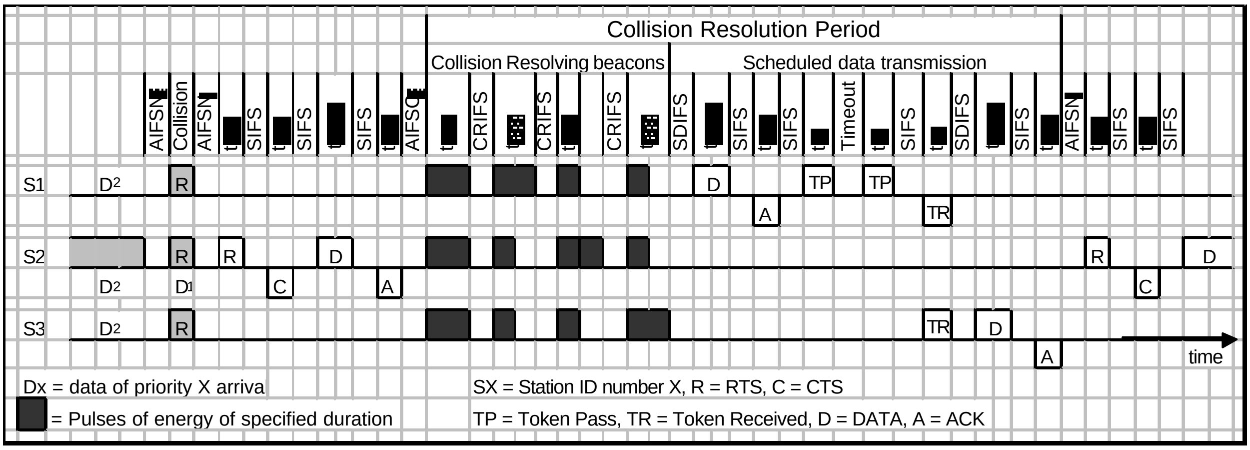


Figure 4.10 TP message passing methods in scheduled data transmission

On the other hand, after *collision resolving beacons phase*, if the station with ID number *l* has CR counter value greater than 1 then the station has to wait to send the collided data until it receives a TP message. During this time, the station decrement its' CR counter by one for each channel idle time of SDIFS. If a TP message is received before the CR counter reaches to zero, the station sends its' collided data as explained in section 4.1. Then the station tries to send a TP massage to another collided station in line as explained above, if it is not the last station involved in the collision resolution. Else if TP message is not received before the CR counter reaches to zero, the station assumes that the chance to send the collided data is lost in the collision resolution and will try to send the data again after the channel is idle for a time period of AIFSN$_i$.

TP and TR are important frames in the scheduled data transmission phase. These frames are used in passing the token among the collided stations. Upon reception of the token by a station, chance for transmission is granted in the scheduled data transmission phase. However, if the token is not received before the CR counter of the station goes to zero, the chance to transmit in the scheduled data transmission phase is lost.

Furthermore, if timeout has not occurred but a TR is not received correctly then the TP has to be sent to the next station in line as before. After reception of the TP, the receiving station has to send a TR to the sender of the TP. After sending the TR, if the next channel idle time is less than SDIFS then the station has to assume that the token has been lost and has to try gain to send the data after $AIFSN_i$. If this happens that means, the previous station that held the token had sent the token to another station. This might happen, if TR or TP message has an error in it due to channel problems like fading, or if a station could not respond to a TP message due to a hidden or an exposed station.

The MAC scheme presented here is designed in such a way that higher priority data has a higher chance to access the channel by using different IFSs. If a station is ready to transmit a higher priority data then a newly arrived or collided lower priority data will not be able to access the channel before the higher priority data. Hence service differentiation is guaranteed in the proposed MAC scheme.

4.2.2 New Stations and ID numbers

New stations use a similar method to the IEEE 802.11 DCF protocol to access the channel. This method is used to resolve collisions between two or more new stations arriving to the BSS at the same time. In this method, a new station senses the channel before initiating a packet transmission. If the channel is idle for a time period greater than an Authentication Initiation Inter Frame Space (AIIFS), then the new station broadcasts the authentication initiation frame (AIF). The AIF contains information about the new station such as address, and capability information. The AIIFS must be greater than the $AIFSN_N$ (the longest AIFSN) so that the new stations will not be added to the BSS if the BSS is busy with existing stations. This will prevent overloading of the BSS.

Otherwise, if the new station senses that the channel is busy then the transmission is deferred and a backoff process is entered. Specifically, the station computes a random value in the range of 0 to the so-called Contention Window (CW). A backoff time interval is computed using this random value: $T_{backoff}$ = Rand(0,CW).T_{slot}, where T_{slot} is the slot_time [Visser and Zarki 1995]. This backoff interval is then used to initialize the backoff timer. This timer is decremented only when the channel is idle. The timer is frozen when the channel is busy. Each time the channel becomes idle for a time period longer than AIIFS, the backoff timer is periodically decremented once every slot time.

As soon as the backoff timer expires, the new station accesses the channel. A collision would occur when two or more new stations start transmission simultaneously in the same slot. An Authentication Confirmation Frame (ACF) is used to notify the new station that the frame it sent has been received. If an ACF is not received before the channel is again idle for an AIIFS time, the new station assumes that the AIF is lost and reenters the backoff process. To reduce the probability of collisions, after each unsuccessful transmission attempt, the CW is doubled until a predefined maximum (CW_{max}) is reached as in IEEE 802.11 DCF protocol. A new station adding procedure is shown in Figure 4.11.

Stations in a given BSS must respond when they receive an authentication initiation frame from a new station. If the authentication is successful, an authentication confirmation frame must be sent to the new station after channel is idle for a time period of short inter frame space (SIFS) channel idle time but if the authentication is unsuccessful, an authentication rejection frame must be sent to the new station in a similar manner. Each of these frames must be acknowledged by the new station.

If an acknowledgement frame is not received from the new station and timeout has occurred then the stations in the BSS assumes that a collision has occurred. This collision is resolved by using the same procedure used in the proposed protocol. Before entering the collision resolution the stations must sense the channel. If the channel is idle for a time period of $AIFSC_A$ (Arbitrary Inter Frame Space for Authentication Collision Resolution), the stations send a CRB (Collision Resolution Beacon). Then the collision is resolved by sending beacons as in the protocol. $AIFSC_A$ must be less than AIIFS so that newly arrived stations do not have a chance to access the channel before the new station is added to the BSS. And $AIFSC_A$ must be greater than the $AIFSN_P$ so that new stations do not interfere with transmissions of already existing stations in the BSS.

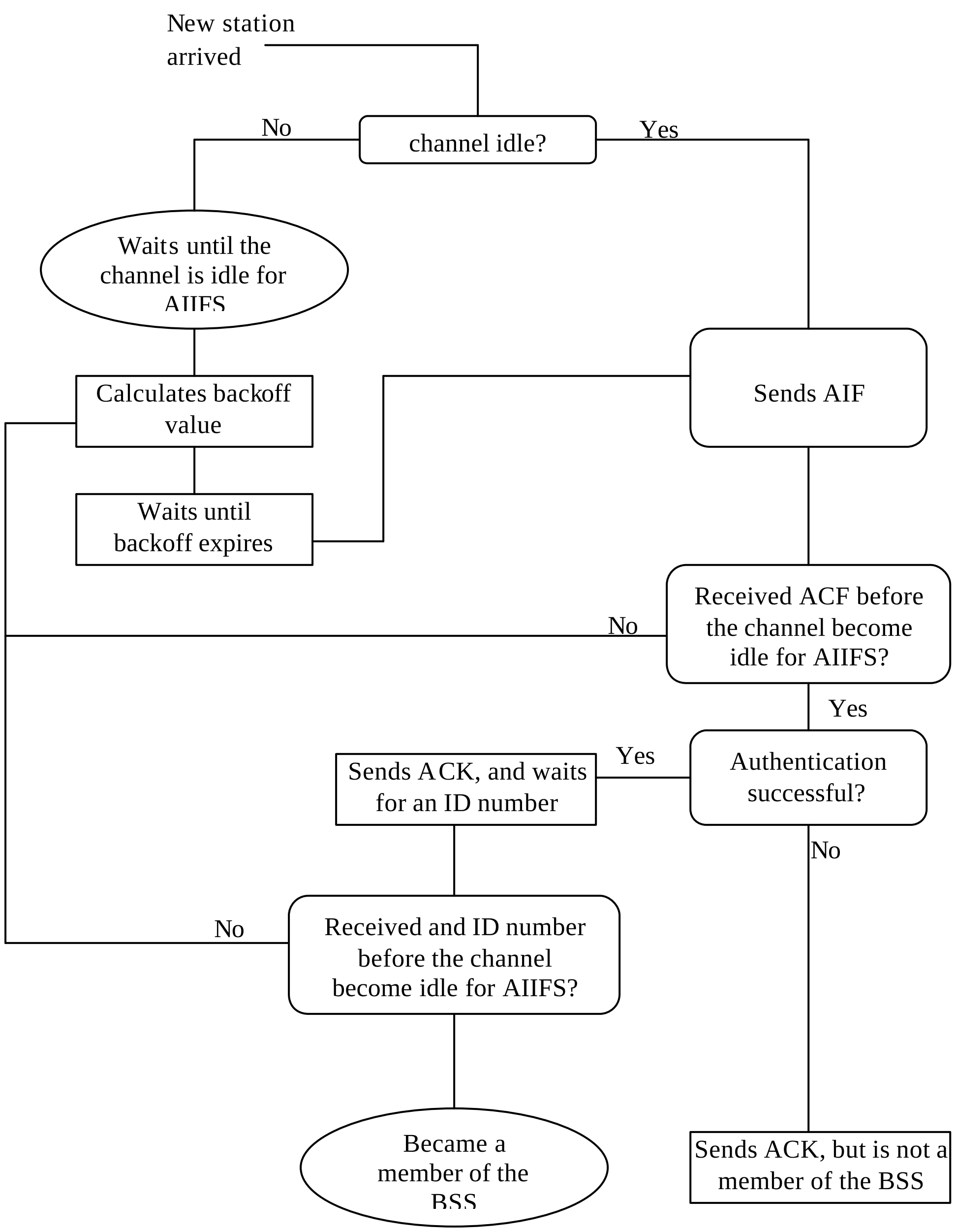


Figure 4.11 Procedure to add a station to a given BSS

The station that has the token after collision resolution will send an authentication response frame to the new station. The token is passed to the next station in line only if an acknowledgement frame is not received from the new station. A stations in the BSS sends an ACF to the new station to inform whether the authentication is unsuccessful (if the new station does not have authority to add into the BSS) or successful. For each case, the new station has to respond to the ACF by sending an ACK. If the authentication is unsuccessful that means the new station cannot be a member of the given BSS, due security issues. However, if the authentication is successful, after the acknowledgement frame the new station waits for a station ID number until the channel is idle for AIIFS. If within this time the new station does not receive a station ID number then the new station has to restart the authentication process again.

After successful authentication, before sending a station ID number to the new station, the station in the BSS has to inform all other stations in the BSS that a new station has been added to the BSS. This information includes new station ID number, new station address, and other relevant information about the new station. This is done immediately after the reception of the acknowledgement frame from the new station. After informing all the stations, the station then sends a frame containing station ID number to the new station. The frame also contains information about all the other stations in the BSS. From this time, the new station is a member of the BSS. This procedure is shown in Figure 4.12.

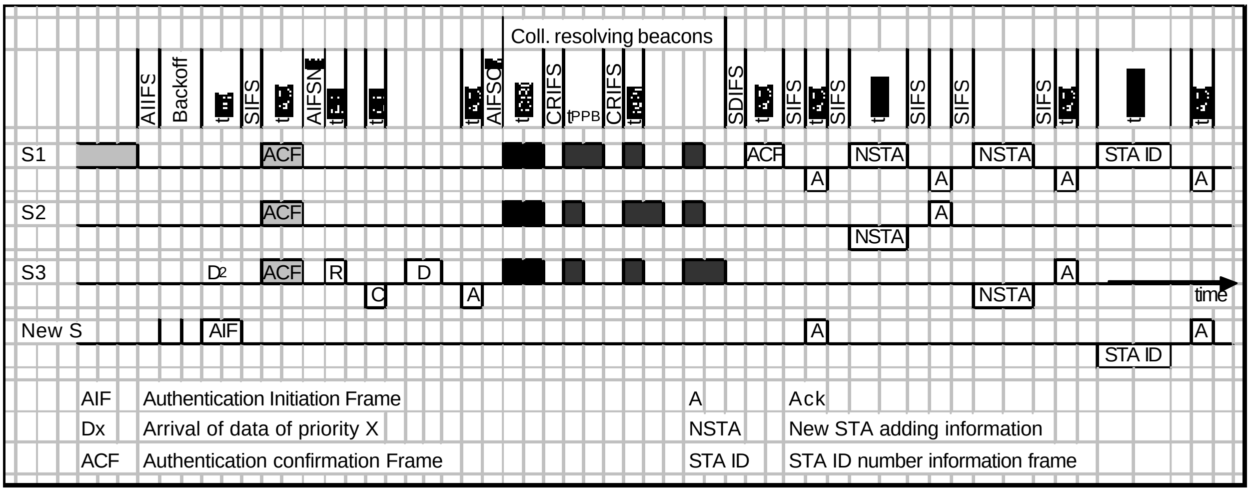


Figure 4.12 Adding a new station into the BSS

If the information about the new station could not be sent to any of the stations in the BSS, then the station sending this information has to try again. This is done by sending CRB after $AISFC_A$ channel idle time. Even after this time, if the non-responding stations did not respond then the station tries to remove the non-responding stations from the BSS. The non-responding stations are removed by sending removal information to all other stations in the BSS. This is shown in Figure 4.13. The figure shows that S2 is non-responding. S1 sent new station adding information (NSTA frame) twice to the S2 but S2 did not respond. As a result, S1 removes the S2 from the BSS by sending station removing information (SR frame) to S3. The SR frame contains the new ID number to the S3, and also information about the removed station, S2. After removing the non-responding stations from the BSS then the station sends the new station ID number to the new station as shown in the figure.

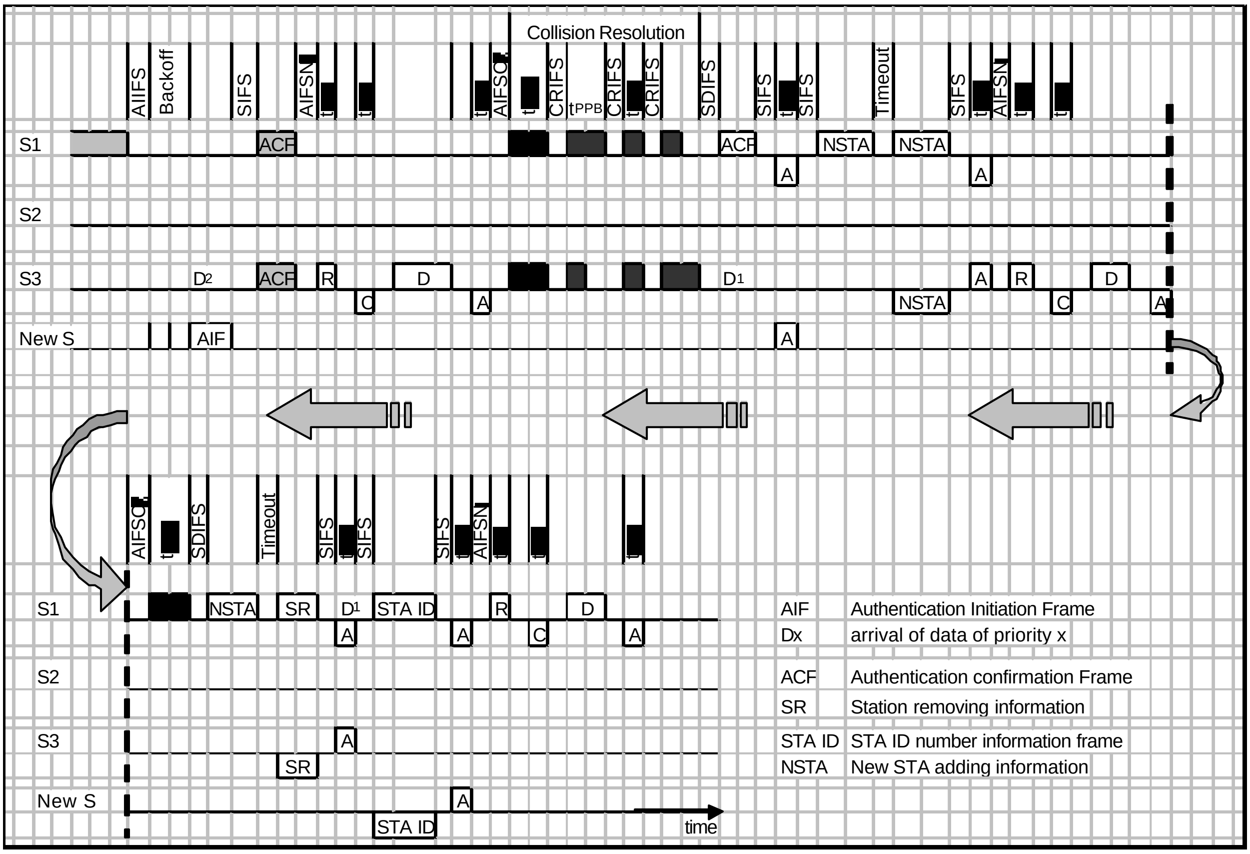


Figure 4.13 Removing non-responding stations from the BSS

Each station keeps a list of stations that are in the BSS as shown in Figure 4.6. This list includes the addresses of the stations and their corresponding station ID numbers. When a packet is received from a sender the receiving station has to check whether the address of the sender is in the list. If not, the receiver has to inform the sender that the sender is no longer a member of the BSS. If this happens the sender has to try to join the BSS by using the method given above.

In a rare case, when a removed station becomes available but it does not know that it has been removed from the BSS, then data collisions may occur between the removed station and the station that took the same position. As a result, acknowledgements and a TR may not be received in scheduled data transmission periods or the chance to transmit scheduled data in scheduled data transmission periods may be lost. This may also happen due to hidden and exposed stations in the BSS. The next section explains methods to deal with these problems.

4.2.3 Enhancements

As mentioned in the last section, acknowledgement and a TR may not be received in scheduled data transmission periods or chance to transmit data in scheduled data transmission periods may be lost because of problems with hidden or exposed stations. These problems may also arise when a station is removed but it does not know that it has been removed from a given BSS. In this case, the removed station tries to send data in the normal procedure because as far as it is concerned it is still a member of the BSS.

Preventing these problems is important as they really can affect the performance of the network. As in other WLANs, problems due to the hidden and exposed stations can be reduced by using RTS/CTS data transmission mechanism and also by ensuring the channel sensing range to be much higher than the receiving range as explained in chapter 2.

Removed station problem can be prevented by using the following procedure. A station does not take part in collision resolutions if the station does not receive an acknowledgement and a TR within a few consecutive scheduled data transmission phases. Else if a station has lost chances to transmit scheduled data during a few consecutive scheduled data transmission phase then again the station does not take part in collision resolutions. After not taking part in collision resolutions, the station has to check whether it is still a member of the BSS by using the random backoff method. The method is exactly the same as the method for adding a new station, if the station is not a member of the BSS yet. However, if the station is still a member of the BSS, when the AIF is successful, the STAID frame is used to notify it rather then the ACF. This is shown in Figure 4.14.

In the figure, S4 does not receive an acknowledgement and a TR in a few consecutive scheduled data transmission phases so the station is not sure whether it is still a member of the BSS. As a result, the station is checking whether it is a member of the BSS or not. To do that, the station sends an AIF using the random backoff procedure. As far as other stations are concerned, S4 is still a member of the BSS, so S1 responds to the AIF by sending STAID frame after collision resolving beacons. When S4 receives the STAID frame, the station knows that it is still a member of the BSS and also obtains its ID number from the STAID frame.

So by using this procedure, those stations which are unsure whether they are member of the BSS or not can check whether they are still included in the BSS. By using this enhancement, the MAC scheme will performs better.

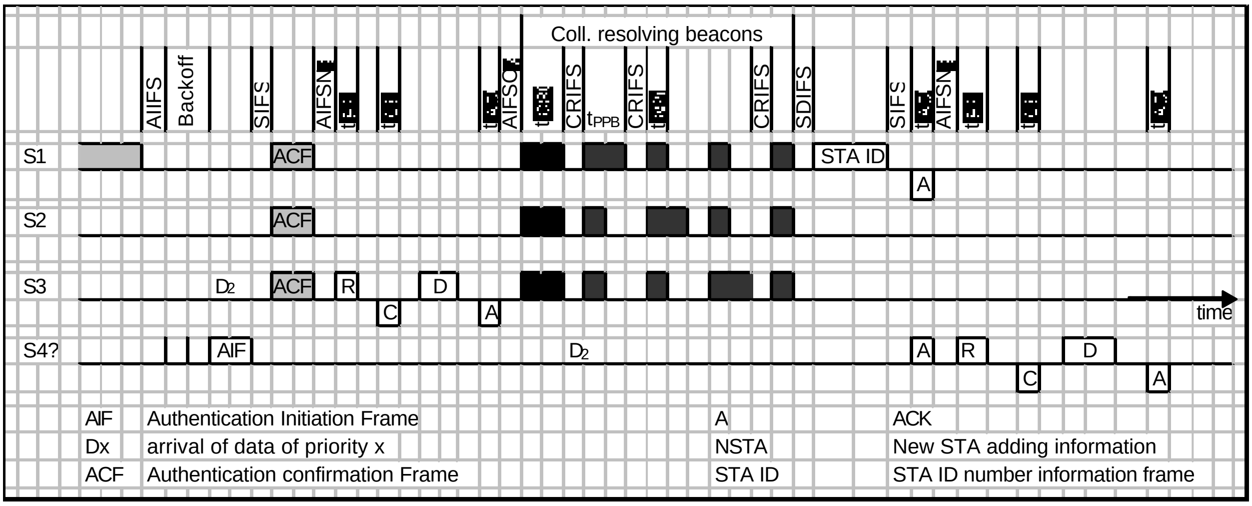


Figure 4.14 A station is checking whether it is still a member of the BSS

4.3 Summary

In this chapter a new distributed MAC scheme is proposed for WLANs to support QoS by means of service differentiation. In the scheme, stations use CSMA for channel access, where collisions between stations are resolved by sending a set of beacons. The *virtual collision* is resolved by a scheduler inside the station.

The performance of the proposed MAC scheme is evaluated analytically in chapter 5. In addition, the performance is also evaluated by using simulation in chapter 6. This chapter also includes the simulation model used and the results obtained are compared with analytical results obtained in chapter 5. Chapter 6 also presents comparison between simulation results obtained for the proposed MAC scheme with that of the IEEE 802.11 and the IEEE 802.11e protocols.

CHAPTER 5

5.0 Mathematical Model of the Proposed MAC Scheme

This chapter presents a detailed mathematical model of the proposed MAC scheme. In the model, a maximum of two priority classes are considered to simplify the model and three different possible scenarios are presented. These scenarios are

1. All the stations in the Basic Service Set (BSS) [IEEE 802.11 1999] have only priority one traffic.
2. All the stations in the BSS have both priority one and priority two traffic but the channel is not fully occupied.
3. All the stations in the BSS have both prio rity one and priority two traffic and the channel is fully occupied.

The model is derived to estimate the average delay experienced by priority one and two packets under different conditions. A few assumptions are also made to simplify the analytical model:

A1. *Average traffic arrival rate for every station is equal,* the average arrival rate of priority one traffic for each station is l_1 packets/second and the average arrival rate of priority two traffic for each station is l_2 packets/second.

A2. *Packets arrive at stations according to Poisson processes, i.e.* at each station interarrival times of priority one packets are is independent from that of another station and is exponentially distributed with parameter l_1. Similarly interarrival times of priority two packets are independent from that of another station and are exponentially distributed with parameter l_2. No assumptions are made about the service demands of the packets. Hence, traffic service times have a general distribution.

A3. No packets are lost due to transmission errors such as fading, so channels are perfect.

A4. Every station in the BSS is in the transmission and receiving range of all other stations, so there are no Hidden Terminals.

A5. *Packets of* every priority class have queues with infinite buffer space, so no packets are lost due to a full buffer.

The sections 5.1, 5.2, and 5.3 describe the three different possible scenarios that may happen in a given BSS. Section 5.1 describes a model to estimate the average delay experienced by priority one packets when the stations in the BSS have only priority one traffic. Section 5.2 and 5.3 describe models to estimate average delay of priority one and two packets when the stations have both the traffic. However, section 5.2 assumes that the channel is not fully saturated but in section 5.3 it is saturated due to priority two traffic. In addition, in section 5.4, an expression to estimate the saturation throughput of priority one traffic, is derived when the packet payload and the number of stations in a given BSS are given.

5.1 Stations have only priority one traffic

In this case, only priority one traffic is considered so that traffic arrival rates for other traffic categories are considered to be zero. The MAC scheme for this situation can be modeled as shown in Figure 5.1. Under the assumptions listed earlier, the MAC layer of each station may be represented by an M/G/1 queue [Bertsekas and Gallager 1992], [Kleinrock 1975], [Kleinrock 1976], [Kleinrock and Gail 1996]. Traffic arrives at the MAC layer of each station with an average rate of λ_1 packets per second. The traffic serving rate μ_1 is the average traffic service rate with respect to all of the station. That is, if the average time taken for a station to serve a packet is x seconds, then $m_1 = 1/x$.

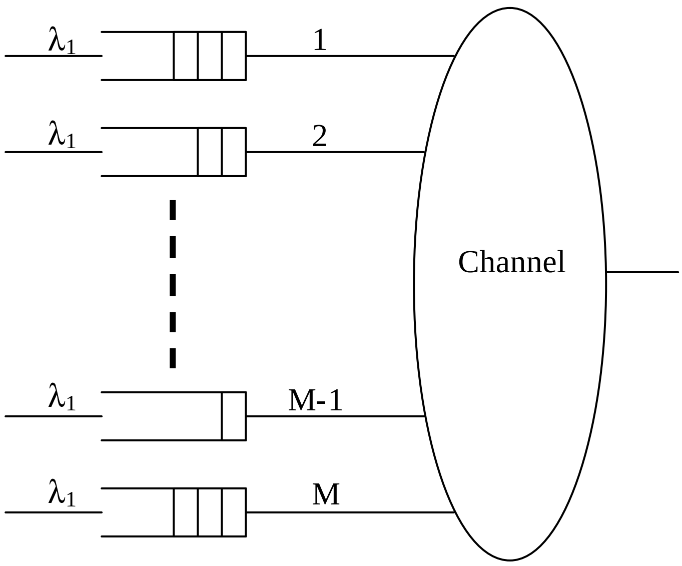


Figure 5.1 The MAC Model for the situation when all the stations in the BSS have only Priority one traffic

When a packet arrives at the MAC layer of a station, three possible situations may arrise:

a. The station may not be serving a packet and the channel is free
b. The station may not be serving a packet but the channel is busy
c. The station may be busy serving a packet

As each station is modeled as an M/G/1 queue, the only parameters that need to be found are the average service time of the m^{th} station, x_m, and the second moment [3] of the service time of the m^{th} station, $\overline{x_m^2}$ [Bertsekas and Gallager 1992]. When this information is known, the average traffic service rate with respect to all the stations, x, can be found as:

$$x = \sum_{m=1}^{M} \frac{x_m}{M} \tag{5.1}$$

where M is the number of stations in the BSS. Furthermore, from the properties of the M/G/1 queue, we can get the average delay of the packets of the m^{th} station, T_m, and the average number of packets in the queue and server of the m^{th} station, N_m [2], [7]:

$$T_m = x_m + \frac{l_1 \overline{x_m^2}}{2(1-r_m)} \tag{5.2}$$

$$N_m = r_m + \frac{l_1^2 \overline{x_m^2}}{2(1-r_m)} \tag{5.3}$$

where $r_m = l_1 x_m$. To find the average service time of the m^{th} station, x_m, and the second moment of the service time of the m^{th} station, $\overline{x_m^2}$, the three situations mentioned above are now considered in turn.

5.1.1 Both station and channel are free

In this situation, when a packet arrives to the MAC layer of the m^{th} station, the station is not serving any packets (station is empty) and the channel is free. As a result, the service time for this packet depends on the number of stations that are going to collide with the m^{th} station. Figure 5.2 shows this situation.

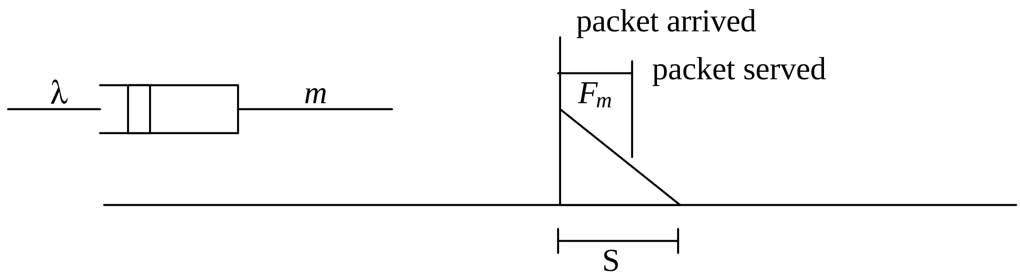


Figure 5.2 When a packet arrives to a station the channel and the station is empty

In the figure, S represents the time taken to serve all the collided stations and F_m is the time taken to serve the packet of the m^{th} station. F_m is a random variable and it depends on the number of stations that are going to collide with the m^{th} station and also which stations are going to collide.

5.1.1.1 Average number of collided stations

As shown in Figure 5.2, the service time of the packets arriving at the m^{th} station in this situation is F_m. This service time is a random variable and depends on the number of stations that are going to collide with the m^{th} station. The average number of stations collided including the m^{th} station, N_r can be given as:

$$N_r = \sum_{i=1}^{M} i p_i \tag{5.4}$$

where $p_i = Pr\{i \text{ } stations\, ready \mid i \geq 1\}$ is the probability that there is i ready stations in the channel given that there is at least one station ready in the channel, and M is the number of stations in the channel. Following [Gross and Harris 1974], using the laws of conditional probability,

$$p_i = \frac{\Pr\{i \text{ } stations\, ready\ and\ i \geq 1\}}{\Pr\{i \geq 1\}}$$

$$p_i = \frac{p'_i}{\sum_{i=1}^{M} p'_i} \qquad (i \geq 1)$$

where p'_i is the unconditional probability that there is i stations ready in the channel. As each station is represented by an M/G/1 queue, the probability that a station is ready (*or* probability that a station has non-empty server) can be given as $r = l_1 x$. Then from binomial law, the p'_i can be given as:

$$p'_i = C_i^M (1-r)^{M-i}\, r^i$$

where C_x^y is the binomial coefficient. Furthermore, the probability that the channel has at least one station ready, $\Pr\{i \geq 1\}$ can be given as:

$$\begin{aligned}\Pr\{i \geq 1\} &= 1 - \textit{probability that no station is ready}\\ &= 1 - \Pr\{i < 1\}\\ &= 1 - C_0^M (1-r)^M\, r^0\\ &= 1-(1-r)^M\end{aligned}$$

Hence p_i can be expressed as:

$$p_i = \frac{C_i^M (1-r)^{M-i}\, r^i}{1-(1-r)^M} \tag{5.5}$$

5.1.1.2 Calculating the probability of the situation

If there is N_r collided station in a collision then the time taken to serve the collision, $S(N_r)$ can be found from the parameters of the MAC scheme as:

If $N_r = 1$

$$S(N_r) = AIFSN_1 + 3 * SIFS + (RTS + CTS + D + MACHeader + PHYHeader + ACK)/C$$

Else if $N_r > 1$

$$\begin{aligned}S(N_r) = {} & AIFSN_1 + RTS/C + AIFSC_1 + CRB + CRIFS * M + PPB * N_r\\ & + NPB * (M - N_r) + ((D + MACHeader + PHYHeader + ACK + TP + TR)/C + 3 * SIFS\\ & + SDIFS) * (N_r - 1) + ((D + MACHeader + PHYHeader + ACK)/C + SIFS + SDIFS)\end{aligned}$$

where the above parameters are described in Table 5.0.

Table 5.0 MAC parameter descriptions

Parameters	Description	Units
SIFS	Short Inter Frame Space duration	seconds
RTS	Request To Send frame length	bits
CTS	Clear To Send frame length	bits
D	Data payload length	bits
C	Channel bit rate	bits/second
MACHeader	MAC layer header length	bits
PHYHeader	Physical layer header length	bits
ACK	Acknowledgement frame length	bits
CRB	Collision resolution beacon duration	seconds
CRIFS	Collsion Resolution IFS duration	seconds
PPB	Packet Present Beacon duration	seconds
NPB	No Packet Beacon duration	seconds
TP	Token Pass frame length	bits
TR	Token Receive frame length	bits
SDIFS	Scheduled Data IFS duration	seconds

S is the average time taken to serve all the collided stations in a given collision. The collision service time *S*, and the second moment of the collision service time $\overline{S^2}$, can be found as:

$$
\begin{aligned}
S &= \sum_{N_r=1}^{M}\left(time\,taken\,to\,serve\,the\,collision\;with\,Nr\;station \times p_{N_r}\right) \\
&= \sum_{N_r=1}^{M} S(N_r)p_{N_r} \\
&= \sum_{N_r=1}^{M} S(N_r)\frac{\left(C_{N_r}^{M}(1-r)^{M-N_r}r^{N_r}\right)}{\left(1-(1-r)^{M}\right)}
\end{aligned}
$$

and

$$
\begin{aligned}
\overline{S^2} &= \sum_{N_r=1}^{M}\left(\left[time\;taken\;to\;serve\;the\;collision\,with\;Nr\;station\right]^2 \times p_{N_r}\right) \\
&= \sum_{N_r=1}^{M} S(N_r)S(N_r)\,p_{N_r} \\
&= \sum_{N_r=1}^{M} S(N_r)S(N_r)\frac{\left(C_{N_r}^{M}(1-r)^{M-N_r}r^{N_r}\right)}{\left(1-(1-r)^{M}\right)}
\end{aligned}
$$

where p_{N_r} is the probability that there is N_r ready stations in the channel given that the channel has at least one station ready, M is the number of stations in the channel and $S(N_r)$ is defined above.

As S is the average time taken to serve all the collided stations for a given collision, and on average N_r stations are collided in each collision, then it is known that on average N_r packets are served in S seconds. If x_m is the average service time for the m^{th} station then the average service time, x, with respect to all the stations is given in equation (5.1).

Over any given time interval, t, the amount of busy time of the server at the m^{th} station, $STA_{m\,busy}$ can be found as:

$$STA_{m\,busy} = rt = x_m l_1 t \text{ seconds}$$

Considering all the stations in the channel, the total number of packets arriving in this time interval, t, is Mlt. As on average N_r stations are collided in each collision, the total number of collisions served during the time interval, t, is Mlt / N_r. From above, average time taken to serve all the collided stations in each collision is S seconds, therefore the total time required to serve all the packets arriving in this time interval is $MlSt / N_r$. This is also equivalent to the total time that the channel is busy, C_{busy}, during that time interval, t. Hence C_{busy} is given by:

$$C_{busy} = \frac{MlSt}{N_r} \tag{5.6}$$

This information is shown in Figure 5.3.

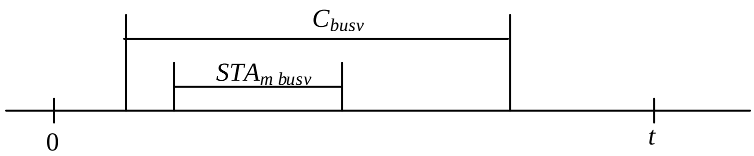


Figure 5.3 Status of the channel during a time interval t.

As mentioned before, in this event, when a packet arrives to the MAC layer of the m^{th} station, the station is not serving any packets (station is empty) and the channel is free. From the figure,

$$P\{\,m^{th} \textit{ station is empty and the channel is free}\}=P\{\textit{the channel is free}\}$$
$$P\{\textit{the channel is free}\}=1-P\{\textit{the channel is busy}\}=p_{C_{free}}$$

Hence the probability of the channel being free, $p_{C_{free}}$ is equal to the probability of this event happening and this can be given as

$$\begin{aligned} p_{C_{free}} &= \frac{\text{Amount of time that the channel is free during the time interval , t}}{\text{Duration of the time interval , t}} \\ &= \frac{t - \frac{MlS}{N_r} t}{t} \\ &= 1 - \frac{MlS}{N_r} \end{aligned} \quad (5.7)$$

5.1.1.3 Calculating the average service time for this situation

As shown in Figure 5.2, in this situation, the service time of the packet arrived to the m^{th} station is F_m. To find an expression for F_m, first consider the following example.

Say there are $M = 6$ stations in the channel, and 3 of them are collided, so that $N_r = 3$. To find the value of F_m for this example, we need to know the probabilities of the m^{th} station (one of the stations that has collided) being serviced at 1^{st}, 2^{nd} or 3^{rd} position.

If $m = 1$, the probability of the m^{th} station being serviced at the 1^{st} position will definitely be equal to 1 and the probability of this station being serviced at other positions will be equal to zero. Likewise, if $m = 3$, the probability of the m^{th} station being serviced at the 1^{st} position will be equal to the probability that first 2 stations are not members of the set of collided stations.

$$\begin{aligned} &P\{\text{this station been serviced at the first position}\} \\ &= P\{\text{first 2 stations are not included in the collided stations}\} \\ &= \frac{\left[\begin{array}{l} \text{choosing zero station among first 2 stations} \times \text{choosing other} \\ \text{2 collided stations among remaining 3 stations (} 4^{th}, 5^{th} \text{ and } 6^{th} \text{ station)} \end{array}\right]}{\left[\begin{array}{l} \text{number of different ways to choose other 2 collided stations} \\ \text{among remaining stations (} 1^{st}, 2^{nd}, 4^{th}, 5^{th} \text{ and } 6^{th} \text{ station)} \end{array}\right]} \\ &= \frac{C_0^2 C_2^3}{C_2^5} \end{aligned}$$

where C_x^y is the binomial coefficient. The probability of this station being serviced at the second position will be equal to the probability that there is one collided station among the first 2 stations and the probability that there is one collided station among the 3 remaining stations.

$$
\begin{aligned}
&P\{\text{this station been serviced at the second position}\}\\
&= P\{\text{there is one collided station among first 2 staions}\}\times\\
&\ P\{\text{there is one collided station among remaining 3 stations}\}\\
&= \frac{\begin{bmatrix}\text{choosing 1 station among first 2 stations} \times \text{choosing other}\\ \text{collided station among remaining 3 stations } (4^{th}, 5^{th} \text{ and } 6^{th} \text{ station})\end{bmatrix}}{\begin{bmatrix}\text{number of different ways to choose other 2 collided stations}\\ \text{among remaining stations } (1^{st}, 2^{nd}, 4^{th}, 5^{th} \text{ and } 6^{th} \text{ station})\end{bmatrix}}\\
&= \frac{C_1^2 C_1^3}{C_2^5}
\end{aligned}
$$

The probability of this station being serviced at the third position will be equal to the probability that there are 2 collided stations among the first 2 stations and the probability that there is no collided station among the 3 remaining stations.

$$
\begin{aligned}
&P\{\text{this station been serviced at the third position}\}\\
&= P\{\text{there are 2 collided stations among first 2 staions}\}\times\\
&\ P\{\text{there is no collided station among remaining 3 stations}\}\\
&= \frac{\begin{bmatrix}\text{choosing 2 station among first 2 stations} \times \text{choosing no}\\ \text{collided station among remaining 3 stations } (4^{th}, 5^{th} \text{ and } 6^{th} \text{ station})\end{bmatrix}}{\begin{bmatrix}\text{number of different ways to choose other 2 collided stations}\\ \text{among remaining stations } (1^{st}, 2^{nd}, 4^{th}, 5^{th} \text{ and } 6^{th} \text{ station})\end{bmatrix}}\\
&= \frac{C_2^2 C_0^3}{C_2^5}
\end{aligned}
$$

The probability of this station being serviced at the other positions will be equal to zero because only 3 stations are collided. Likewise, we can find all these probabilities for all the stations and the result is summarized in Table 5.1.

Table 5.1 Summarized results showing the probability of a station being service at a given position

		m^{th} station					
		1	2	3	4	5	6
Probability of m^{th} station been served at	1st	1	$\frac{C_0^1 C_2^4}{C_2^5}$	$\frac{C_0^2 C_2^3}{C_2^5}$	$\frac{C_0^3 C_2^2}{C_2^5}$	0	0
	2nd	0	$\frac{C_1^1 C_1^4}{C_2^5}$	$\frac{C_1^2 C_1^3}{C_2^5}$	$\frac{C_1^3 C_1^2}{C_2^5}$	$\frac{C_1^4 C_1^1}{C_2^5}$	0
	3rd	0	0	$\frac{C_2^2 C_0^3}{C_2^5}$	$\frac{C_2^3 C_0^2}{C_2^5}$	$\frac{C_2^4 C_0^1}{C_2^5}$	1
	4th	0	0	0	0	0	0
	5th	0	0	0	0	0	0
	6th	0	0	0	0	0	0

The time taken to finish serving a station which is served at the y^{th} position, F_m^y is another parameter that is required for calculating F_m. From the properties of the MAC, this timing can be given as below for a general case.

If N_r stations are ready and the m^{th} station is served at the y^{th} position ($1 \le y \le N_r$) then the time taken to serve the station, F_m^y, is

If $N_r = 1$

$$F_m^y = AIFSN_1 + 3 * SIFS + (RTS + CTS + D + MACHeader + PHYHeader + ACK)/C$$

Else if $N_r > 1$ and $y = 1$

$$F_m^y = AIFSN_1 + RTS/C + AIFSC_1 + CRB + CRIFS * M + PPB * N_r + NPB * (M - N_r) + ((D + MACHeader + PHYHeader + ACK)/C + SIFS + SDIFS)$$

Else if $N_r > 1$ and $y > 1$

$$F_m^y = AIFSN_1 + RTS/C + AIFSC_1 + CRB + CRIFS * M + PPB * N_r + NPB * (M - N_r) + ((D + MACHeader + PHYHeader + ACK)/C + SIFS + SDIFS) + ((D + MACHeader + PHYHeader + ACK + TP + TR)/C + 3 * SIFS + SDIFS) * (y - 1)$$

where the MAC parameters are specified in the Table 5.0.

These are the only pieces of information required to calculate the time taken to serve the packet of the m^{th} station, F_m. From the example, the time taken to serve the packet of the 3^{rd} station, F_3 can be calculated by using the probabilities given in table 5.1 as below:

$$F_3 = \frac{C_0^2 C_2^3}{C_2^5} F_3^1 + \frac{C_1^2 C_1^3}{C_2^5} F_3^2 + \frac{C_2^2 C_0^3}{C_2^5} F_3^3 = \sum_{y=1}^{N_r} \frac{C_{y-1}^{N_r-1} C_{N_r-y}^{M-m}}{C_{N_r-1}^{M-1}} F_m^y$$

By using this information, for other stations can also be calculated. Similar reasoning can be used to calculate F_m for a general case and the results obtained are given below.

If a packet of the m^{th} station is collided, the average time taken to serve the collided packet in that collision serving time, F_m, is given by:

If N_r £ $M/2$ *and* 1 £ m £ $N_r - 1$

$$F_m = \sum_{y=1}^{m} \frac{C_{y-1}^{m-1} C_{N_r-y}^{M-m}}{C_{N_r-1}^{M-1}} F_m^y \qquad (5.8\text{-}1)$$

Else if N_r £ $M/2$ *and* N_r £ m £ $M-N_r + 1$

$$F_m = \sum_{y=1}^{N_r} \frac{C_{y-1}^{m-1} C_{N_r-y}^{M-m}}{C_{N_r-1}^{M-1}} F_m^y \qquad (5.8\text{-}2)$$

Else if N_r £ $M/2$ *and* $M-N_r + 2$£ m £ M

$$F_m = \sum_{y=1}^{M-m+1} \frac{C_{N_r+y+m-M-2}^{m-1} C_{M-m-y+1}^{M-m}}{C_{N_r-1}^{M-1}} F_m^{N_r+y+m-M-1} \qquad (5.8\text{-}3)$$

Else if $N_r > M/2$ *and* 1 £ m £ $M - N_r + 1$

$$F_m = \sum_{y=1}^{m} \frac{C_{y-1}^{m-1} C_{N_r-y}^{M-m}}{C_{N_r-1}^{M-1}} F_m^y \qquad (5.8\text{-}4)$$

Else if $N_r > M/2$ *and* $M - N_r + 2$ £ m £ $N_r - 1$

$$F_m = \sum_{y=1}^{M-N_r+1} \frac{C_{m-(M-N_r+2)+y}^{m-1} C_{M-m+1-y}^{M-m}}{C_{N_r-1}^{M-1}} F_m^{m-(M-N_r+2)+y+1} \qquad (5.8\text{-}5)$$

Else if $N_r > M/2$ *and* N_r £ m £ M

$$F_m = \sum_{y=1}^{M-m+1} \frac{C_{m-(M-N_r+2)+y}^{m-1} C_{M-m+1-y}^{M-m}}{C_{N_r-1}^{M-1}} F_m^{m-(M-N_r+2)+y+1} \qquad (5.8\text{-}6)$$

Hence the average service time of the packets in this situation, $x_{m1} = F_m$, where F_m is given in equation (5.8).

5.1.2 The station is empty but the channel is busy

In this situation, when a packet arrives to the MAC layer of the m^{th} station, the station is not serving any packets (station is empty) but the channel is busy. As a result, the service time for this packet depends on the amount of time remaining to serve previously collided stations and the number of stations that are going to collide with the m^{th} station. Figure 5.4 shows this scenario.

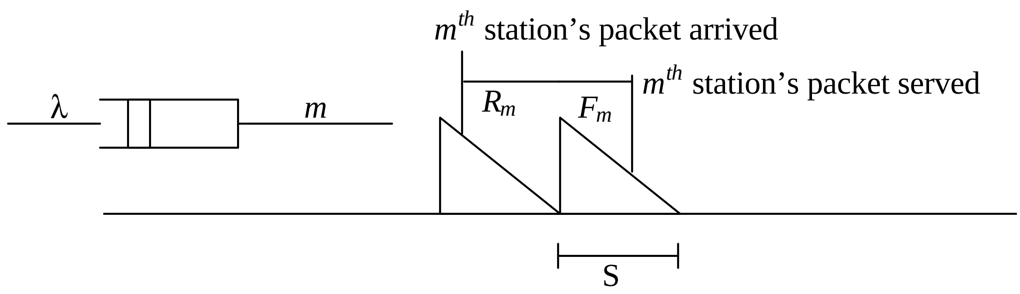


Figure 5.4 The scenario when a packet arrives to a station, the channel is busy but the station is empty

In the figure, S represents the average time taken to serve all the collided stations in a given collision and F_m is the time taken to serve the packet of the m^{th} station and R_m is the time remaining to serve previously collided stations [Bertsekas and Gallager 1992]. The average service time of this event, $x_{m2} = R_m + F_m$, where F_m is given in equation (5.8) and R_m needs to be found.

5.1.2.1 Calculating the probability of the situation

In this situation, when a packet arrives to the MAC layer of the m^{th} station, the station is not serving any packets but the channel is busy. From Figure 5.3,

$$P\{\ m^{th}\ station\ is\ empty\ and\ the\ channel\ is\ busy\}$$
$$=\ P\{the\ channel\ is\ busy\}\,P\{\ m^{th}\ station\ is\ empty\,|\,the\ channel\ is\ busy\}$$

From section 5.1.1.2, the amount of busy time of the server of the m^{th} station is $STA_{m\ busy} = x_m\, l t$ seconds, where x_m is the average service time for the m^{th} station, and l_1 is the average packet arrival rate. Hence from this

$$P\{\ m^{th}\ station\ busy\ \} = x_m\, l_1$$

$$P\{\ m^{th}\ station\ is\ empty\,|\,the\ channel\ is\ busy\ \} = \frac{\dfrac{M l_1 S}{N_r} - x_m\, l_1}{\dfrac{M l_1 S}{N_r}}$$

Therefore, probability of this event happening

$$P\{\ m^{th}\ station\ is\ empty\ and\ the\ channel\ is\ busy\} = \frac{M l_1 S}{N_r} - x_m\, l_1 \qquad (5.9)$$

5.1.2.2 Calculating the time remaining to serve previously collided stations, R_m

R_m can be calculated by using graphical augment as in [Bertsekas and Gallager 1992]. Figure 5.5 shows the residual service time $r(t)$ (i.e., the remaining time for completion of service for collided stations at time t) as a function of t.

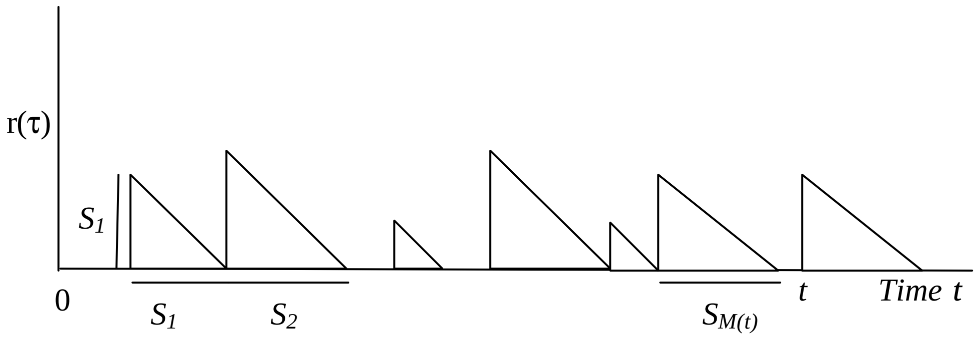


Figure 5.5 Residual service time

However in this situation we are only interested in the time when the m^{th} station is empty but the channel is busy. From section 5.1.2.1, in a given time interval t, the amount of time that the m^{th} station is empty but the channel is busy, $t' = \frac{M l_1 S t}{N_r} - x_m l_1 t$. Figure 5.5 can be changed to suit for this situation. The new figure for this situation is shown in Figure 5.6.

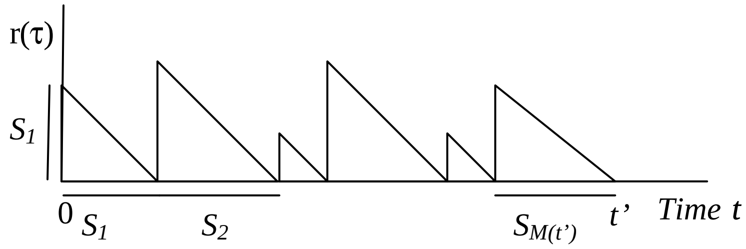


Figure 5.6 Residual service time of the situation, when a packet arrives to a station, the station is empty but the channel is busy

The total number of packets arriving in a time interval, t is $M l_1 t$ and it is known that on average N_r packets are served in S seconds of channel busy time. The total amount of the m^{th} station busy time is $x_m l_1 t$ at a given time, t. From this information, the number of packets served during the m^{th} station busy time is $N_r x_m l_1 t / S$. So the number of packets that need to be served in t' time is $(M l_1 - N_r x_m l_1 / S)t$. Hence

$$
\begin{aligned}
&\textit{Average number of collisions need to be serviced in } t' \textit{ time}, M(t') \\
&= \frac{(M l_1 - N_r x_m l_1 / S)t}{N_r} \\
&= (M l_1 / N_r - x_m l_1 / S)t
\end{aligned}
$$

Now from Figure 5.6, consider a time t' for which $r(t') = 0$. The time average of $r(t)$ in the interval [0, t'] is

$$\frac{1}{t'}\int_0^{t'} r(t)dt = \frac{1}{t'}\sum_{i=1}^{M(t')}\frac{1}{2}S_i^2$$

where $M(t')$ is the number of collisions served within [0, t'], and S_i is the service time of i^{th} collided packets. The equation can also be written as

$$\frac{1}{t'}\int_0^{t'} r(t)dt = \frac{M(t')}{t'} \frac{\sum_{i=1}^{M(t')} \frac{1}{2} S_i^2}{M(t')}$$

Assuming the limit below exists and obtain

$$\lim_{t' \to \infty} \frac{1}{t'}\int_0^{t'} r(t)dt = \frac{1}{2} \lim_{t' \to \infty} \frac{M(t')}{t'} \lim_{t' \to \infty} \frac{\sum_{i=1}^{M(t')} S_i^2}{M(t')}$$

Assuming that the time averages can be replaced by ensemble averages, the last limit on the right is the second moment of the collision service time. Thus the equation becomes

$$\lim_{t' \to \infty} \frac{1}{t'}\int_0^{t'} r(t)dt = \frac{1}{2} \frac{(Ml_1 / N_r - x_m l_1 / S)t}{\frac{Ml_1 St}{N_r} - x_m l_1 t} \overline{S^2} = \frac{\left[\frac{Ml_1}{N_r} - \frac{x_m l_1}{S}\right]\overline{S^2}}{2\left[\frac{Ml_1 S}{N_r} - x_m l_1\right]}$$

Hence, the time remaining to serve previously collided stations, R_m

$$R_m = \frac{\left[\frac{Ml_1}{N_r} - \frac{x_m l_1}{S}\right]\overline{S^2}}{2\left[\frac{Ml_1 S}{N_r} - x_m l_1\right]} \qquad (5.10)$$

The average service time of the packets in this situation, $x_{m2} = R_m + F_m$, where F_m is given in equation (5.8) and R_m is given in the equation (5.10).

5.1.3 The station is busy in service

In this situation, when a packet arrives to the MAC layer of m^{th} station, the station is serving a packet so the channel must be busy. Figure 5.7 shows this scenario. As shown in the figure, in this situation, the time taken to service the arrived packet is S seconds. Therefore the average service time of this event, $x_{m3} = S$. The S is the average time taken to serve all the collided stations in a given collision and is given in section 5.1.1.2.

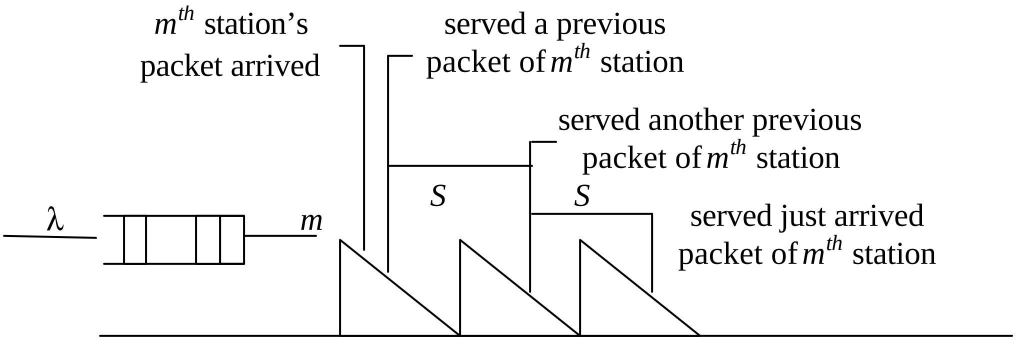


Figure 5.7 The scenario showing a station is busy when a packet arrives to the station

5.1.3.1 Calculating the Probability of the situation

In this situation, when a packet arrives to the MAC layer of the m^{th} station, the station is busy. From Figure 5.3,

$$P\{\ m^{th}\ station\, is\ busy\}$$
$$= \frac{x_m\, l_1 t}{t} = x_m\, l_1$$

Therefore, probability of this event happening

$$P\{\ m^{th}\ station\, busy\} = x_m\, l_1 \qquad (5.11)$$

5.1.4 Calculation of the average service time of a station

The above three sections (5.1.1, 5.1.2 and 5.1.3) describe the three possible situations that may happen when a packet arrives to a station. The probabilities of these events and average service time of the packets accompanied with these events are also derived in the above three sections. The summary of these is given in Table 5.2.

Table 5.2 Summarized results of the above three sections

		Probability of the Event	Average service time
Events	Station is empty & Channel is empty	$1 - \frac{M l_1 S}{N_r}$	$x_{m1} = F_m$
	Station is empty & Channel is busy	$\frac{M l_1 S}{N_r} - x_m l_1$	$x_{m2} = R_m + F_m$
	Station is busy	$x_m l_1$	$x_{m3} = S$

From the information in Table 5.2, an equation for the average service time of the m^{th} station, x_m can be given as:

$$
\begin{aligned}
x_m &= \left[1 - \frac{M l_1 S}{N_r}\right] F_m + \left[\frac{M l_1 S}{N_r} - x_m l_1\right](R_m + F_m) + x_m l_1 S \\
&= F_m - x_m l_1 F_m + \left[\frac{M l_1 S}{N_r} - x_m l_1\right] R_m + x_m l_1 S
\end{aligned}
$$

Then by substituting R_m given in equation (5.10)

$$
\begin{aligned}
x_m &= F_m - x_m l_1 F_m + \left[\frac{M l_1 S}{N_r} - x_m l_1\right] \frac{\left[\frac{M l_1}{N_r} - \frac{x_m l_1}{S}\right]\overline{S^2}}{2\left[\frac{M l_1 S}{N_r} - x_m l_1\right]} + x_m l_1 S \\
&= F_m - x_m l_1 F_m + \left[\frac{M l_1}{2N_r} - \frac{x_m l_1}{2S}\right]\overline{S^2} + x_m l_1 S
\end{aligned}
$$

Hence the average service time of the m^{th} station, x_m is

$$
x_m = \frac{F_m + \frac{M l_1 \overline{S^2}}{2N_r}}{1 + l_1 F_m + \frac{l_1 \overline{S^2}}{2S} - l_1 S} \tag{5.12}
$$

When x_m is known the average service time, x, for all the stations can be found by using equation (5.1)

$$x = \sum_{m=1}^{M} \frac{x_m}{M}$$

As F_m, N_r and S are parameters dependent on x, the equations (5.1) can be solved by the substitution method.

5.1.5 Average Delay of the packets, T_1

After solving equation (5.1), the corresponding value for x_m is known. Then the second moment of the service time of the m^{th} station, $\overline{x_m^2}$ can be found from table 5.2 [Stark and Woods 2002].

$$\overline{x_m^2} = \left[1 - \frac{M l_1 S}{N_r}\right] F_m{}^2 + \left[\frac{M l_1 S}{N_r} - x_m l_1\right](R_m + F_m)^2 + x_m l_1 S^2 \tag{5.13}$$

From properties of the M/G/1 queue [Bertsekas and Gallager 1992], we can now obtain the average delay of the packets of the m^{th} station, T_m:

$$T_m = x_m + \frac{l_1 \overline{x_m^2}}{2(1 - r_m)}$$

As the average arrival rate of packets for all the stations are equal, on average the total number of packets being served by each station must be equal. So the average delay of the packets, T can be given as:

$$T = \sum_{m=1}^{M} \frac{T_m}{M} \tag{5.14}$$

Analytical results obtained for the average delay of the packets, T is shown in Figure 5.8. The MAC parameters used in the analytical model is shown in Table 5.3.

Table 5.3: The MAC parameters

Data packet payload (D)		1280	
MAC header		272 bits	
PHY header		128 bits	
ACK		112 bits + PHY header	
RTS		180 bits + PHY header	
CTS		112 bits + PHY header	
TP		112 bits + PHY header	
TR		112 bits + PHY header	
Channel Bit rate (C)		2 Mbps	
Slot Time	20 µs	CRB	150 µs
SIFS	10 µs	$AIFSC_2$	90 µs
CRIFS	20µs	AIFSN2	110 µs
SDIFS	30 µs	CRB	150 µs
$AIFSC_1$	50 µs	PPB	30 µs
$AIFSN_1$	70 µs	NPB	10 µs

The figure shows the results obtained for different Poisson arrival rates. From the figure, it shows that the average packet delay increases as the number of stations in the BSS increase. The figure also shows that when arrival rate of the packets at the stations increase, the average packet delay also increases. The reason being, as the number of stations and arrival rate of the packets increase, there will be more collisions and the number of stations involved in each collision increases. As a result, the average delay of the packets increase.

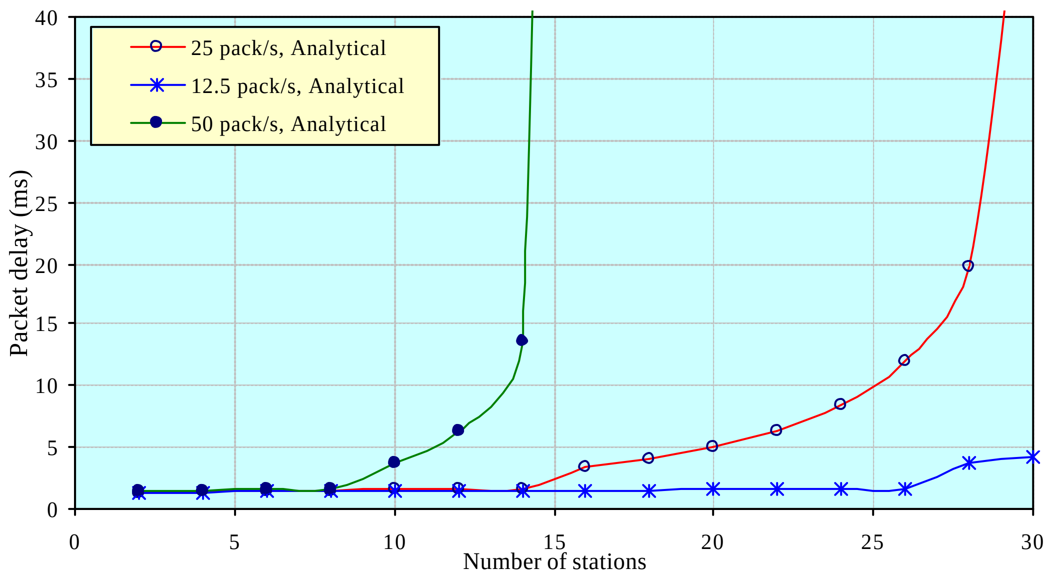


Figure 5.8 Average delay experienced by the priority one packets

5.2 Stations have priority one and two traffic but the channel is not fully occupied

In this case, priority one and two traffic are considered. The MAC scheme can be modeled as shown in Figure 5.9. For each priority class traffic, it is assumed that traffic arrives to the stations according to a Poisson process, traffic service times have a general distribution and every priority has a queue with infinite buffer space. With the help of these assumptions, the MAC layer of each traffic category of a station is represented by an M/G/1 queue [Bertsekas and Gallager 1992], [Kleinrock 1975], [Kleinrock 1976], [Kleinrock and Gail 1996]. As shown in the figure, traffic arrive to priority one and priority two queues with average rates of l_1 and l_2 packets per second, respectively. Traffic serving rates, m_1 and m_2 are the average traffic service rates with respect to a station for priority one and priority two, respectively. That means, for example, if an average time taken for a station to serve a priority one packet is x second then $m_1 = 1/x$.

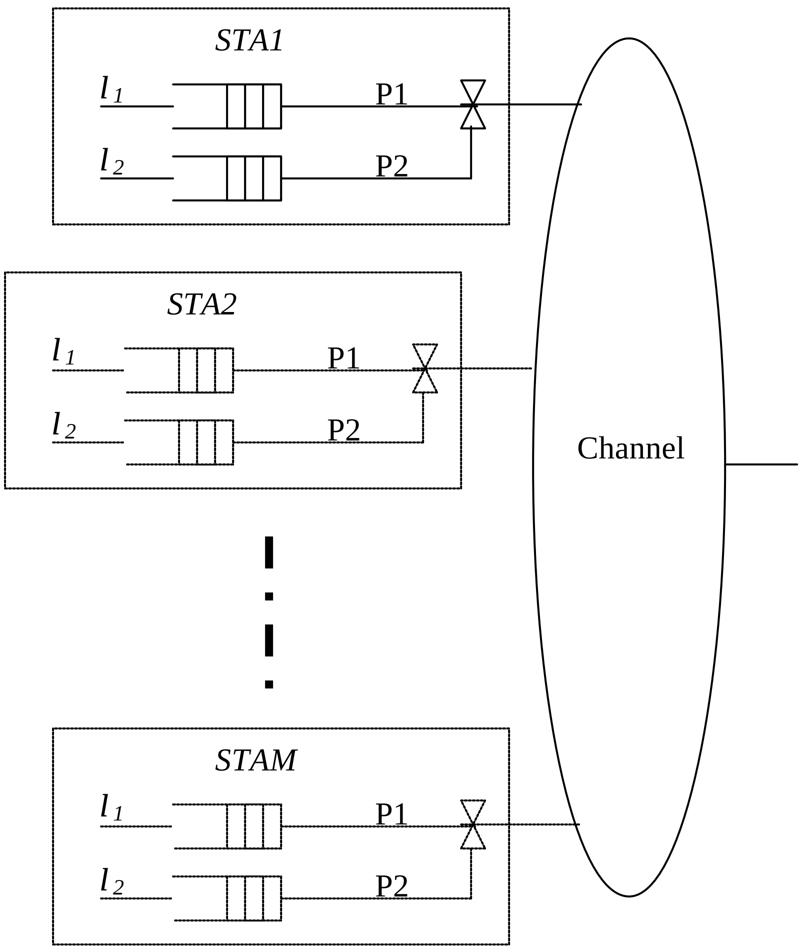


Figure 5.9 MAC model for stations, which has two traffic categories

5.2.1 Priority one traffic

First consider priority one traffic only. When a priority one packet arrives to the MAC layer of a station, three possible situations may arise:

- Priority one queue is not serving a packet and the channel is free
- Priority one queue is not serving a packet but the channel is busy
- Priority one queue is serving a packet

As each priority traffic of a station is modeled as an M/G/1 queue, the only parameters that need to be found are the average service time of the m^{th} station, x_m, and the second moment of the service time of the m^{th} station, $\overline{x_m^2}$. To find these parameters, the above three situations are now considered in turn.

5.2.1.1 Priority one queue is not serving a packet and the channel is free

In this situation, when a packet arrives to the priority one queue of the m^{th} station, the station is not serving any packets (station is empty) and the channel is free. As a result, the service time for this packet depends on the number of stations that are going to collide with the m^{th} station with priority one packets. Figure 5.10 shows this scenario.

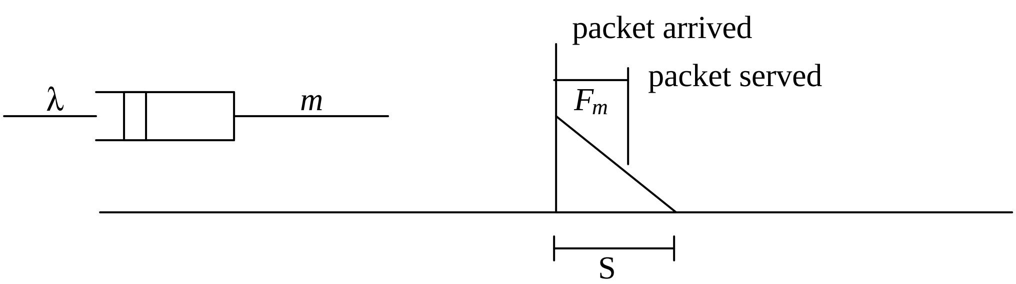


Figure 5.10 When a packet arrives to the priority one queue of a station the channel and the queue is empty

In the figure, S represents the average time taken to serve all the collided stations in a given collision and F_m is the time taken to serve the packet of the m^{th} station. As before, F_m is a random variable and it depends on the number of stations that are going to collide with the m^{th} station and also which stations are going to collide. The value for F_m is given in the equation (5.8).

5.2.1.1.1 Average number of priority two collided stations

As for the priority one (*P1*) case, the average number of stations collided with priority two (*P2*) traffic, N_{r2} can be given as

$$N_{r2} = \sum_{i=1}^{M} i p_{2i} \tag{5.15}$$

where p_{2i} is the probability that there is i ready stations with $P2$ traffic in the channel, given that the channel is not free and M is the number of stations in the channel. From this definition, p_{2i} can be given as [Gross and Harris 1974]:

$$p_{2i} = Pr\{i \text{ stations ready with } P2 \text{ traffics} \mid i \geq 1\}$$

Using the laws of conditional probability,

$$p_{2i} = \frac{Pr\{i \text{ stations ready with } P2 \text{ traffics and } i \geq 1\}}{Pr\{i \geq 1\}}$$

$$p_{2i} = \frac{p'_{2i}}{\sum_{i=1}^{M} p'_{2i}} \qquad (i \geq 1)$$

where p'_{2i} is the probability that there is i stations ready in the channel. As the $P2$ queue of each station is represented by an M/G/1 queue, the probability that a station is ready is $r_2 = l_2 x_2$, where x_2 is the average service time of the $P2$ packets. Hence,

$$p_{2i} = \frac{C_i^M (1-r_2)^{M-i} r_2^{\,i}}{1 - C_0^M (1-r_2)^M r_2^{\,0}} = \frac{C_i^M (1-r_2)^{M-i} r_2^{\,i}}{1-(1-r_2)^M} \tag{5.16}$$

5.2.1.1.2 Calculating the probability of this situation

As S is the time taken to serve all priority one collided stations, and on average N_r stations are collided, then it is known that N_r packets are served in S seconds. If x_m is the average service time of priority one packets of the m^{th} station then the average service time of priority one packets, x, for all the stations is given in equation (5.1).

$$x = \sum_{m=1}^{M} \frac{x_m}{M}$$

If there is N_{r2} collided station in a collision with priority two traffic then the time taken to serve the collision, $S(N_{r2})$ can be found from the parameters of the MAC scheme as:

If $N_{r2} = 1$

$$S(N_{r2}) = AIFSN_2 + 3*SIFS + (RTS + CTS + D + MACHeader + PHYHeader + ACK)/C$$

Else if $N_{r2} > 1$

$$\begin{aligned} S(N_{r2}) &= AIFSN_2 + RTS/C + AIFSC_2 + CRB + CRIFS*M + PPB*N_{r2} \\ &+ NPB*(M - N_{r2}) + ((D + MACHeader + PHYHeader + ACK + TP + TR)/C + 3*SIFS \\ &+ SDIFS)*(N_{r2} - 1) + ((D + MACHeader + PHYHeader + ACK)/C + SIFS + SDIFS) \end{aligned}$$

where the above parameters are described in Table 5.0. S_2 is the time taken to serve all priority two collided stations in a given collision. The $P2$ collision service time S_2 and the second moment of the $P2$ collision service time $\overline{S_2^2}$, can be found as:

$$\begin{aligned} S_2 &= \sum_{N_r=1}^{M} \left(time\, taken\, to\, serve\, the\, collision\, with\, Nr2\, station \times p_{2N_{r2}} \right) \\ &= \sum_{N_{r2}=1}^{M} S(N_{r2}) p_{2N_{r2}} \\ &= \sum_{N_{r2}=1}^{M} S(N_{r2}) \left(C_{N_{r2}}^{M} (1 - r_2)^{M - N_{r2}} r_2^{N_{r2}} \right) \Big/ \left(1 - (1 - r_2)^M \right) \end{aligned}$$

and

$$\begin{aligned} \overline{S_2^2} &= \sum_{N_{r2}=1}^{M} \left([time\, taken\, to\, serve\, the\, collision\, with\, Nr2\, station]^2 \times p_{2N_{r2}} \right) \\ &= \sum_{N_{r2}=1}^{M} S(N_{r2}) S(N_{r2}) p_{2N_{r2}} \\ &= \sum_{N_{r2}=1}^{M} S(N_{r2}) S(N_{r2}) \left(C_{N_{r2}}^{M} (1 - r_2)^{M - N_{r2}} r_2^{N_{r2}} \right) \Big/ \left(1 - (1 - r_2)^M \right) \end{aligned}$$

where $p_{2N_{r2}}$ is the probability that there is N_{r2} ready stations with $P2$ traffic in the channel, given that the channel is not free, M is the number of stations in the channel and $S(N_{r2})$ is defined above.

On average N_{r2} stations are collided with priority two packets, then it is known that N_{r2} packets are served in S_2 seconds. If x_{m2} is the average service time of priority two packets of the m^{th} station then the average service time of priority two packets, x_2, for all the stations is given by:

$$x_2 = \sum_{m=1}^{M} \frac{x_{m2}}{M} \tag{5.17}$$

The amount of time the channel is busy due to priority one traffic at a given time interval, t, is given in equation (5.6).

$$C_{busy} = \frac{Ml_1 St}{N_r}$$

If the arrival rate of priority two traffic for the m^{th} station is l_2 packets per second and as N_{r2} priority two packets are served in S_2 seconds and there is M stations in the channel, the total time required to serve all the priority two packets arriving in this time or the total time that the channel is busy due to priority two traffic, C_{busy2} is given by:

$$C_{busy2} = \frac{Ml_2 S_2 t}{N_{r2}} \tag{5.18}$$

This information is shown in Figure 5.11.

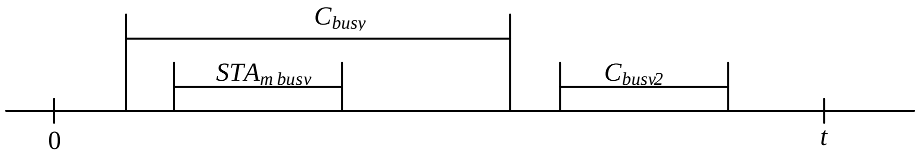


Figure 5.11 Status of the channel during a time interval t.

In the figure, $STA_{m\ busy}$ represents the amount of time that priority one server of the m^{th} station is busy and is given in section (5.1.1.2). As mentioned before, when a packet arrives to the priority one queue of the m^{th} station, the station is not serving any packet and the channel is free. From Figure 5.10,

$$P\{\ m^{th}\ \text{station's } P1 \text{ server is empty and the channel is free}\} = P\{\text{the channel is free}\}$$

$$P\{\text{the channel is free}\} = 1 - P\{\text{the channel is busy}\} = p_{C_{free2}}$$

Hence the probability that the channel is free, $p_{C_{free2}}$ is the probability of this event happening and can be given as:

$$p_{C_{free2}} = 1 - \frac{MlS}{N_r} - \frac{Ml_2S_2}{N_{r2}} \tag{5.19}$$

So the probability of this event is $p_{C_{free2}}$ and the average service time of this event, $x_{m1} = F_m$, where F_m is given in equation (5.8).

5.2.1.2 Priority one queue is not serving a packet but the channel is busy

In this situation, when a packet arrives to the priority one queue of the m^{th} station, the server of the queue is empty but the channel is busy. As a result, the service time for this packet depends on the amount of time remaining to serve previously collided *P1* stations or the amount of time remaining to serve previously collided *P2* stations and also the number of stations that are going to collide with the m^{th} station with *P1* traffic. Figure 5.12 shows this scenario.

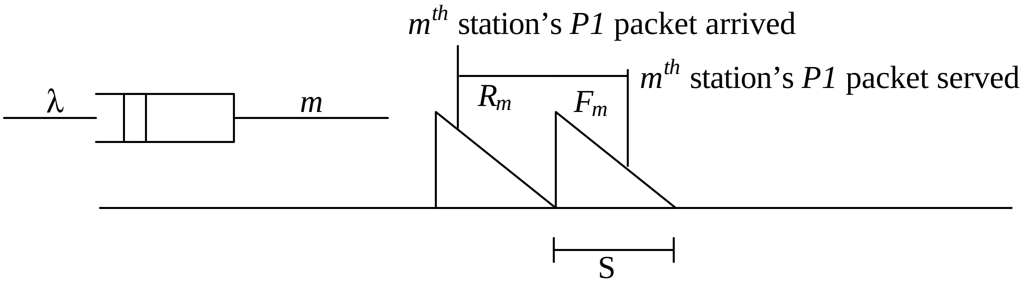


Figure 5.12 When a P1 packet arrives to a station the channel is busy but the station's P1 server is empty

In the figure, S represents the average time taken to serve all the $P1$ collided stations in a given collision and F_m is the time taken to serve the $P1$ packet of the m^{th} station and R_m is the time remaining to serve previously collided stations with either $P1$ or $P2$ traffic. The average service time of this event, $x_{m2} = R_m + F_m$, where F_m is given in equation (5.8) and R_m needs to be found.

5.2.1.2.1 Calculating the Probability of this situation

In this event, when a packet arrives to the priority one queue of the m^{th} station, the server of the queue is empty but the channel is busy. From Figure 5.11,

$$
\begin{aligned}
&P\{\ m^{th}\ station's\ P1\ server\ is\ empty\ and\ the\ channel\ is\ busy\} \\
&= P\{the\ channel\ is\ busy\}\ P\{\ m^{th}\ station's\ P1\ server\ is\ empty \,|\, the\ channel\ is\ busy\}
\end{aligned}
$$

From section 5.1.1.2, the amount of busy time of $P1$ server of the m^{th} station is $STA_{m\ busy} = x_m l_1 t$ seconds, where x_m is the average service time of $P1$ packets of the m^{th} station, and l_1 is the average packet arrival rate of $P1$. Hence from this

$$P\{\ m^{th}\ station's\ P1\ server\ busy\ \} = x_m l_1$$

$$P\{the\ channel\ is\ busy\ \} = \frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}}$$

$$P\{\ m^{th}\ station's\ P1\ server\ is\ empty \,|\, the\ channel\ is\ busy\ \} = \frac{\dfrac{M l_1 S}{N_r} + \dfrac{M l_2 S_2}{N_{r2}} - x_m l_1}{\dfrac{M l_1 S}{N_r} + \dfrac{M l_2 S_2}{N_{r2}}}$$

Therefore, probability of this event happening

$$
\begin{aligned}
&P\{\ m^{th}\ station's\ P1\ server\ is\ empty\ and\ the\ channel\ is\ busy\} \\
&= \frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}} - x_m l_1
\end{aligned}
\tag{5.20}
$$

5.2.1.2.2 Calculation of the time remaining to serve previous collided stations, R_m

R_m can be calculated by using graphical augment as in [Bertsekas and Gallager 1992]. Figure 5.13 shows the residual service time $r(t)$ (i.e., the remaining time for completion of service for collided stations at time t) as a function of t.

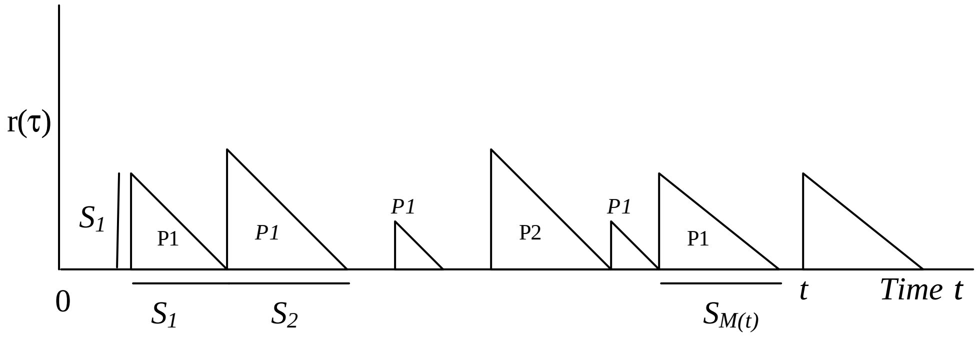


Figure 5.13 Residual service time

However in this situation we are only interested in the time when *P1* server of the m^{th} station is empty but the channel is busy. From section 5.2.1.2.1, at a given time t, the amount of time that the *P1* server of the m^{th} station is empty but the channel is busy, $t' = \frac{Ml_1 St}{N_r} + \frac{Ml_2 S_2 t}{N_{r2}} - x_m l_1 t$. Figure 5.13 can be changed to suit this situation. The new figure for this situation is shown in Figure 5.14.

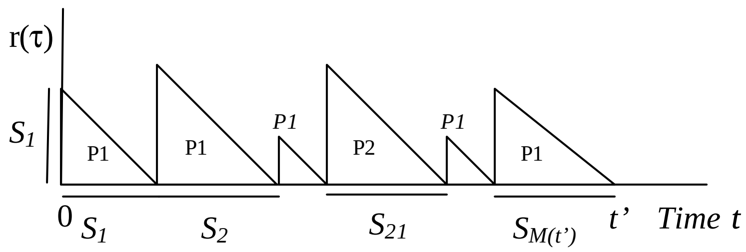


Figure 5.14 Residual service time for this situation

The total number of *P1* packets arriving during a time interval, t is $Ml_1 t$ and it is known that on average N_r, *P1* packets are served in S seconds of channel busy time. The total amount of the *P1* server busy time of the m^{th} station is $x_m l_1 t$ over the time t. From this information, the number of *P1* packets served during the *P1* server busy time of the m^{th} station is $N_r x_m l_1 t/S$. Hence, the number of *P1* packet that need to be served within the time interval t' is $(Ml_1 - N_r x_m l_1/S)t$. Hence,

$$\text{Average number of } P1 \text{ collisions need to be serviced in } t' \text{ time}, M(t')$$
$$= \frac{(Ml_1 - N_r x_m l_1/S)t}{N_r}$$
$$= (Ml_1 / N_r - x_m l_1/S)t$$

Similarly, the total number of *P2* packets arriving within a time interval, t is Ml_2t and it is known that on average N_{r2}, *P2* packets are served in S_2 seconds of channel busy time. So the number of *P2* packets that need to be serviced within time *t'* is Ml_2t. Hence,

$$\textit{Average number of P2 collisions that need to be serviced in t' time}, M2(t') = \frac{Ml_2t}{N_{r2}}$$

Now from Figure 5.14, consider a time *t'* for which $r(t') = 0$. The time average of $r(t)$ in the interval $[0, t']$ is

$$\frac{1}{t'}\int_0^{t'} r(t)dt = \frac{1}{t'}\sum_{i=1}^{M(t')}\frac{1}{2}S_i^2 + \frac{1}{t'}\sum_{i=1}^{M2(t')}\frac{1}{2}S_{2i}^2$$

where $M(t')$ is the number of *P1* collisions served within $[0, t']$ and $M2(t')$ is the number of *P2* collisions served within $[0, t']$, S_i is the service time of the i^{th} collided *P1* packets and S_{2i} is the service time of the i^{th} collided *P2* packets. The equation can also be written as

$$\frac{1}{t'}\int_0^{t'} r(t)dt = \frac{M(t')}{t'}\frac{\sum_{i=1}^{M(t')}\frac{1}{2}S_i^2}{M(t')} + \frac{M2(t')}{t'}\frac{\sum_{i=1}^{M2(t')}\frac{1}{2}S_{2i}^2}{M2(t')}$$

Assuming the limit below exists, we obtain

$$\lim_{t'\to\infty}\frac{1}{t'}\int_0^{t'} r(t)dt = \frac{1}{2}\lim_{t'\to\infty}\frac{M(t')}{t'}\lim_{t'\to\infty}\frac{\sum_{i=1}^{M(t')}S_i^2}{M(t')} + \frac{1}{2}\lim_{t'\to\infty}\frac{M2(t')}{t'}\lim_{t'\to\infty}\frac{\sum_{i=1}^{M2(t')}S_{2i}^2}{M2(t')}$$

Assuming that time averages can be replaced by ensemble averages, $\lim_{t'\to\infty}\frac{\sum_{i=1}^{M(t')}S_i^2}{M(t')}$ is the second moment of the *P1* collision service time, $\overline{S^2}$, and $\lim_{t'\to\infty}\frac{\sum_{i=1}^{M2(t')}S_{2i}^2}{M2(t')}$ is the second moment of the *P2* collision service time, $\overline{S_2^2}$. Thus the equation becomes

$$\lim_{t'\to\infty}\frac{1}{t'}\int_0^{t'} r(t)\,dt$$

$$=\frac{1}{2}\frac{\left(Ml_1/N_r - x_m l_1/S\right)t}{\dfrac{Ml_1St}{N_r}+\dfrac{Ml_2S_2t}{N_{r2}}-x_m l_1 t}\overline{S^2}+\frac{1}{2}\frac{\dfrac{Ml_2t}{N_{r2}}}{\dfrac{Ml_1St}{N_r}+\dfrac{Ml_2S_2t}{N_{r2}}-x_m l_1 t}\overline{S_2^{\,2}}$$

$$=\frac{\left[\dfrac{Ml_1}{N_r}-\dfrac{x_m l_1}{S}\right]\overline{S^2}+\dfrac{Ml_2}{N_{r2}}\overline{S_2^{\,2}}}{2\left[\dfrac{Ml_1S}{N_r}+\dfrac{Ml_2S_2}{N_{r2}}-x_m l_1\right]}$$

Hence, the time remaining to serve previous collided stations, R_m

$$R_m=\frac{\left[\dfrac{Ml_1}{N_r}-\dfrac{x_m l_1}{S}\right]\overline{S^2}+\dfrac{Ml_2}{N_{r2}}\overline{S_2^{\,2}}}{2\left[\dfrac{Ml_1S}{N_r}+\dfrac{Ml_2S_2}{N_{r2}}-x_m l_1\right]} \tag{5.21}$$

The average service time of this event, $x_{m2} = R_m + F_m$, where F_m is given in equation (5.8) and R_m is given in equation (5.21).

5.2.1.3 Priority one queue is serving a packet

In this situation, when a packet arrives to *P1* queue of the m^{th} station, the *P1* server is serving a packet so the channel must be busy. Figure 5.15 shows this scenario.

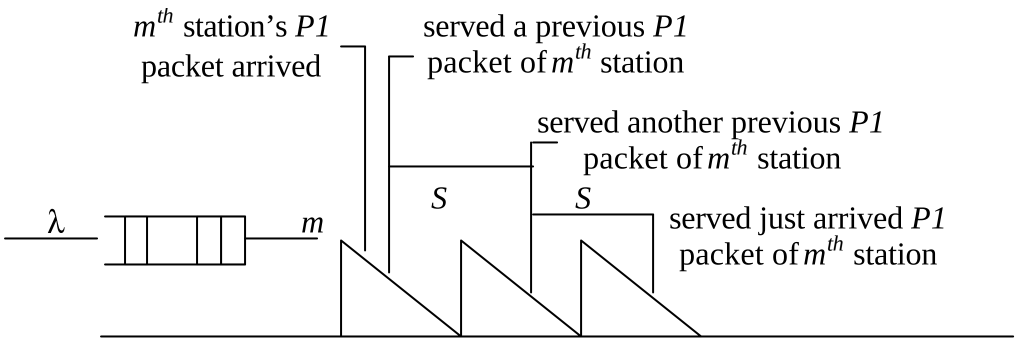


Figure 5.15 When a *P1* packet arrives to a station *P1* server is busy

As shown in the figure, in this situation, the time taken to service the arrived *P1* packet is *S* seconds. Therefore, the average service time of this event, $x_{m3} = S$, where *S* is the time taken to serve all the collided stations of *P1* in a given *P1* collision. *S* is given in section (5.1.2).

5.2.1.3.1 Calculating the probability of this situation

In this situation, when a packet arrives to the *P1* queue of the m^{th} station, *P1* server is serving a packet. From Figure 5.11,

$$P\{\ m^{th}\ station's\ P1 server\ is\ busy\}$$
$$= \frac{x_m l_1 t}{t} = x_m l_1$$

Therefore, the probability of this event happening

$$P\{\ m^{th}\ station's\ P1\ server\ is\ busy\} = x_m l_1 \qquad (5.22)$$

5.2.1.4 Average Service time of *P1* packets

The above three sections (5.2.1.1, 5.2.1.2 and 5.2.1.3) describe the three possible situations that may happen when a *P1* packet arrives to a station. The probabilities of these situations and the average service time of the packets accompanied with these situations are also derived in the above three sections. A summary of these results are given in Table 5.4.

Table 5.4 Summarized results

		Probability of the Event	Average service time
Events	Station's *P1* server is empty & Channel is empty	$1 - \frac{Ml_1 S}{N_r} - \frac{Ml_2 S_2}{N_{r2}}$	$x_{m1} = F_m$
	Station's *P1* server is empty & Channel is busy	$\frac{Ml_1 S}{N_r} + \frac{Ml_2 S_2}{N_{r2}} - x_m l$	$x_{m2} = R_m + F_m$
	Station's *P1* server is busy	$x_m l_1$	$x_{m3} = S$

From the information in the table, an equation for the average service time of *P1* packets of the m^{th} station, x_m can be given as

$$\begin{aligned} x_m &= \left[1 - \frac{M l_1 S}{N_r} - \frac{M l_2 S_2}{N_{r2}}\right] F_m + \left[\frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}} - x_m l_1\right](R_m + F_m) + x_m l_1 S \\ &= F_m - x_m l_1 F_m + \left[\frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}} - x_m l_1\right] R_m + x_m l_1 S \end{aligned}$$

Then by substituting R_m given in equation (5.21)

$$\begin{aligned} x_m &= F_m - x_m l_1 F_m + \left[\frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}} - x_m l\right] \frac{\left[\frac{M l_1}{N_r} - \frac{x_m l_1}{S}\right]\overline{S^2} + \frac{M l_2}{N_{r2}}\overline{S_2{}^2}}{2\left[\frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}} - x_m l_1\right]} + x_m l_1 S \\ &= F_m - x_m l_1 F_m + \left[\frac{M l_1}{2N_r} - \frac{x_m l_1}{2S}\right]\overline{S^2} + \frac{M l_2}{2N_{r2}}\overline{S_2{}^2} + x_m l_1 S \end{aligned}$$

Hence the average service time of the m^{th} station, x_m is

$$x_m = \frac{F_m + \frac{M l_1 \overline{S^2}}{2N_r} + \frac{M l_2}{2N_{r2}}\overline{S_2{}^2}}{1 + l_1 F_m + \frac{l_1 \overline{S^2}}{2S} - l_1 S} \tag{5.23}$$

When x_m is known, the average service time, x for all the stations can also be found by using equation (5.1)

$$x = \sum_{m=1}^{M} \frac{x_m}{M}$$

where x_m is given in the equation (5.23). As F_m, N_r, $\overline{S^2}$ and S are parameters dependent on x, but N_{r2}, $\overline{S_2{}^2}$ and S_2 are parameters that depend on x_2. Therefore, equation (5.1) cannot be solved as we have only one equation with two unknowns. To solve this equation, *P2* traffic has to be analyzed and this is done in the following sections.

After solving equation (5.1), the corresponding value for x_m given in equation (5.23) is known. When x_m is known, the second moment of the service time of *P1* packets of the m^{th} station, $\overline{x_m^2}$ can be found from table 5.4.

$$\overline{x_m^2} = \left[1 - \frac{Ml_1S}{N_r} - \frac{Ml_2S_2}{N_{r2}}\right]F_m{}^2 + \left[\frac{Ml_1S}{N_r} + \frac{Ml_2S_2}{N_{r2}} - x_m l_1\right](R_m + F_m)^2 + x_m l_1 S^2 \quad (5.24)$$

From the properties of M/G/1 queue [2], the average number of packets in *P1* queue and server of the m^{th} station, N_m can be found:

$$N_m = r_m + \frac{l_1{}^2 \overline{x_m^2}}{2(1 - r_m)} \quad (5.25)$$

where $r_m = l_1 x_m$.

5.2.2 Priority two traffic

Now consider priority two traffic. When a priority two packet arrives to the MAC layer of a station, again there are three situations that may occur:

- Priority two queue is not serving a packet and the channel is free
- Priority two queue is not serving a packet but the channel is busy
- Priority two queue is serving a packet

As the MAC layer for priority two is also modeled as M/G/1 queue, the only parameters that need to be found are the average service time of the *P2* packets of the m^{th} station, x_{m2}, and the second moment of the service time of the *P2* packets of the m^{th} station, $\overline{x_{m2}^2}$. The above three situations are considered here to find those parameters.

5.2.2.1 Priority two queue is not serving a packet and the channel is free

In this situation, when a packet arrives to the priority two queue of the m^{th} station, the station is not serving any packet (station is empty) and the channel is free. As a result, the service time for this packet depends on the number of stations that are going to collide with the m^{th} station with priority two packets, average number of *P1* packets in the *P1* system (queue and server), and average number of *P1* packets arriving during this time. Figure 5.16 shows this scenario.

In the figure, *S* is the time taken to serve a collision of *P1* traffic, *f* is the initial *P2* collision detection time, given by $f = AIFSN_2 + RTS$, $P1_{SA}$ is the time taken to serve the *P1* packets in the system and any newly arrived *P1* packets after *P2* collision has occurred and $F_{m2}-f$ is the time taken to serve the *P2* packet of the m^{th} station after collision resolution of *P2* has been started.

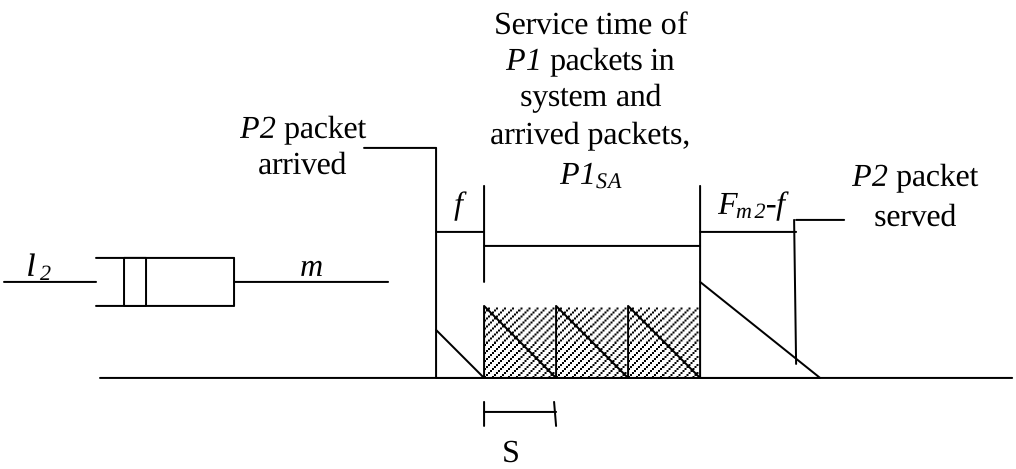


Figure 5.16 When a packet arrives to the priority two queue of a station the channel and the queue is empty

For similar reasons given in the section (5.1.1.3), F_{m2} can be derived as

If N_{r2} £ $M/2$ *and* 1 £ m £ $N_{r2} - 1$

$$F_{m2} = \sum_{y=1}^{m} \frac{C_{y-1}^{m-1} C_{N_{r2}-y}^{M-m}}{C_{N_{r2}-1}^{M-1}} F_{m2}^{y} \tag{5.26-1}$$

Else if N_{r2} £ $M/2$ *and* N_{r2} £ m £ $M-N_{r2} + 1$

$$F_{m2} = \sum_{y=1}^{N_{r2}} \frac{C_{y-1}^{m-1} C_{N_{r2}-y}^{M-m}}{C_{N_{r2}-1}^{M-1}} F_{m2}^{y} \qquad (5.26\text{-}2)$$

Else if N_{r2} *£ M/2 and M-*N_{r2} *+ 2£ m £ M*

$$F_{m2} = \sum_{y=1}^{M-m+1} \frac{C_{N_{r2}+y+m-M-2}^{m-1} C_{M-m-y+1}^{M-m}}{C_{N_{r2}-1}^{M-1}} F_{m2}^{N_{r2}+y+m-M-1} \qquad (5.26\text{-}3)$$

Else if N_{r2} *> M/2 and 1 £ m £ M -* N_{r2} *+ 1*

$$F_{m2} = \sum_{y=1}^{m} \frac{C_{y-1}^{m-1} C_{N_{r2}-y}^{M-m}}{C_{N_{r2}-1}^{M-1}} F_{m2}^{y} \qquad (5.26\text{-}4)$$

Else if N_{r2} *> M/2 and M -* N_{r2} *+ 2 £ m £* N_{r2} *- 1*

$$F_{m2} = \sum_{y=1}^{M-N_{r2}+1} \frac{C_{m-(M-N_{r2}+2)+y}^{m-1} C_{M-m+1-y}^{M-m}}{C_{N_{r2}-1}^{M-1}} F_{m2}^{m-(M-N_{r2}+2)+y+1} \qquad (5.26\text{-}5)$$

Else if N_{r2} *> M/2 and* N_{r2} *£ m £ M*

$$F_{m2} = \sum_{y=1}^{M-m+1} \frac{C_{m-(M-N_{r2}+2)+y}^{m-1} C_{M-m+1-y}^{M-m}}{C_{N_{r2}-1}^{M-1}} F_{m2}^{m-(M-N_{r2}+2)+y+1} \qquad (5.26\text{-}6)$$

where F_{m2}^{y} is defined as:

If N_{r2} stations are collided with *P2* traffic, and the *P2* packet of the *m*th station is served in the *y*th position ($1 \le y \le N_{r2}$) then the time taken to serve the station, F_{m2}^{y} is;

If $N_{r2} = 1$

$$F_{m2}^{y} = AIFSN_2 + 3 * SIFS + (RTS + CTS + D2 + MACHeader + PHYHeader + ACK)/C$$

Else if $N_{r2} > 1$ *and* $y = 1$

$$F_{m2}^{y} = AIFSN_2 + RTS/C + AIFSC_2 + CRB + CRIFS * M + PPB * N_{r2} + NPB * (M - N_{r2}) + ((D2 + MACHeader + PHYHeader + ACK)/C + SIFS + SDIFS)$$

Else if $N_{r2} > 1$ *and* $y > 1$

$$F_{m2}^{y} = AIFSN_2 + RTS/C + AIFSC_2 + CRB + CRIFS * M + PPB * N_{r2} + NPB * (M - N_{r2})$$
$$+ ((D2 + MACHeader + PHYHeader + ACK)/C + SIFS + SDIFS)$$
$$+ ((D2 + MACHeader + PHYHeader + ACK + TP + TR)/C + 3 * SIFS + SDIFS) * (y - 1)$$

where C is the channel rate, $D2$ is the data payload of P2 packets and other parameters are specified in the MAC scheme.

5.2.2.1.1 Calculating the probability of this situation

If $P2$ arrival rates for the m^{th} station is l_2 packets per second, then over a given time, t, the amount of busy time of the $P2$ server of the m^{th} station is $STA_{m2\,busy} = x_{m2} l_2 t$ seconds, where x_{m2} is the average service time of the $P2$ packets of the m^{th} station. The Venn Diagram [Stark and Woods 2002] in Figure 5.17 shows the amount of channel time spent in different areas.

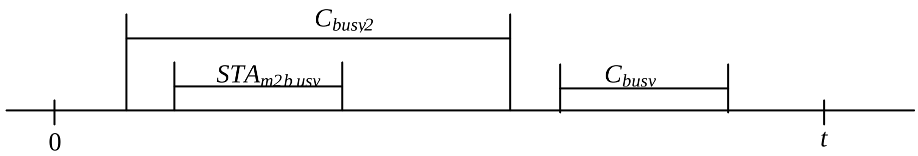


Figure 5.17 Status of the channel during a time interval t.

In the figure, C_{busy} is given in equation (5.6) and C_{busy2} is given in equation (5.18). In this situation, when a packet arrives to the priority two queue of the m^{th} station, the station is not serving any packet and the channel is free. From Figure 5.17,

$$P\{\ m^{th} \text{ station's } P2 \text{ server is empty and the channel is free}\} = P\{\text{the channel is free}\}$$
$$P\{\text{the channel is free}\} = 1 - P\{\text{the channel is busy}\} = p_{C_{free2}}$$

Hence the probability that the channel is free, $p_{C_{free2}}$ is the probability of this event happening and is given in equation (5.19).

$$p_{C_{free2}} = 1 - \frac{M l S}{N_r} - \frac{M l_2 S_2}{N_{r2}}$$

5.2.2.1.2 Average service time of this event

As shown in Figure 5.16, $P1_{SA}$ is the time taken to serve $P1$ packets in the system and newly arrived $P1$ packets after a $P2$ collision has occurred. The average number of $P1$ packets in $P1$ queue and the server of a station, N_{S1} can be given as:

$$N_{S1} = \sum_{m=1}^{M} \frac{N_m}{M} \tag{5.27}$$

where N_m is given in equation (5.25). Then the number of $P1$ collisions that need to be served due to packets in the $P1$ queue and server of the stations is $\frac{N_{S1}M}{N_r}$, where N_r is the average number of stations collided with $P1$ packets.

The number of $P1$ collisions that need to be served due to newly arrived $P1$ traffic is $\frac{M l_1 (P1_{SA} + f)}{N_r}$. From this information and with the help of Figure 5.16, $P1_{SA}$ can be given as:

$$P1_{SA} = \frac{N_{S1} MS}{N_r} + \frac{M l_1 (P1_{SA} + f) S}{N_r}$$

$$\therefore P1_{SA} = \frac{\frac{N_{S1} MS}{N_r} + \frac{M l_1 Sf}{N_r}}{1 - \frac{M l_1 S}{N_r}} = \frac{N_{S1} MS + M l_1 Sf}{N_r - M l_1 S} \tag{5.28}$$

Hence from Figure 5.16, the average service time of this event, x_{m2-1} can be given as

$$x_{m2-1} = f + P1_{SA} + F_{m2} - f = F_{m2} + \frac{N_{S1} MS + M l_1 Sf}{N_r - M l_1 S} \tag{5.29}$$

5.2.2.2 Priority two queue is not serving a packet but the channel is busy

In this situation, when a packet arrives to the priority two queue of the m^{th} station, *P2* server of the m^{th} station is not serving any packets but the channel is busy. The channel may be busy due to either of *P1* or *P2* traffic. As a result, the service time for this packet depends on the amount of time remaining to complete the on-going collision resolution, R_{m2}, the number of stations that are going to collide with the m^{th} station with *P2* packets, N_{r2}, average number of *P1* packets in the *P1* queue and server, N_{S1}, and average number of *P1* packets arriving during this time. Figure 5.18 shows this scenario.

In the figure, S is the time taken to serve a collision of *P1* traffic, f is the initial *P2* collision detection time, and $f = AIFSN_2 + RTS$, $P1_{SA}$ is the time taken to serve the *P1* packets in the system and newly arrived *P1* packets after *P2* collision has occurred and $F_{m2}-f$ is the time taken to serve the *P2* packet of the m^{th} station after collision resolution of *P2* has started.

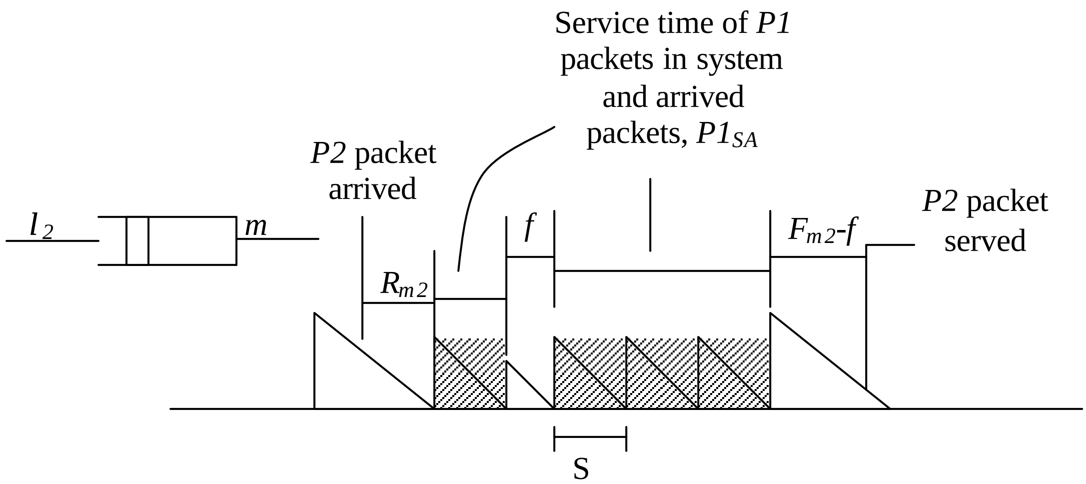


Figure 5.18 When a packet arrives to the P2 queue of a station, the P2 server is empty but the channel is busy

5.2.2.2.1 Time remaining to complete the ongoing collision resolution, R_{m2}

As before, R_{m2} can be calculated by using graphical augment. Figure 5.19 shows the residual service time $r(t)$ (i.e., the remaining time for completion of service for collided stations at time t) as a function of t.

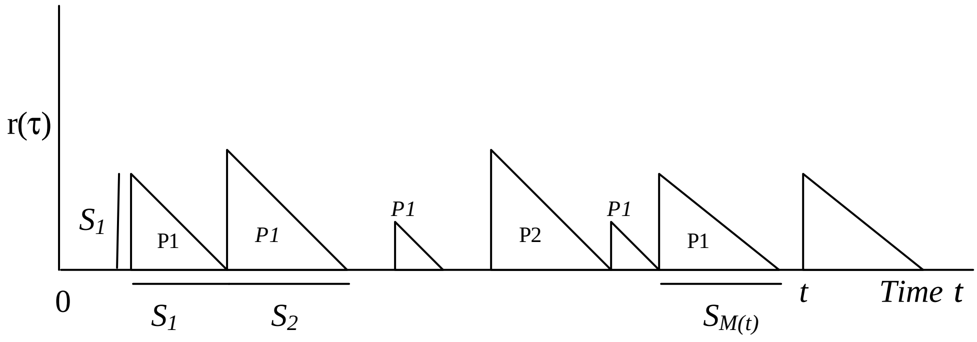


Figure 5.19 Residual service time

However in this situation we are only interested in the time when the $P2$ server of the m^{th} station is empty but the channel is busy. From figure 5.17, within a given time t, the amount of time that the $P2$ server of the m^{th} station is empty but the channel is busy, $t_2^{'} = \frac{M l_1 S t}{N_r} + \frac{M l_2 S_2 t}{N_{r2}} - x_{m2} l_2 t$.

Figure 5.19 can be changed to suit for this situation. The new figure for this situation is shown in figure 5.20.

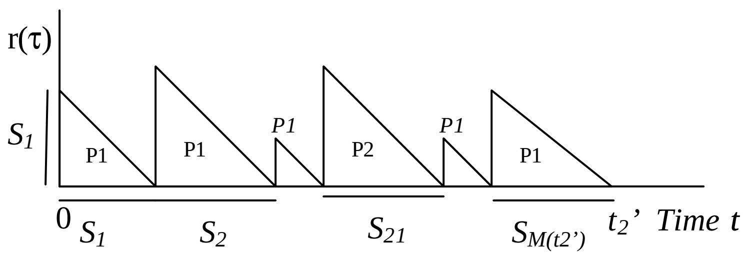


Figure 5.20 Residual service time for this situation

The total number of $P1$ packets arriving in time t is $M l_1 t$ and it is known that on average N_r, $P1$ packets are served in S seconds of channel busy time. Hence the number of $P1$ packets that need to be serviced within time t_2' is $M l_1 t$. So

$$\textit{Average number of } P1 \textit{ collisions need to be serviced in } t_2^{'} \textit{ time}, M(t_2^{'})$$
$$= \frac{M l_1 t}{N_r}$$

Similarly, the total number of $P2$ packets arriving in time, t is $M l_2 t$ and it is known that on average N_{r2}, $P2$ packets are served in S_2 seconds of channel busy time. The total amount of the $P2$ server busy time of the m^{th} station is $x_{m2} l_2 t$ at a given time, t, where x_{m2} is the average service time of *P2* packets of the m^{th} station. From this information, the number of packets served during the *P2* server busy time of the m^{th} station is $N_{r2}x_{m2}l_2t/S_2$. Hence the number of *P2* packets that need to be served in t_2' is $(Ml_2 - N_{r2}x_{m2}l_2t/S_2)t$. Therefore,

$$
\begin{aligned}
&Average\,number\ of\ P2\,collisions\ need\ to\,be\,serviced\ in\ t_2'\ time, M(t_2') \\
&= \frac{(Ml_2 - N_{r2}x_{m2}l_2/S_2)t}{N_{r2}} \\
&= (Ml_2/N_{r2} - x_{m2}l_2/S_2)t
\end{aligned}
$$

Now from Figure 5.20, consider a time t_2' for which $r(t_2') = 0$. The time average of $r(t)$ in the interval $[0, t_2']$ is

$$\frac{1}{t_2'}\int_0^{t_2'} r(t)dt = \frac{1}{t_2'}\sum_{i=1}^{M(t_2')}\frac{1}{2}S_i^2 + \frac{1}{t_2'}\sum_{i=1}^{M2(t_2')}\frac{1}{2}S_{2i}^2$$

where $M(t_2')$ is the number of *P1* collisions served within $[0, t_2']$ and $M2(t_2')$ is the number of *P2* collisions served within $[0, t_2']$, S_i is the service time of the i^{th} collided *P1* packets and S_{2i} is the service time of the i^{th} collided *P2* packets. The equation can also be written as

$$\frac{1}{t_2'}\int_0^{t_2'} r(t)dt = \frac{M(t_2')}{t_2'}\frac{\sum_{i=1}^{M(t_2')}\frac{1}{2}S_i^2}{M(t_2')} + \frac{M2(t_2')}{t_2'}\frac{\sum_{i=1}^{M2(t_2')}\frac{1}{2}S_{2i}^2}{M2(t_2')}$$

Assuming the limit below exists, we obtain

$$\lim_{t_2'\to\infty}\frac{1}{t_2'}\int_0^{t_2'} r(t)dt = \frac{1}{2}\lim_{t_2'\to\infty}\frac{M(t_2')}{t_2'}\lim_{t_2'\to\infty}\frac{\sum_{i=1}^{M(t_2')}S_i^2}{M(t_2')} + \frac{1}{2}\lim_{t_2'\to\infty}\frac{M2(t_2')}{t_2'}\lim_{t_2'\to\infty}\frac{\sum_{i=1}^{M2(t_2')}S_{2i}^2}{M2(t_2')}$$

Assuming that time averages can be replaced by ensemble averages, $\lim_{t_2^{'}\to\infty}\frac{\sum_{i=1}^{M(t_2^{'})} S_i^2}{M(t_2^{'})}$ is the second moment of the *P1* collision service time, $\overline{S^2}$, and $\lim_{t_2^{'}\to\infty}\frac{\sum_{i=1}^{M2(t_2^{'})} S_{2i}^2}{M2(t_2^{'})}$ is the second moment of the *P2* collision service time, $\overline{S_2^{\,2}}$. Thus the equation becomes

$$\lim_{t_2^{'}\to\infty}\frac{1}{t_2^{'}}\int_0^{t_2^{'}} r(t)\,dt$$
$$=\frac{1}{2}\frac{\frac{Ml_1t}{N_r}}{\frac{Ml_1St}{N_r}+\frac{Ml_2S_2t}{N_{r2}}-x_{m2}l_2t}\overline{S^2}+\frac{1}{2}\frac{\left(Ml_2/N_{r2}-x_{m2}l_2/S_2\right)t}{\frac{Ml_1St}{N_r}+\frac{Ml_2S_2t}{N_{r2}}-x_{m2}l_2t}\overline{S_2^{\,2}}$$
$$=\frac{\frac{Ml_1}{N_r}\overline{S^2}+\left[\frac{Ml_2}{N_{r2}}-\frac{x_{m2}l_2}{S_2}\right]\overline{S_2^{\,2}}}{2\left[\frac{Ml_1S}{N_r}+\frac{Ml_2S_2}{N_{r2}}-x_{m2}l_2\right]}$$

Hence time remaining to complete the on going collision resolution, R_{m2}

$$R_{m2}=\frac{\frac{Ml_1}{N_r}\overline{S^2}+\left[\frac{Ml_2}{N_{r2}}-\frac{x_{m2}l_2}{S_2}\right]\overline{S_2^{\,2}}}{2\left[\frac{Ml_1S}{N_r}+\frac{Ml_2S_2}{N_{r2}}-x_{m2}l_2\right]} \qquad (5.30)$$

5.2.2.2.2 Calculating the probability of this situation

In this situation, when a packet arrives to the priority two queue of the m^{th} station, *P2* server of the m^{th} station is not serving any packet but the channel is busy. From Figure 5.17,

$$P\{\ m^{th}\ \text{station's P2 server is empty and the channel is busy}\}$$
$$=\ P\{\text{the channel is busy}\}\ P\{\ m^{th}\ \text{station's P2 server is empty}\,|\,\text{the channel is busy}\}$$

From section 5.2.2.1.1, the amount of busy time of $P2$ server of the m^{th} station is $STA_{m2\ busy}= x_{m2} l_2 t$ seconds. Where x_{m2} is the average service time of $P2$ packets of the m^{th} station, and l_2 is the average packet arrival rate of $P2$. Hence from this:

$$P\{\ m^{th}\ station's\ P2\ server\ busy\ \} = x_{m2} l_2$$

$$P\{the\ channel\ is\ busy\ \} = \frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}}$$

$$P\{\ m^{th}\ station's\ P2\ server\ is\ empty | the\ channel\ is\ busy\ \} = \frac{\frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}} - x_{m2} l_2}{\frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}}}$$

Therefore, probability of this event happening

$$P\{\ m^{th}\ station's\ P2\ server\ is\ empty\ and\ the\ channel\ is\ busy\}$$
$$= \frac{M l_1 S}{N_r} + \frac{M l_2 S_2}{N_{r2}} - x_{m2} l_2 \qquad (5.31)$$

5.2.2.2.3 Average service time of this event

As shown in Figure 2.18, $P1_{SA}$ is the time taken to serve $P1$ packets in the system and newly arrived $P1$ packets after the $P2$ packet arrival to the m^{th} station. The number of $P1$ collision that need to be served due to $P1$ packets in queues and servers of the stations is $\frac{N_{S1} M}{N_r}$, where N_{S1} is given in equation (5.27) and number of $P1$ collision that need to be served due to newly arrived $P1$ traffics is $\frac{M l_1 (P1_{SA} + f + R_{m2})}{N_r}$. From this information and with the help of Figure 5.18, $P1_{SA}$ can be given as:

$$P1_{SA} = \frac{N_{S1} MS}{N_r} + \frac{M l_1 (P1_{SA} + f + R_{m2}) S}{N_r}$$

$$\therefore P1_{SA} = \frac{\frac{N_{S1} MS}{N_r} + \frac{M l_1 Sf}{N_r} + \frac{M l_1 S R_{m2}}{N_r}}{1 - \frac{M l_1 S}{N_r}} = \frac{N_{S1} MS + M l_1 Sf + M l_1 S R_{m2}}{N_r - M l_1 S} \qquad (5.32)$$

Hence from Figure 2.18, the average service time of this event, $x_{m2\text{-}2}$ can be given as:

$$x_{m2-2} = R_{m2} + f + P1_{SA} + F_{m2} - f = R_{m2} + F_{m2} + \frac{N_{S1}MS + Ml_1Sf + Ml_1SR_{m2}}{N_r - Ml_1S} \tag{5.33}$$

5.2.2.3 Priority two queue is serving a packet

In this situation, when a packet arrives to the *P2* queue of the m^{th} station, the server is serving a packet so the channel must be busy. Figure 5.21 shows this scenario.

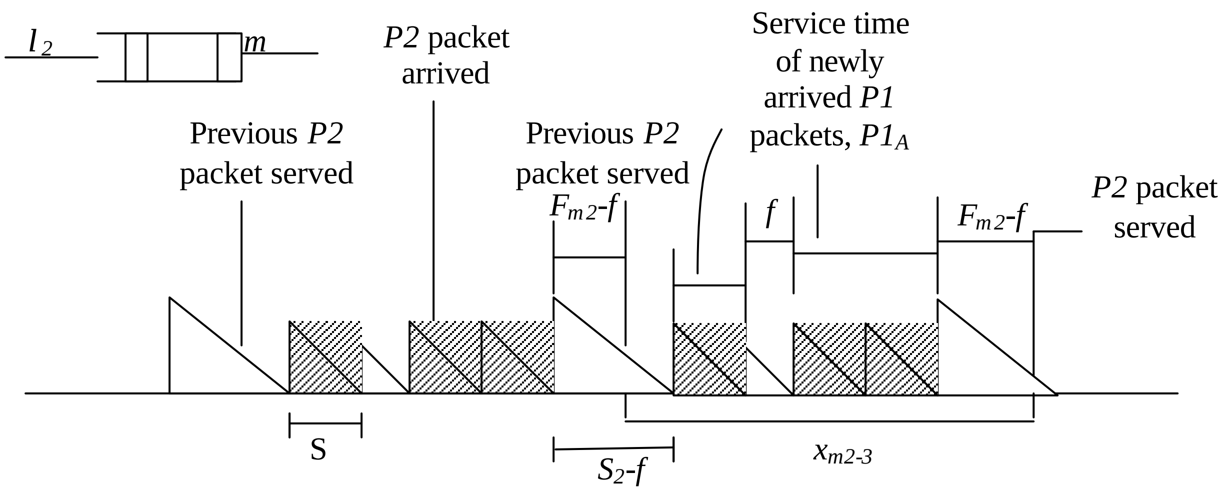


Figure 5.21 When a packet arrives to P2 queue of a station the P2 server is busy

In the figure, *S* is the time taken to serve a collision of *P1* traffic, *f* is the initial *P2* collision detection time, and $f = AIFS2N + RTS$, $P1_A$ is the time taken to serve newly arrived *P1* packets after *P2* collision has occurred, F_{m2}–f is the time taken to serve the *P2* packet of the m^{th} station after collision resolution of *P2* has been started and S_2–f is the time taken to serve P2 collisions neglecting the initial collision detection time *f*.

As shown in the figure, in this situation, the time taken to service the arrived *P2* packet is x_{m2-3} seconds. A *P2* packet cannot be served when there are *P1* packets in the queues and servers. After previous *P2* packets have been served, the only packets waiting to be served are the newly arrived *P1* packets.

As shown in Figure 5.21, $P1_A$ is the time taken to serve the newly arrived *P1* packets. The number of *P1* collisions that need to be served due to the newly arrived P1 traffic is $\frac{Ml_1(P1_A + S_2)}{N_r}$. From this information and help of Figure 5.21, $P1_A$ can be obtained:

$$P1_A = \frac{Ml_1(P1_A + S_2)S}{N_r}$$

$$\therefore P1_A = \frac{\frac{Ml_1SS_2}{N_r}}{1 - \frac{Ml_1S}{N_r}} = \frac{Ml_1SS_2}{N_r - Ml_1S} \tag{5.34}$$

Hence, from Figure 5.21, the average service time of this event, $x_{m2\text{-}3}$ is given by:

$$x_{m2-3} = S_2 - F_{m2} + f + P1_A + F_{m2} - f = S_2 + \frac{Ml_1SS_2}{N_r - Ml_1S} = \frac{N_rS_2}{N_r - Ml_1S} \tag{5.35}$$

5.2.2.3.1 Finding the probability of this situation

In this situation, when a packet arrives to *P2* queue of the m^{th} station, the server is serving a packet. From Figure 5.17,

$$P\{\ m^{th}\ station'\ s\ P2\ server\ is\ busy\}$$
$$= \frac{x_{m2}\, l_2 t}{t} = x_{m2}\, l_2$$

Therefore, the probability of this event happening

$$P\{\ m^{th}\ station'\ s\ P2\ server\ is\ busy\} = x_{m2}\, l_2 \tag{5.36}$$

5.2.2.4 Average Service time of *P2* packets

The above three sections (5.2.2.1, 5.2.2.2 and 5.2.2.3) describe the three possible situations that may occur when a *P2* packet arrives to a station. The probabilities of these events and average service time of the packets accompanied with these events are also derived in the above three sections. A summary of these results is given in Table 5.5.

Table 5.5 Summarized results

		Probability of the Event	Average service time
Events	Station's *P2* server is empty & Channel is empty	$1 - \frac{Ml_1S}{N_r} - \frac{Ml_2S_2}{N_{r2}}$	x_{m2-1}
	Station's *P2* server is empty & Channel is busy	$\frac{Ml_1S}{N_r} + \frac{Ml_2S_2}{N_{r2}} - x_{m2}l_2$	x_{m2-2}
	Station's *P2* server is busy	$x_{m2}l_2$	x_{m2-3}

From the information in the table, an equation for the average service time of *P2* packets of the m^{th} station, x_{m2} can be obtained as

$$x_{m2} = \left[1 - \frac{Ml_1S}{N_r} - \frac{Ml_2S_2}{N_{r2}}\right]x_{m2-1} + \left[\frac{Ml_1S}{N_r} + \frac{Ml_2S_2}{N_{r2}} - x_{m2}l_2\right]x_{m2-2} + x_{m2}l_2x_{m2-3}$$

Then by substituting, R_{m2}, x_{m2-1}, x_{m2-2}, and x_{m2-3}, given in equation (5.30), (5.29), (5.33) and (5.35), respectively, an expression for x_{m2} is obtained as

$$x_{m2} = \frac{\left[F_{m2} - f\left[\frac{Ml_1S}{N_r} + \frac{Ml_2S_2}{N_{r2}}\right] + \left[\frac{\left[1 - \frac{Ml_1S}{N_r} - \frac{Ml_2S_2}{N_{r2}}\right]\frac{Ml_1Sf}{N_r} + \frac{Ml_2\overline{S_2^2}}{2N_{r2}} + \frac{Ml_1}{2N_r}\overline{S^2} + \frac{N_{S1}MS}{N_r} + \left[\frac{Ml_1S}{N_r} + \frac{Ml_2S_2}{N_{r2}}\right]f}{\left[1 - \frac{Ml_1S}{N_r}\right]}\right]\right]}{\left[1 - fl_2 + l_2F_{m2} + \left[\frac{\frac{l_2\overline{S_2^2}}{2S_2} + \frac{N_{S1}MSl_2}{N_r} + l_2f - l_2S_2}{\left[1 - \frac{Ml_1S}{N_r}\right]}\right]\right]} \tag{5.37}$$

When x_{m2} is known the average service time of *P2* packets, x_2 for all the station can be found by using equation (5.17)

$$x_2 = \sum_{m=1}^{M} \frac{x_{m2}}{M}$$

where x_{m2} is given in equation (5.37). Now the equation (5.17) and (5.1) {$x = \sum_{m=1}^{M} \frac{x_m}{M}$, where x_m is given in equation (5.23)} can be solved for x and x_2. When these two equations are solved , the corresponding values of x_{m2} given in equation (5.37) can be found. Then the second moment of the service time of $P2$ packets of the m^{th} station, $\overline{x^2_{m2}}$, can be found from table 5.5.

$$\overline{x^2_{m2}} = \left[1 - \frac{Ml_1S}{N_r} - \frac{Ml_2S_2}{N_{r2}}\right] x_{m2-1}x_{m2-1} + \left[\frac{Ml_1S}{N_r} + \frac{Ml_2S_2}{N_{r2}} - x_{m2}l_2\right] x_{m2-2}x_{m2-2} + x_{m2}l_2x_{m2-3}x_{m2-3} \tag{5.38}$$

5.2.2.5 Average Delay of the P1 and P2 packets

After solving the equations, the corresponding values for x_m, $\overline{x^2_m}$, x_{m2} and $\overline{x^2_{m2}}$ can be calculated from the equations given. Then from the properties of the M/G/1 queue [Bertsekas and Gallager 1992], we can now obtain the average delay of $P1$ packets of the m^{th} station, T_m and the average delay of $P2$ packets of the m^{th} station, T_{m2}:

$$T_m = x_m + \frac{lx^2_m}{2(1 - r_m)}$$

$$T_{m2} = x_{m2} + \frac{lx^2_{m2}}{2(1 - r_{m2})}$$

where $r_m = l_1x_m$ and $r_{m2} = l_2x_{m2}$. As the average arrival rate of packets for all the stations are equal, on average the total number of packets being served by each station must be equal. So the average delay of the packets, T_1 and T_2 can be given as:

$$T_1 = \sum_{m=1}^{M} \frac{T_m}{M}$$

$$T_2 = \sum_{m=1}^{M} \frac{T_{m2}}{M}$$

The analytical results obtained for the average delay of the packets, T_1 and T_2 are shown in Figure 5.22.

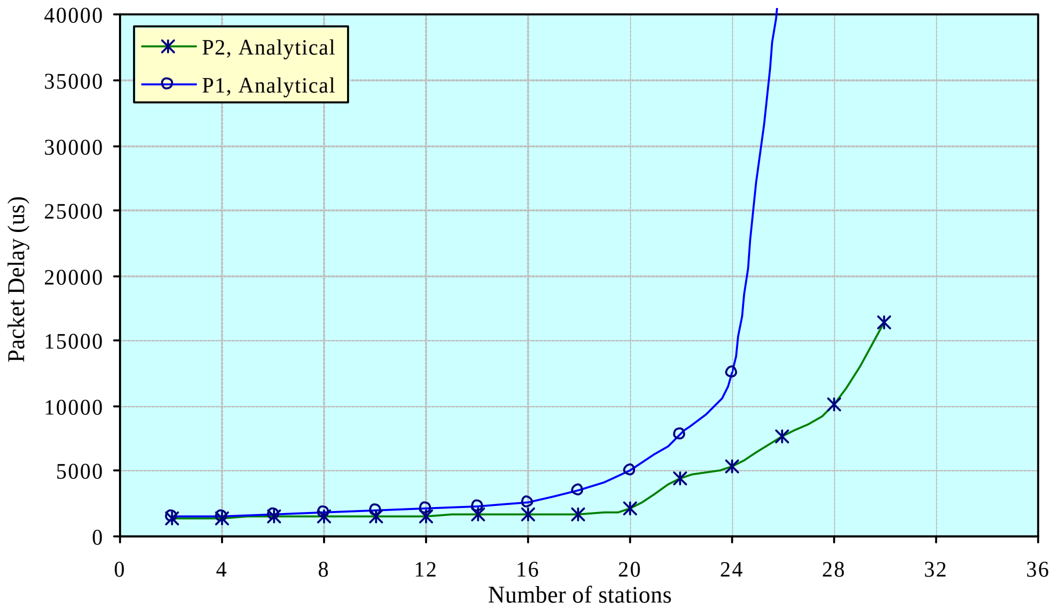


Figure 5.22 Average delay of priority one and two packets

The arrival rates are $l_1 = 12.5$ packets per second and $l_2 = 12.5$ packets per second. From the figure, it is seen that priority one packets have less delay compared to priority two packets. This shows that the proposed MAC scheme supports service differentiation. This analytical model is only suitable when the channel is not fully occupied. When the number of stations goes to 32, the channel is saturated by P2 traffic (i.e. P2 has more traffic than the channel can serve) so the analytical model does not give a reasonable answer.

To find the performance for *P1* traffic after the channel is saturated by *P2* data, this situation has to be considered separately, which is done in section 5.3.

5.3 Stations have priority one and two traffic and the channel is fully occupied

In this case, both priority one and two traffic are considered. However, *P2* traffic arrives faster than the channel can serve. As a result, *P2* traffic at all stations always has non-empty transmission queues and the time that is left in the channel after serving *P1* traffic is used by *P2* traffic. So the channel will always be busy in service and there is no free time left. The MAC scheme can be modeled as shown in Figure 5.9. For *P1* traffic, it has been assumed that traffic arrive to the stations according to the Poisson process, traffic service times have a general distribution and have a queue with an infinite buffer space. With the help of these assumptions, the MAC layer for *P1* traffic of a station is represented by an M/G/1 queue queue [Bertsekas and Gallager 1992], [Kleinrock 1975], [Kleinrock 1976], [Kleinrock and Gail 1996].

As shown in Figure 5.9, traffic arrives to priority one and priority two queues with an average rates of l_1 and l_2 packets per second, respectively. Traffic serving rates m_1 and m_2 are the average service rates with respect to a station for priority one and priority two traffic, respectively.

In this situation, *P2* packets arrive faster than the channel can serve so the service time of *P2* packets will not be bounded so we need to find only *P1* service time, x. When a priority one packet arrives to the MAC layer of a station, the channel will always be busy due to P2 traffic saturation. So there are only two situations that may arise:

- Priority one queue is not serving a packet but the channel is busy
- Priority one queue is serving a packet

As at each station P1 traffic is modeled as an M/G/1 queue, the only parameters that need to be found are the average service time of *P1* packets of the m^{th} station, x_m, and the second moment of the service time of *P1* packets of the m^{th} station, $\overline{x_m^2}$. To find these parameters, those two situations are considered.

5.3.1 Priority one queue is not serving a packet but the channel is busy

In this situation, when a packet arrives to the priority one queue of the m^{th} station, the server of the queue is empty but the channel is busy. As a result, the service time for this packet depends on the amount of time remaining to serve previously collided stations with *P1* traffic or the amount of time remaining to serve previously collided stations with *P2* traffic and also the number of stations that are going to collide with the m^{th} station with *P1* traffic. Figure 5.23 shows this scenario.

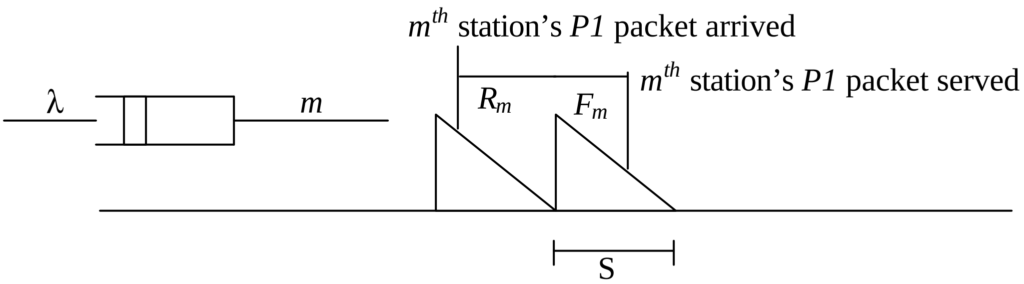


Figure 5.23 When a P1 packet arrives to a station the channel is busy but the station's P1 server is empty

In the figure, S represents the average time taken to serve all the collided stations with $P1$ traffic in a given collision and F_m is the time taken to serve the $P1$ packet of the m^{th} station and R_m is the time remaining to serve previous collided stations either with $P1$ or $P2$ traffic. The average service time of this event, $x_{m2} = R_m + F_m$, where F_m is given in equation (5.8) and R_m needs to be found.

5.3.1.1 Finding the probability of this situation

At a given time, t, the amount of time the channel is busy due to priority one traffic is given in equation (5.6).

$$C_{busy} = \frac{M l_1 S t}{N_r}$$

And all the remaining time is occupied by P2 traffic, so the total time that the channel is busy due to priority two traffic, C_{busy2} can be given as:

$$C_{busy2} = \left[1 - \frac{M l_1 S}{N_r}\right] t \qquad (5.39)$$

The amount of busy time of $P1$ server of the m^{th} station is $STA_{m\ busy} = x_m l_1 t$ seconds. This information is shown in Figure 5.24.

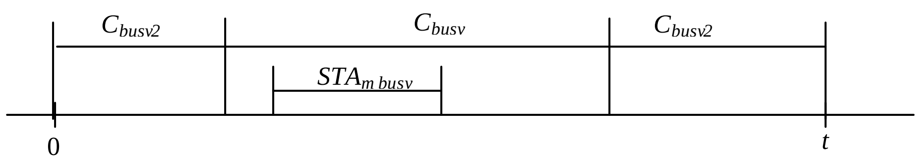


Figure 5.24 Status of the channel during a time interval *t*.

As mentioned before, in this situation when a packet arrives to *P1* queue of the m^{th} station, the station is not serving any packet but the channel is busy. From the figure,

$$\begin{aligned}&P\{\ m^{th}\ station's\ P1\,server\ is\ empty\,and\ the\,channel\ is\ busy\}\\&=\ P\{the\ channel\ is\,busy\}\ P\{\ m^{th}\ station's\ P1\ server\ is\ empty\,|\,the\,channel\ is\ busy\}\end{aligned}$$

And

$$\begin{aligned}&P\{\ m^{th}\ station's\ P1\,server\,busy\ \} = x_m\, l_1\\&P\{the\ channel\ is\,busy\ \} = 1\\&P\{\ m^{th}\ station's\ P1\,server\ is\ empty\,|\,the\,channel\ is\ busy\ \} = 1 - x_m\, l_1\end{aligned}$$

Therefore, probability of this event happening

$$\begin{aligned}&P\{\ m^{th}\ station's\ P1\,server\ is\,empty\ and\ the\ channel\ is\ busy\}\\&= 1 - x_m\, l_1\end{aligned} \tag{5.40}$$

5.3.1.2 Time remaining to complete the ongoing collision resolution, R_m

As before, R_m can be calculated by using graphical augment as in [2]. Figure 5.25 shows the residual service time, $r(t)$ (i.e., the remaining time for completion of service for collided stations at time t) as a function of t.

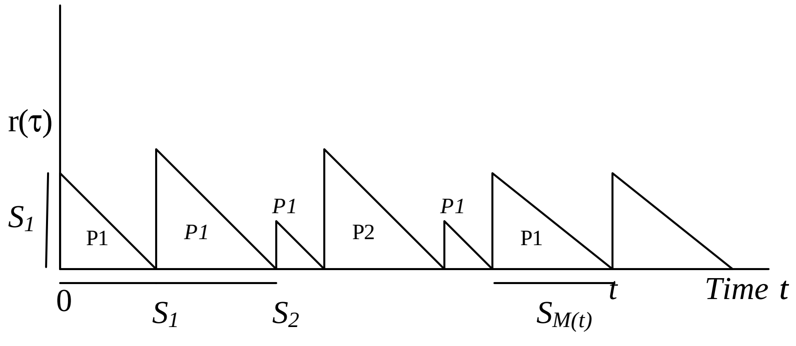


Figure 5.25 Residual service time

However in this situation we are only interested in the time when *P1* server of the m^{th} station is empty but the channel is busy. From Figure 5.24, within a given time interval, *t*, the amount of time that the *P2* server of the m^{th} station is empty but the channel is busy, $t' = t - x_m l_1 t$. Figure 5.25 can be changed to suit for this situation. The new figure for this situation is shown in Figure 5.26

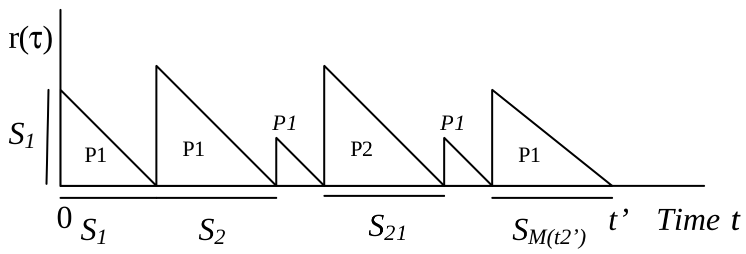


Figure 5.26 Residual service time for this situation

The total number of *P1* packets arriving within time, *t* is Ml_1t, and it is known that on average N_r, priority one packets are served in *S* seconds of channel busy time. Hence the number of *P1* packets that need to be served in *t'* is $Ml_1t - \frac{x_m l_1 N_r t}{S}$, where $\frac{x_m l_1 N_r t}{S}$ is the number of *P1* packets served during *P1* server of m^{th} station busy time. So

$$\textit{Average number of P1 collisions need to be serviced in } t' \textit{ time}, M(t')$$
$$= \frac{Ml_1t}{N_r} - \frac{x_m l_1 t}{S}$$

Similarly, the total number of *P2* packets arriving in time *t* is Ml_2t and it is known that on average N_{r2}, *P2* packets are served in S_2 seconds of channel busy time. But all the stations have always non-empty P2 transmission queue, therefore $N_{r2} = M$. The total time available for *P2* traffic is $\left[1 - \frac{Ml_1S}{N_r}\right]t$. Hence;

$$\textit{Average number of P2 collisions serviced in } t' \textit{ time}, M2(t')$$
$$= \left[\frac{1}{S_2} - \frac{Ml_1S}{S_2N_r}\right]t$$

Now from Figure 5.26, consider a time *t'* for which $r(t') = 0$. The time average of $r(t)$ in the interval [0, t'] is

$$\frac{1}{t'}\int_0^{t'} r(t)\,dt = \frac{1}{t'}\sum_{i=1}^{M(t')}\frac{1}{2}S_i^2 + \frac{1}{t'}\sum_{i=1}^{M2(t')}\frac{1}{2}S_{2i}^2$$

where $M(t')$ is the number of $P1$ collisions served within [0, t'] and $M2(t')$ is the number of $P2$ collisions served within [0, t'], S_i is the service time of the i^{th} collided $P1$ packets and S_{2i} is the service time of the i^{th} collided $P2$ packets. The equation can also be written as

$$\frac{1}{t'}\int_0^{t'} r(t)\,dt = \frac{M(t')}{t'}\frac{\sum_{i=1}^{M(t')}\frac{1}{2}S_i^2}{M(t')} + \frac{M2(t')}{t'}\frac{\sum_{i=1}^{M2(t')}\frac{1}{2}S_{2i}^2}{M2(t')}$$

Assuming the limit below exists, we obtain

$$\lim_{t'\to\infty}\frac{1}{t'}\int_0^{t'} r(t)\,dt = \frac{1}{2}\lim_{t'\to\infty}\frac{M(t')}{t'}\lim_{t'\to\infty}\frac{\sum_{i=1}^{M(t')}S_i^2}{M(t')} + \frac{1}{2}\lim_{t'\to\infty}\frac{M2(t')}{t'}\lim_{t'\to\infty}\frac{\sum_{i=1}^{M2(t')}S_{2i}^2}{M2(t')}$$

Assuming that time averages can be replaced by ensemble averages, $\lim_{t'\to\infty}\frac{\sum_{i=1}^{M(t')}S_i^2}{M(t')}$ is the second moment of the $P1$ collision service time, $\overline{S^2}$, and $\lim_{t'\to\infty}\frac{\sum_{i=1}^{M2(t')}S_{2i}^2}{M2(t')}$ is the second moment of the $P2$ collision service time, $\overline{S_2^2}$. Thus the equation becomes

$$\begin{aligned}
&\lim_{t'\to\infty}\frac{1}{t'}\int_0^{t'} r(t)\,dt \\
&= \frac{1}{2}\frac{\frac{Ml_1t}{N_r} - \frac{x_m l_1 t}{S}}{t - x_m l_1 t}\overline{S^2} + \frac{1}{2}\frac{\left[\frac{1}{S_2} - \frac{Ml_1S}{S_2N_r}\right]t}{t - x_m l_1 t}\overline{S_2^{\,2}} \\
&= \frac{\left[\frac{Ml_1}{N_r} - \frac{x_m l_1}{S}\right]\overline{S^2} + \left[\frac{1}{S_2} - \frac{Ml_1S}{S_2N_r}\right]\overline{S_2^{\,2}}}{2\left[1 - x_m l_1\right]}
\end{aligned}$$

Hence, the time remaining to complete the ongoing collision resolution, R_m

$$R_m = \frac{\left[\frac{Ml_1}{N_r} - \frac{x_m l_1}{S}\right]\overline{S^2} + \left[\frac{1}{S_2} - \frac{Ml_1 S}{S_2 N_r}\right]\overline{S_2^{\,2}}}{2[1 - x_m l_1]} \tag{5.41}$$

The average service time of this event, $x_{m1} = R_m + F_m$, where F_m is given in equation (5.8) and R_m is given in equation (5.41).

5.3.2 Priority one queue is serving a packet

In this situation, when a packet arrives to *P1* queue of the m^{th} station, the *P1* server is serving a packet so the channel must be busy. Figure 5.27 shows this scenario.

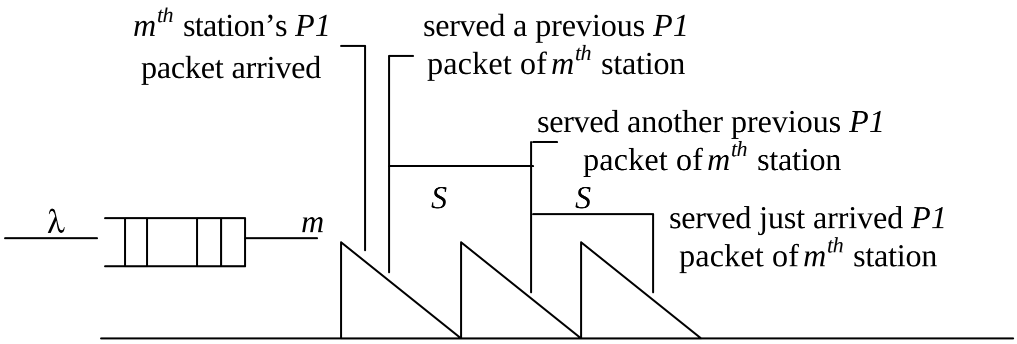


Figure 5.27 When a packet arrives to a station the station is busy

As shown in the figure, in this situation, the time taken to service the arrived *P1* packet is S seconds. Therefore the average service time of this event, $x_{m2} = S$, where S is the time taken to serve all the collided stations with *P1* traffic. S is given in section (5.1.3).

5.3.2.1 Finding the probability of this situation

In this event, when a packet arrives to *P1* queue of the m^{th} station, the *P1* server is serving a packet. From Figure 5.24,

$$P\{\, m^{th} \text{ station's } P1 \text{ server is busy}\}$$

$$= \frac{x_m l_1 t}{t} = x_m l_1$$

Therefore, probability of this event happening

$$P\{\ m^{th}\ station's\ P1\ server\ is\ busy\} = x_m l_1 \tag{5.42}$$

5.3.3 Average Service time of a station

The above two sections (5.3.1 and 5.3.2) describe the two possible situations that may happen when a *P1* packet arrives to a station. The probabilities of these events and average service time of the packets accompanied with these events are also derived in the above two sections. The summary of these are given in Table 5.6.

Table 5.6 Summarized results of the above three sections

		Probability of the Event	Average service time
Events	Station's *P1* server is empty but Channel is busy	$1 - x_m l_1$	$x_{m1} = R_m + F_m$
	Station Station's *P1* server is busy	$x_m l_1$	$x_{m2} = S$

From the information in the table, an equation for the average service time of m^{th} station, x_m can be given as:

$$\begin{aligned} &x_m \\ &= [1 - x_m l_1](R_m + F_m) + x_m l_1 S \\ &= F_m - x_m l_1 F_m + [1 - x_m l_1]R_m + x_m l_1 S \end{aligned}$$

Then by substituting R_m given in equation (5.41)

$$x_m$$

$$= F_m - x_m l_1 F_m + [1 - x_m l_1] \frac{\left[\frac{M l_1}{N_r} - \frac{x_m l_1}{S}\right]\overline{S^2} + \left[\frac{1}{S_2} - \frac{M l_1 S}{S_2 N_r}\right]\overline{S_2{}^2}}{2[1 - x_m l_1]} + x_m l_1 S$$

$$= F_m - x_m l_1 F_m + \left[\frac{M l_1}{2N_r} - \frac{x_m l_1}{2S}\right]\overline{S^2} + \left[\frac{1}{2S_2} - \frac{M l_1 S}{2S_2 N_r}\right]\overline{S_2{}^2} + x_m l_1 S$$

Hence the average service time of the m^{th} station, x_m is

$$x_m = \frac{F_m + \frac{M l_1 \overline{S^2}}{2N_r} + \left[\frac{1}{2S_2} - \frac{M l_1 S}{2S_2 N_r}\right]\overline{S_2{}^2}}{1 + l_1 F_m + \frac{l_1 \overline{S^2}}{2S} - l_1 S} \tag{5.43}$$

When x_m is known the average service time, x for all the station can be found by using the equation (5.1)

$$x = \sum_{m=1}^{M} \frac{x_m}{M}$$

where x_m is given in equation (5.43). As F_m, N_r and S are parameters that depend on x, the equations (5.1) can be solved by a substitution method.

5.3.4 Average Delay of P1 packets

After solving equation (5.1), the corresponding value for x_m is known. Then the second moment of the service time of m^{th} station, $\overline{x_m^2}$ can be found from Table 5.6.

$$\overline{x_m^2} = [1 - x_m l_1](R_m + F_m)^2 + x_m l_1 S^2 \tag{5.44}$$

From the properties of M/G/1 queue, we can now obtain the average delay of *P1* packets of the m^{th} station, T_m:

$$T_m = x_m + \frac{l x_m^2}{2(1-r_m)}$$

As the average arrival rate of packets for all the stations are equal, on average the total number of packets being served by each station must be equal. So the average delay of *P1* packets, T_1 can be given as:

$$T_1 = \sum_{m=1}^{M} \frac{T_m}{M}$$

As shown in Figure 5.22, the model given in the section (5.2) cannot be used when *P2* saturates the channel. The model shows that *P2* traffic saturates the channel when the number of stations reaches to 32 with the given arrival rates. When this happens, the model given in this section has to be used. The analytical results obtained by combining the two models is shown in Figure 5.28.

From the figure, it is seen that priority one packets have less delay compared to priority two packets. This shows that the proposed MAC scheme supports service differentiation. The model given in section 5.2 is used if the channel is not saturated by *P2* traffic otherwise, the model given in section 5.3 is used.

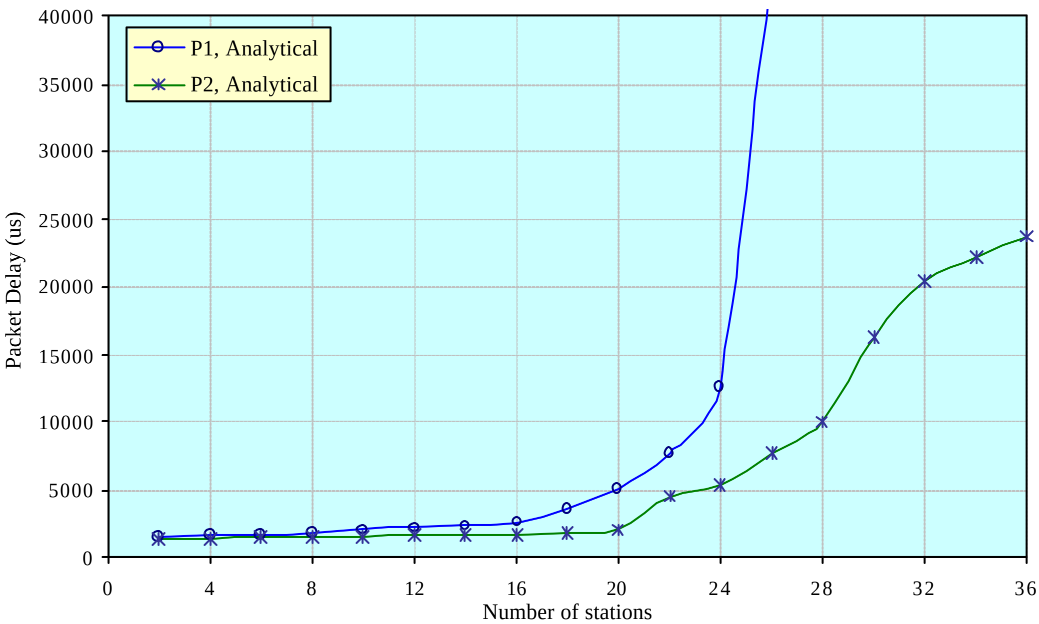


Figure 5.28 Average delay of priority one and two packets

5.4 Saturation throughput

The saturation throughput is defined as the throughput achieved in the system in saturation conditions, i.e., when all the stations of the BSS have non-empty transmission queues. The throughput is the proportion of time spent in transmission of useful data. This is an important performance measure of a MAC scheme. To obtain the saturation throughput, let us assume that the saturation condition is reached, so that every station of the BSS has the same priority data ready for sending. This situation is shown in Figure 5.29. The figure shows that the BSS has 3 stations and every station has data of P_1 ready to be sent.

In the figure, the same situation is repeated during the next period in time. The duration of that period of time, $P_{Duration}$, for a BSS that has M stations and has data of P_1 is given by:

$$P_{Duration} = AIFSN_1 + RTS/C + AIFSC_1 + CRB + M \times CRIFS + M \times PPB$$
$$+ (M-1)(SDIFS + ACK/C + D/C + TP/C + TR/C + 3 \times SIFS) + (SDIFS + ACK/C + D/C + SIFS)$$

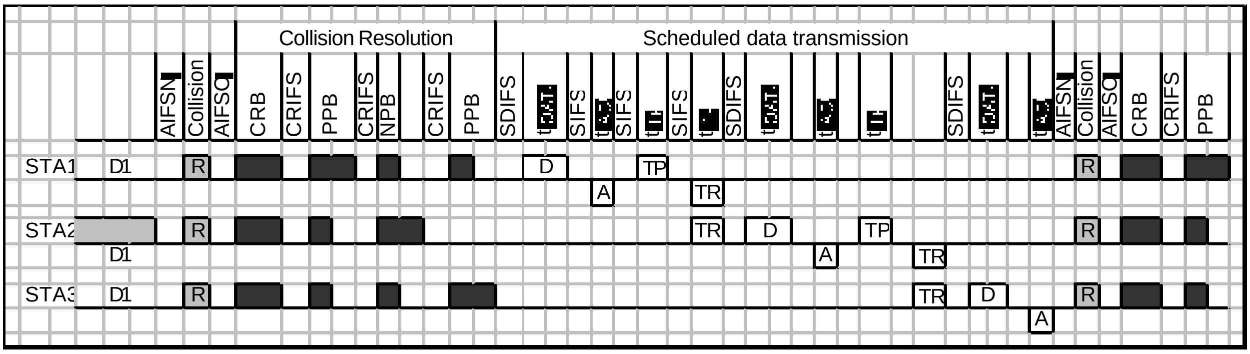


Figure 5.29 The BSS is in saturation condition with priority 1 data

The only useful information (data payload) sent by the stations is that contained in data frames. The data frames themselves contain control bits (MAC header and PHY header) and data payload bits. The time taken to transmit the payload data, T_I, during that period of time is given as:

$$T_I = M \times (D - h)/C$$

where h is the sum of PHY and MAC header bits. Then the saturation throughput can be obtained as:

$$SaturationThroughput = \frac{I_T}{P_{Duration}} \tag{5.45}$$

Figure 5.30 shows the results obtained for the saturation throughput. The parameters used in the calculation are shown in Table 5.7. These parameter values are typical values taken from [IEEE 802.11 1999]. The figure shows that saturation throughput is almost independent on the number of stations in the BSS and depends on the payload length of the data packets.

Table 5.7 Parameters of the proposed MAC scheme

Channel Bit rate (C)		2 Mbps	PHY header		128 bits
Data packet payload (D-h)		vary[in bits]	MAC header		272 bits
ACK	240 bits	Slot Time	20 µs	CRB	150 µs
RTS	288 bits	SIFS	10 µs	$AIFSC_2$	90 µs
CTS	240 bits	CRIFS	20µs	$AIFSN_2$	110 µs
TP	240 bits	SDIFS	30 µs	PPB	30 µs
TR	240 bits	$AIFSC_1$	50 µs	NPB	10 µs
TimeOut	20 µs	$AIFSN_1$	70 µs		

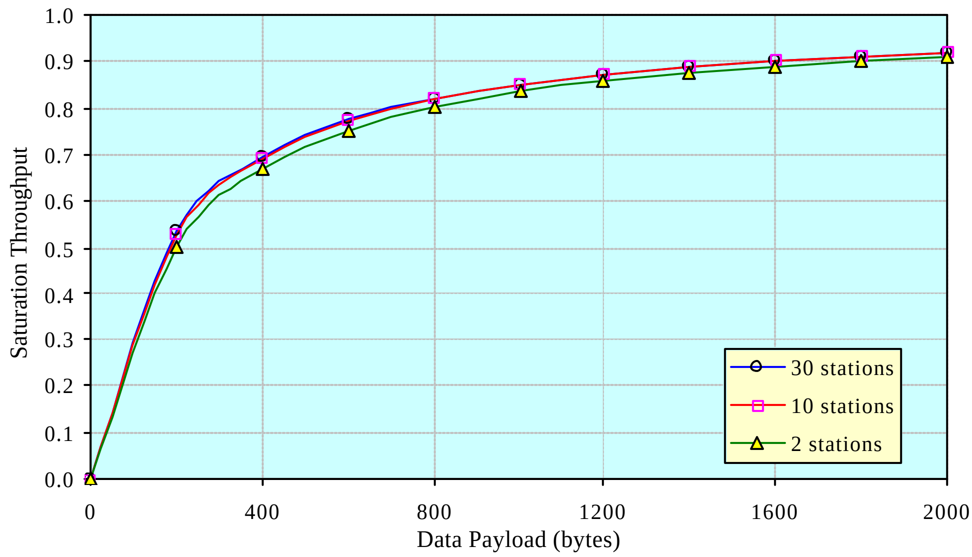


Figure 5.30 Saturation throughput of the proposed MAC scheme

5.5 Summary

The chapter presented a mathematical model for the proposed MAC scheme. In the model, a maximum of two priority classes of traffic are considered and evaluated. The results obtained for the average delay of the P1 and P2 traffic is presented for different conditions. The results show that P1 traffic experiences more delay compared to P2 traffic. Hence, the MAC scheme supports service differentiation.

An expression is also derived in this chapter for saturation throughput in the proposed MAC. The results show that saturation throughput is very much dependent on the data packet payload and almost independent of the number of stations in the BSS.

The analytical results obtained in this chapter are compared with the results obtained by simulation in chapter 6.

6.0 Simulation Models and Results

In this chapter we present a performance evaluation study of the proposed MAC scheme as well as the IEEE 802.11 DCF, and IEEE 802.11e EDCF MAC schemes by means of stochastic simulation. Simulation models were developed using MATLAB. The chapter presents the simulation models used and the results obtained. Results obtained by simulation and analytically for the proposed MAC scheme are also compared. We also present a comparison study of the proposed MAC scheme with IEEE 802.11 DCF and IEEE 802.11e EDCF.

6.1 Introduction

Simulation models for the all considered MAC schemes (IEEE 802.11 DCF, IEEE 802.11e EDCF, and the proposed MAC scheme) were designed and implemented in MATLAB. The details of the IEEE 802.11 DCF, IEEE 802.11e EDCF, and the proposed MAC scheme models are described in sections 6.4, 6.5 and 6.6, respectively. For each of the schemes, performance was evaluated in the sense of throughput, service differentiation, and real-time traffic delay, defined as follows:

- **Maximum saturation throughput**: The maximum saturation throughput is the throughput achieved under a given MAC scheme when the scheme is operating under saturated conditions. In our simulations, saturated conditions were achieved by generating traffic so fast that each station in a given BSS (Basic Service Set) always had a non-empty transmission queue. These simulations were done assuming perfect channels, so that no packet was lost due to transmission error. Having determined the number of data packets, N_D, serviced by the channel during a given time interval, t, the saturation throughput, T_{sat}, can be expressed as:

$$T_{sat} = \frac{N_D.PL}{t} \tag{6.1}$$

where PL is the time taken to transmit a data packet payload.

- **Throughput with imperfect channels**: This throughput can be determined by running simulation under the presence of channel errors. Having determined the number of data packets, $N_{D,er}$, served by a given channel during time interval, *t*, the throughput, T^{er} can be expressed as:

$$T_{er} = \frac{N_{D,er}.PL}{t}$$

  where *PL* is the time taken to transmit a data packet payload.

- **Quality of Service (QoS):** The QoS supported by the MAC schemes was evaluated by means of service differentiation achieved by them. Service differentiation is a performance measure used to see how good or bad a given traffic stream performs compared to other traffic streams. In our simulations, each station in the BSS generated two streams of voice traffic. One stream was serviced with higher priority than the other. The service differentiation was assessed by comparing the average delay associated with voice data packets of each of these two priority classes of data.

- **Real-time traffic performance**: In our simulations, each station generated either ON-OFF or Continuous Bit Rate (CBR) voice traffic. In each case, delays associated with the voice packets were calculated for varying number of stations in the BSS.

Three different types of data traffic were generated: Continuous Bit Rate (CBR) traffic, ON-OFF traffic and background data traffic. The CBR and ON-OFF traffic represented voice data. Background data traffic was modelled by Poisson processes.

In our simulation models the wireless channels were represented by the 2-State Markov chain models. Such a channel can be either in a good or bad state. The bit error rate (BER) in the good state would always be lower then in the bad state. More details about the channel models are given in section 6.3.

The following assumptions simplifying the simulation models were made:

AS1. Only homogeneous, i.e. networks with stations generating the same traffic, are simulated.

AS2. It was assumed that every station in the BSS is in the transmission and receiving range of all other stations, so there occur no Hidden Terminal problem.

AS3. Each station has infinite buffer space so no packets were lost due to a full buffer.

The following sections describe the data traffic models, channel error models, simulation models of the MAC schemes, methods used to analyze simulation output data and results obtained in the simulations.

6.2 Data Traffic Models

Each station in the channel can generate three different types of data traffic: CBR traffic, ON-OFF traffic, and/or background data traffic. The stations which transmit real-time data like voice generate CBR or ON-OFF traffic as a model of real-time data. On top of the real-time data, the stations can also generate background data traffic.

In the CBR traffic model, we assumed a generation rate of 32 kb/s, with a packet size of 160 bytes as in [Krishnakumar and Sobrinho 1996]. Thus, the inter-packet time was 40 ms. In all simulated cases, the channel rate was set to 2 Mb/s. In all the MAC schemes considered, the physical layer (PHY) header had a length of 128 bits and the MAC layer header had length of 272 bits. These header lengths are the values used in the current IEEE 802.11 WLAN standard. Following these assumptions, a CBR frame duration, $F_{Duration}$, can be calculated as:

$$F_{Duration} = \frac{P_{size} + PHY_{Header} + MAC_{Header}}{C_{Rate}}$$

where $F_{Duration}$ is the frame duration in µs, P_{size} is a CBR packet size in bits, PHY_{Header} is the PHY header size in bits, MAC_{Header} is the MAC layer header size in bits, and C_{Rate} is the channel rate in Mb/s. Thus:

$$F_{Duration} = \frac{160 * 8 + 128 + 272}{2} = 840$$

So a CBR frame has duration of 840 µs. This is shown in Figure 6.1.

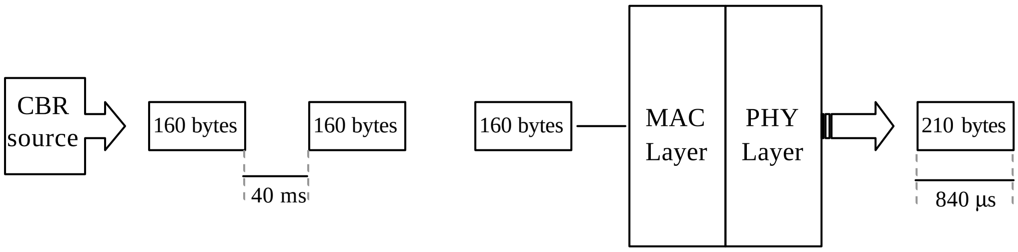


Figure 6.1 CBR traffic generation

In the ON-OFF traffic model, the voice traffic was modeled using a 2-state Markov chain as shown in Figure 6.2. As with CBR, traffic was generated during the ON periods at a rate of 32 kb/s with a packet size of 160 bytes. Hence, as before, the inter-packet time was 40 ms and the duration of the voice frame (including MAC and PHY headers) was 840 µs. In the OFF periods, the traffic source stays idle. The transition rate from ON to OFF, α, and OFF to ON, β, was set to be equal so that the model had exponentially distributed ON and OFF periods of 300 ms average each [Barry et al. 2001].

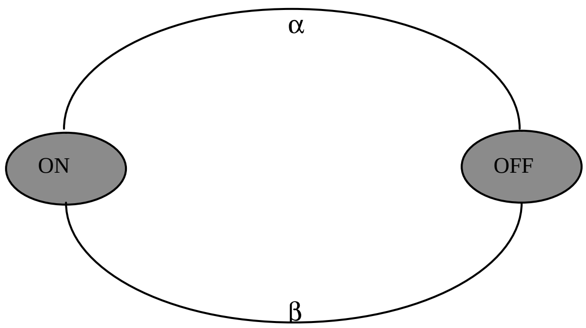


Figure 6.2 2-state Markov chain ON-OFF traffic model

In background data traffic, data packets were generated at each station according to a Poisson process with an average packet generation rate of l packets per second. In the Poisson process, the inter-arrival times are exponentially distributed with an average inter-arrival time of $1/l$ second. The parameter l can be varied to achieve the desired level of data load in the channel. For example, if there are M stations in the channel, and all the stations generate background traffic, then the data load, r_D, in the channel can be calculated as $r_D = lMb_{data}/r_C$. Here b_{data} is the number of bits in the data packet (including headers) and r_C is the channel bit rate. This is shown in Figure 6.3.

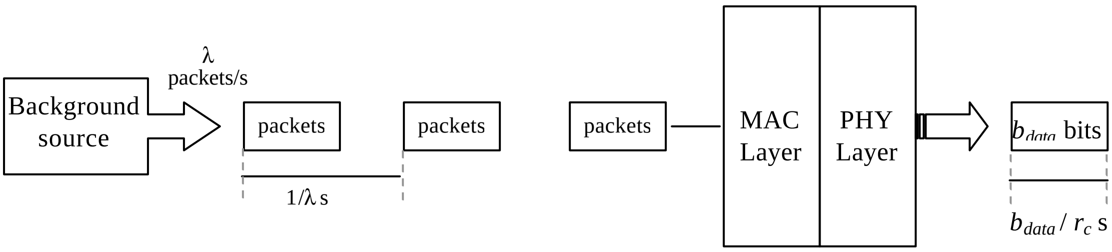


Figure 6.3 Background traffic generation

The codes used to implement these traffic models are attached in Appendix A.1. In the simulations, traffic arrival times were pre-computed for the duration of the maximum simulated time, and these arrival times were stored in an array.

6.3 The wireless channel model

The wireless channel was modeled using the burst error model proposed by Gilbert [Gilbert 1960]. It is a 2-state continuous time Markov chain as shown in Figure 6.4. State G represents the channel in a good state with very low error rate, BER_G. State B represents the channel in a bad state, when it operates in a fading condition with a higher error rate, BER_B. The transition rate from good to bad state is denoted by α, and transition rate from the bad to good state is denoted by β. The times spent in state G will alternate with the times spent in state B. The lengths of the times spent in each state have exponential distributions, with a mean length of $1/\alpha$ for the G state and mean length of $1/\beta$ for the B state.

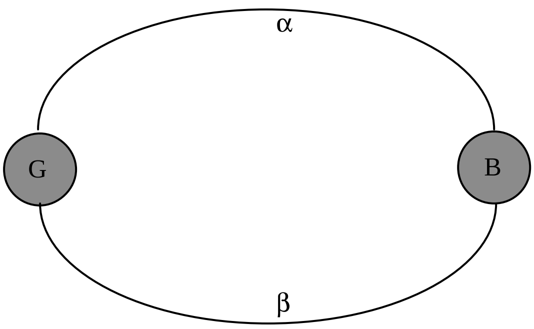


Figure 6.4 Burst error model represented by 2-state continuous-time Markov chain

Depending on the time at which a frame is transmitted, the frame may be transmitted in a good state, in a bad state, or part of the frame in a good state and the other part in a bad state. Thus we can distinguish three cases:

Case 1: When a frame transmission starts, the channel is in the good state, and there is no transition out of this state before the frame transmission completes.

Case 2: When a frame transmission starts, the channel is in the bad state, and there is no transition out of this state before the frame transmission completes.

Case 3: A frame transmission starts with the channel in either state and the channel undergoes one or more transitions before the frame transmission completes.

The probability that the frame is transmitted successfully, *P (success)* is calculated as:

$$P(success) = (1 - BER_G)^{N_G} . (1 - BER_B)^{N_B}$$

where

BER_G is the bit error rate in good states.

BER_B is the bit error rate in bad states.

N_G is the number of bits of the frame transmitted in good states.

N_B is the number of bits of the frame transmitted in bad states.

$N_G + N_B$ = the length of the frame.

If the probability that the frame is transmitted successfully, *P (success)*, for a given N_G and N_B is known, then whether the frame is transmitted successfully can be determined by choosing a random number, r_n, between 0 and 1. If $r_n >$ *P (success)* then the frame is considered to be corrupted and contains one or more errors. Otherwise, the frame is considered to be transmitted successfully. The transition rates used in the simulation are $\alpha = 30\ sec^{-1}$ and $\beta = 10\ sec^{-1}$ and are estimates reported in [Cocker et al. 2001] and [Deng and Chang 1999].

6.3.1 Implementation of the model in MATLAB

The above model was implemented in MATLAB by using two functions. At the start of each simulation, the wireless channel error model is initialized by using the initializing function called INITIALISING_ERROR_MODEL. The function first generated exponentially distributed random numbers for the times spent in state G and the times spent in state B with corresponding mean values. Then it chooses a random number, r_n, between 0 and 1. If $r_n > 0.5$ the simulation starts in a good state or else in a bad state. According to this information, the channel time is partitioned into alternating good and bad periods as shown in Figure 6.5. In the simulation, these good and bad time intervals of the channel have been pre-computed for the duration of the maximum simulated time and stored in an array. Each cell in the array contains a number which indicates when the period is ending. The INITIALISING_ERROR_MODEL function returns this array with information saying whether the simulation starts in a good or a bad state. The code of this function is given in Appendix A.2.

When a frame is transmitted, the function called FADING_ERROR was used to check whether the frame was transmitted successfully. The function uses the following information: frame transmission starting time, frame duration, and information obtained from the function INITIALISING_ERROR_MODEL. The FADING_ERROR function first finds whether the frame transmission starts at a bad or a good state. After that, the function finds the number of bits transmitted in the good and bad states and the probability that the frame is transmitted successfully, *P(success)*. Then this function decides whether the frame contains an error by choosing a random number. The code of the function is included and discussed in Appendix A.2.

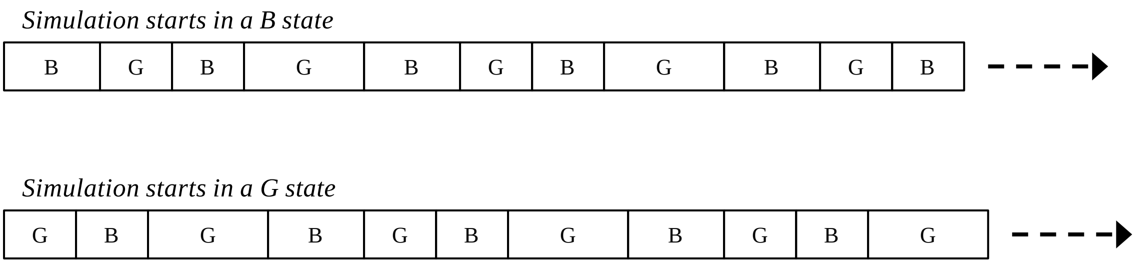


Figure 6.5 Alternating good and bad states in the channel error model

6.4 IEEE 802.11e EDCF

In IEEE 802.11e EDCF, each station can have up to eight different traffic categories (TCs). But to make the simulation model simple, only two TCs were considered as shown in Figure 6.6. As explained in chapter 3, in IEEE 802.11e EDCF, service differentiation between TCs was achieved by using different values of AIFS, CW_{min}, CW_{max}, and PF for different TCs. The following sections present more detail about the simulated scenario, simulation model, and results.

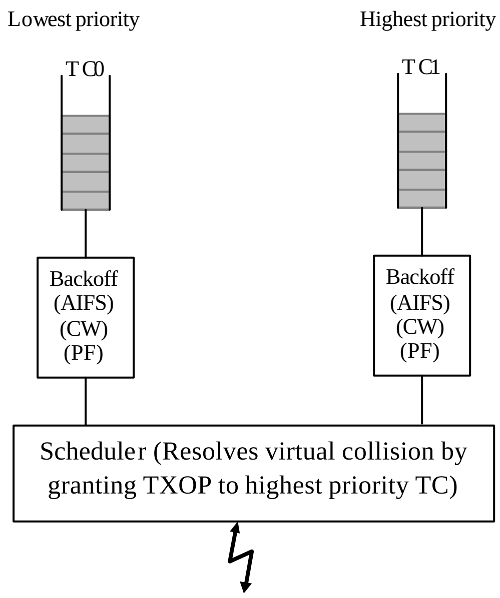


Figure 6.6 Two priority level simulation model of the IEEE 802.11e EDCF

6.4.1 Simulated Scenario

Simulation was done for a single Quality of Service supporting BSS (QBSS). In a QBSS, each station can directly communicate with all other stations as shown in Figure 2.1. The QBSS works as a wireless *ad hoc* network so there is no need of an AP. Each station generates the same types of data traffic depending on the type of simulation.

A wireless channel was modelled as discussed in the section 6.3 and it has been assumed that the transmission power and distances between stations were allocated in such a way that there were no hidden stations in the channel. Furthermore, the RTS/CTS access method was always used to minimize hidden station problems, irrespective of frame length. Table 6.1 shows the values selected for the two priorities of priority 1 (P1) and priority 2 (P2), and parameters of EDCF in the reported simulation studies.

Table 6.1 Parameters of IEEE 802.11e EDCF

EDCF parameters		
MAC header	272 bits	
PHY header	128 bits	
ACK	112 bits + PHY header	
RTS	180 bits + PHY header	
CTS	112 bits + PHY header	
SIFS	10 us	
Slot Time	20 us	
Channel Bit rate (C)	2 Mbps	
Two priority level parameters		
	Priority 1	Priority 2
AIFS	50 us	70 us
CWmin	7	15
CWmax	127	255
PF	2	2

6.4.2 Simulation Model

The model used to simulate IEEE 802.11e EDCF is given in Figure 6.7. Initially, the number of stations in the channel and the maximum simulated time were specified. The method used to estimate the maximum simulated time is explained in section 6.7. Then according to the maximum simulated time of the simulation, traffic was generated and the channel model initialized.

If the current simulated time is not greater than the maximum simulated time, the stations serve the packets generated by using the MAC procedure of the EDCF. Otherwise, the simulation is finished and statistics are analyzed.

In serving packets generated by the stations, the server first finds the time that the packet at the front of the queue of each station needs to be transmitted (due time). The due time for Priority one (P1) and Priority two (P2) were calculated by using the backoff time slots accompanied with each packets. After calculating the due time of the packets at the front of the queues, the server then finds which stations have the smallest due time. If two or more stations have the smallest due time, then a collision will occur and the channel contention will be unsuccessful or if only one station has two packets of different priority with smallest due time then this *virtual collision* is resolved by giving the chance to the higher priority packet.

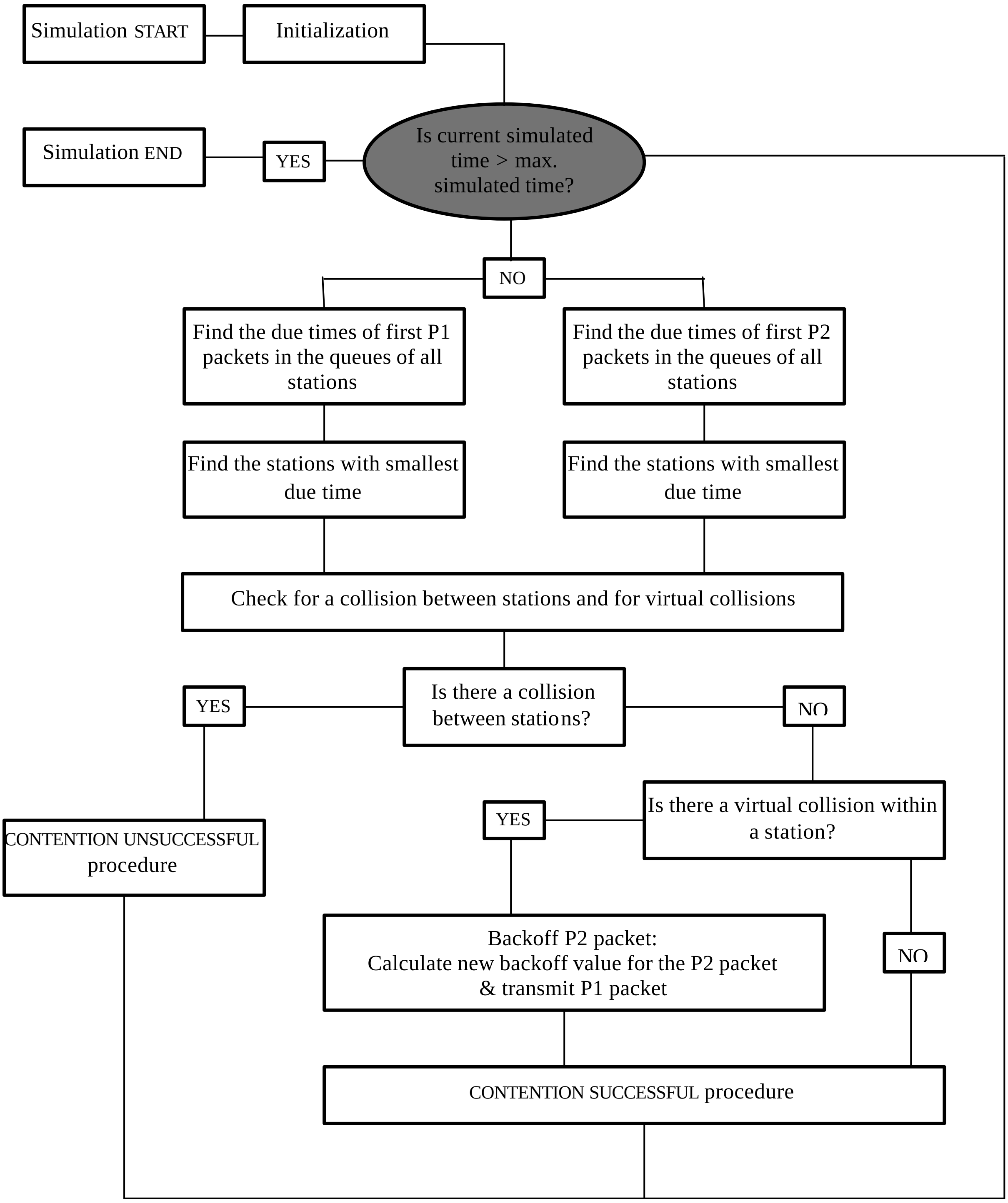


Figure 6.7 Simulation model of the IEEE 802.11e EDCF

When an unsuccessful contention occurs, the un-collided stations decrement the backoff counter and the collided stations check whether the retry limit of the packet is reached. If the retry limit is reached, the packet is discarded and a new packet comes to the front of the queue, but if the retry limit is not reached then a new backoff value is calculated for the collided packet. This is shown in Figure 6.8. After this the server again checks whether the simulation runtime is reached, and if not, the server again finds the due time of the packets at the front of the queue, and repeats the process.

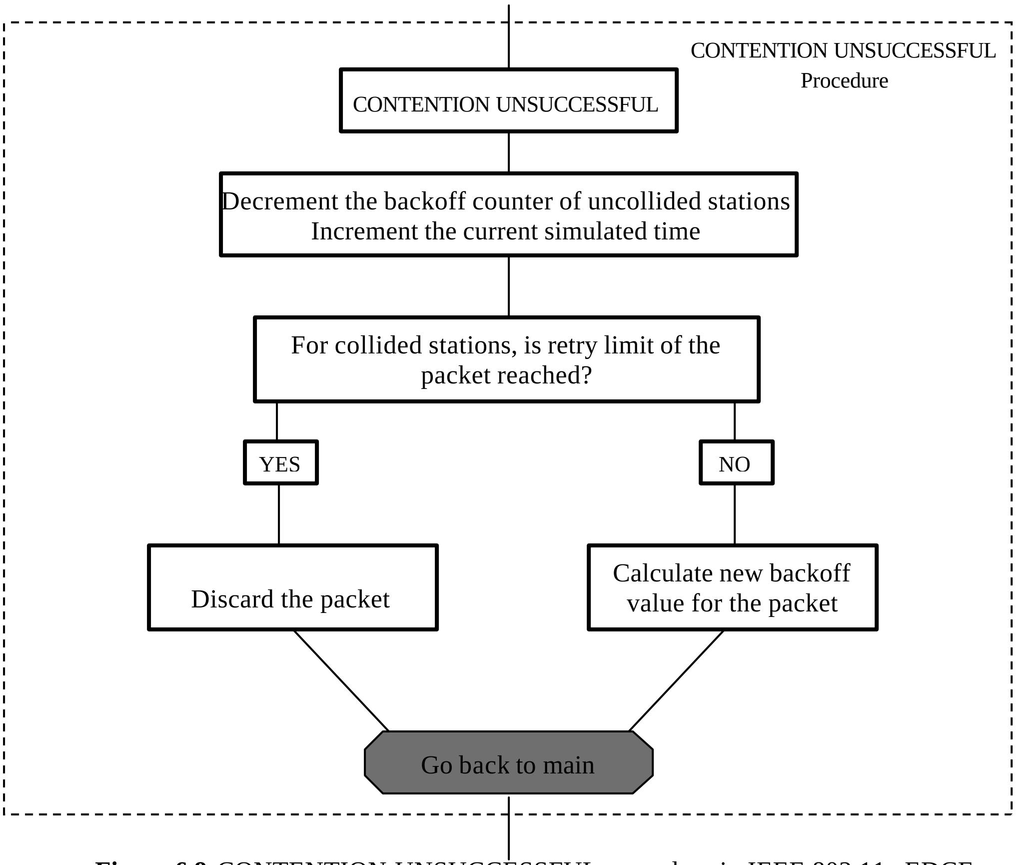


Figure 6.8 CONTENTION UNSUCCESSFUL procedure in IEEE 802.11e EDCF

If there is no collision between stations, i.e. only one station has the smallest due time, then the server checks whether there is a *virtual collision*, i.e. clashes between data packets of different priority within a station. The lower priority (P2) packet needs to backoff if there is a virtual collision and then the higher priority (P1) packet will be transmitted by using a procedure called CONTENTION SUCCESSFUL. If there is no virtual collision within the station, the due packet (either P1 or P2 packet) is transmitted by using the CONTENTION SUCCESSFUL procedure shown in Figure 6.9. After this procedure, the server again checks whether the maximum simulated time is reached, if not the server again starts the actions all over again.

In the CONTENTION SUCCESSFUL procedure, the backoff counters of the stations are updated and the procedure finds the station which has a packet with the smallest due time. The procedure tries to send the packet by using RTS/CTS method. First, the RTS frame is transmitted to the destination station and waits for the CTS frame. Then these two frames are checked for the transmission errors by using the FADING_ERROR procedure. If either of these frames contain en error, the Station Short Retry Count (SSRC) increases by 1 because these are short frames. If the retry limit is reached, the packet is discarded and then the server starts to serve the next packet in the queue. Otherwise, a new backoff value is calculated for the packet and the whole procedure is started again.

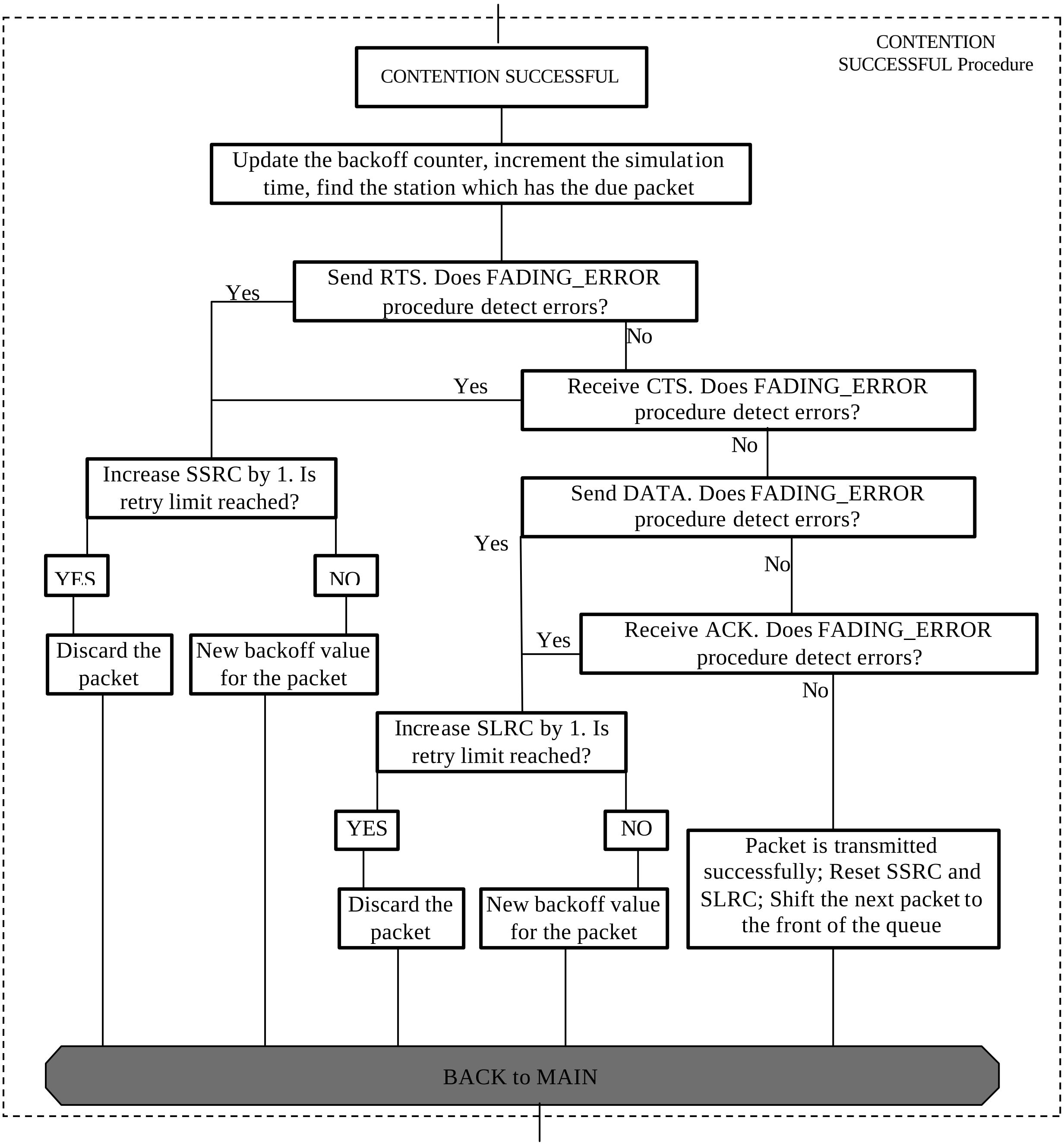


Figure 6.9 CONTENTION SUCCESSFUL procedure in IEEE 802.11e EDCF

If RTS and CTS frames are successful, the station sends the data frame and waits for the ACK frame from the destination station. If either of these frames contain an error, the Station Long Retry Count (SLRC) increases by 1 because normally the data frame is a long frame. If the retry limit is reached, the packet is discarded and then the server starts to serve the next packet in the queue. Otherwise, a new backoff value is calculated for the packet and the whole procedure is started again. On the other hand, if both data and ACK frames are transmitted successfully, the SSRC and SLRC resets and the next packet is brought to the front of the queue and the whole procedure is started again.

The MATLAB codes used to implement these procedures are given in the Appendix A.3. More detail about the model is included in the codes as comments.

6.5 Implementation of IEEE 802.11 DCF

In IEEE 802.11 DCF, different priority data traffic is not supported unlike in IEEE 802.11e EDCF. Thus, each station in the channel can be represented by a single queue as shown in Figure 6.10. The queue may contain a mixture of traffic types like voice and data, but both types of traffic have an equal chance to access the channel.

The model used to simulate IEEE 802.11 DCF is similar to that of IEEE 802.11e EDCF. The major difference is that DCF only has one traffic queue, and with IEEE 802.11 DCF, the backoff counter is reduced by one at the beginning of the first slot interval after the DIFS period. The MATLAB codes implementing the model are given in Appendix A.4. The details of the model are similar to that explained in section 6.4.2.

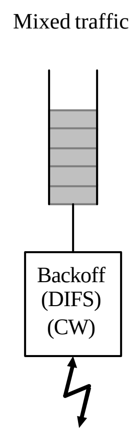


Figure 6.10 Simulation model of the IEEE 802.11 DCF

6.5.1 Simulated Scenario

Performance of a single BSS was simulated. In a BSS, each station can directly communicate with all other stations in the BSS, as shown in the Figure 2.1. The BSS works as a wireless *ad hoc* network, so there is no need of an AP. All stations generate the same traffic appropriate for a given simulated case. The traffic can be pure voice, data or a mix of both.

In the simulation, a wireless channel was modeled as discussed in section 6.3 and it is assumed that the transmission power and distances between stations are such that there are no hidden stations in the channel. Furthermore, RTS/CTS access method was always used to minimize hidden station problems, irrespective of frame length. Table 6.2 shows the DCF parameters taken from [Cocker et al. 2001].

Table 6.2 Parameters of IEEE 802.11 DCF

DCF parameters	
MAC header	272 bits
PHY header	128 bits
ACK	112 bits + PHY header
RTS	180 bits + PHY header
CTS	112 bits + PHY header
SIFS	10 us
Slot Time	20 us
Channel Bit rate (C)	2 Mbps
DIFS	50 us
CWmin	31
CWmax	255
PF	2

6.6 Implementation of the proposed MAC scheme

In the proposed MAC scheme, each station can have different priority level traffic. In order to simplify the simulation model, only two priority levels (P1 and P2) were considered and Figure 6.11 shows the model used in the simulation.

As explained in the chapter 4, service differentiation between different priority levels was achieved by using different values of AIFSN and AIFSC for different priorities. The following sections present details about the simulation.

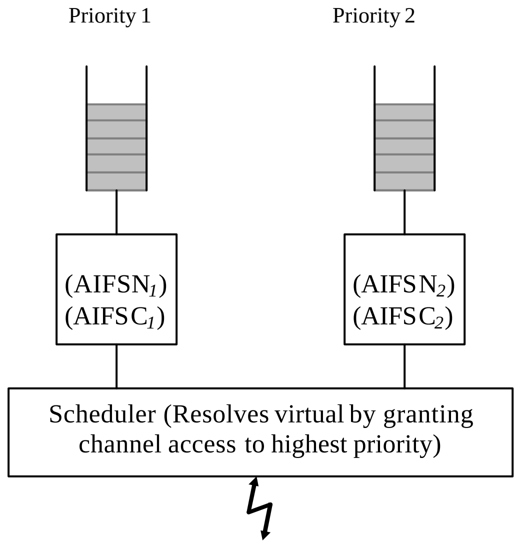


Figure 6.11 Two priority level simulation model of the proposed MAC scheme

6.6.1 Simulated Scenario

As given for the previous MAC schemes, simulation was done for a single BSS. The BSS works as a wireless *ad hoc* network so there is no need of an AP. Each station generates the same types of traffic depending on the parameter that needs to be evaluated by simulation.

A wireless channel was modeled as discussed in section 6.3, and it was assumed that the transmission power and distance between stations are in such a way that there are no hidden stations in the channel. Table 6.3 shows the parameters used to simulate the proposed MAC.

Table 6.3 Parameters of the proposed MAC scheme

Channel Bit rate (C)		2 Mbps	PHY header		128 bits
Data packet payload (D)		vary[in bits]	MAC header		272 bits
ACK	240 bits	Slot Time	20 µs	CRB	150 µs
RTS	288 bits	SIFS	10 µs	$AIFSC_2$	90 µs
CTS	240 bits	CRIFS	20µs	$AIFSN_2$	110 µs
TP	240 bits	SDIFS	30 µs	PPB	30 µs
TR	240 bits	$AIFSC_1$	50 µs	NPB	10 µs
TimeOut	20 µs	$AIFSN_1$	70 µs		

6.6.2 Simulation Model

The simulation model used for the proposed MAC scheme is given in Figure 6.12. Initially the number of stations in the channel and maximum simulated time were specified. Then, according to the maximum simulated time, traffic was generated and the channel model initialized. If the current simulated time is not greater than the maximum simulated time, the stations serve the packets generated by using the proposed MAC procedure. Otherwise, the simulation is finished and the statistics are analyzed.

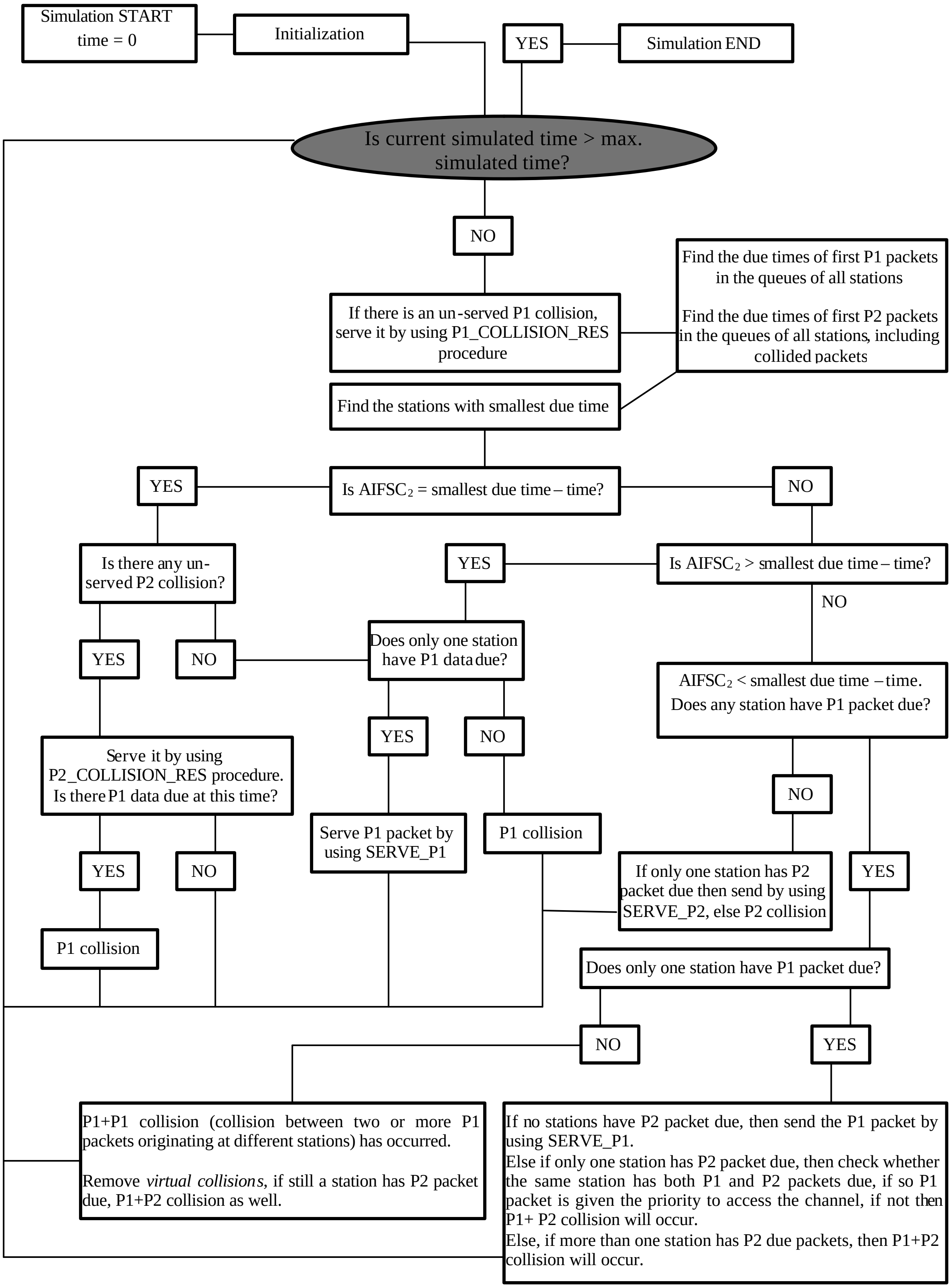


Figure 6.12 Simulation model of the proposed MAC scheme

If the maximum simulated time is not reached, the server first finds whether there is an un-served P1 collision. A P1 collision has occurred if a P1 data packet of a station had collided with a data packet of another station. If there is an un-served P1 collision, the collided packets are served by using the procedure called P1_COLLISION_RES. The detail of the procedure is shown in Figure 6.13. The procedure, first sends a set of beacons to resolve the collision. Then the collided stations send data packets according to station ID number. Each packet transmitted during the collision resolution is checked for transmission errors by using the FADING_ERROR procedure. The P1_COLLISION_RES procedure is attached in the Appendix A.5. More detail about the procedure is included in the codes as comments.

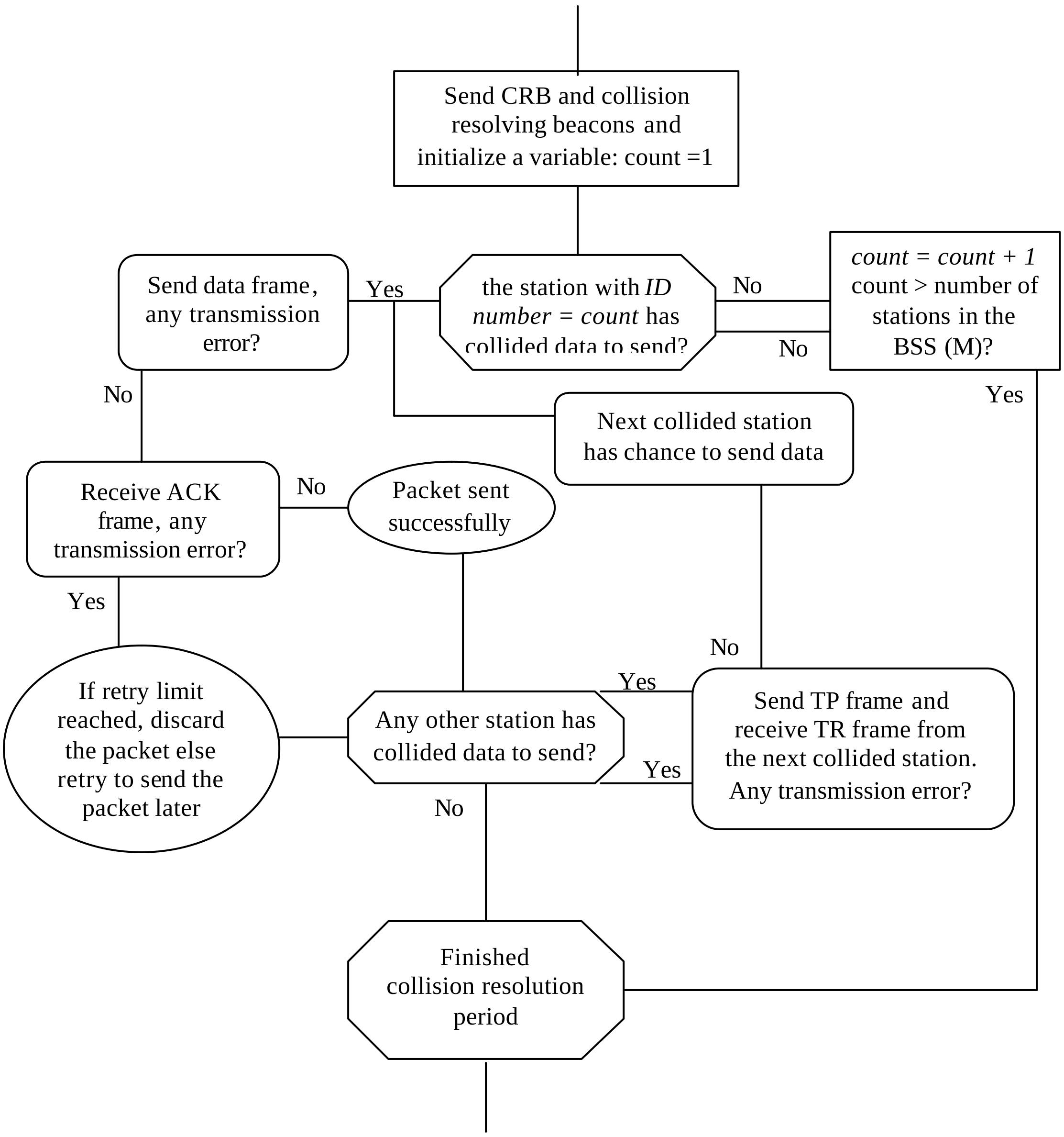


Figure 6.13 Collision Resolution procedure

Having served the collided packets of P1, the server then finds the due time of the packets at the front of the queue of each station. The due times for P1 and P2 packets are calculated according to the arrival times of the packets. If there are collided packets of P2, the due times for these packets are also found at this time. When the due times of all the packets at the front of the queue of each station are known, the server finds the stations which have the packet with the smallest due time.

If the time difference between channel time and the smallest due time of the packets are equal to the $AIFSC_2$ then that means the server may have P2 collided packets and P1 packets due at the time. So the server checks whether there is an un-served collision of P2. The un-served collided packets are then served, if there are any, by using P2_COLLISION_RES procedure. This procedure is identical as the one given in Figure 6.13. The only difference is that P2_COLLISION_RES procedure serves P2 collided packets rather than P1 collided packets. After serving the P2 collided packets, the server checks whether there are any P1 packets which has the smallest due time. If there are P1 packets due at this time, then P1 collision will occur and it will be served in the next period. On other hand, if there is no un-served P2 collision, the server serves the P1 packets due at this time. If there is only one station which has a P1 due packet at this time, the packet is served using SERVE_P1 procedure, or if there is more than one stations which has P1 due packets then P1 collision will occur and the collision is served in the coming period.

However, if the time difference between channel time and the smallest due time of the packets is less than the $AIFSC_2$, that means the server has only P1 packets due at this time. As before, if there is only one station which has a P1 packet due at this time, the packet is served using SERVE_P1 procedure, or if there is more than one station which has P1 packets due then P1 collision will occur and the collision is served in the coming period. The SERVE_P1 procedure sends the data packet by using the RTS/CTS method. The procedure also checks for transmission errors by using the FADING_ERROR procedure. If there is a transmission error, the data packet will be scheduled to retransmit in the next available time according to the procedure of the MAC scheme.

Also, if the time difference between the channel time and the smallest due time of the packets is more than the $AIFSC_2$ that means the server may have P1 and P2 packets due at this time. But there will not be any un-served P2 collided packets. The procedure to serve these packets is explained in Figure 6.12. The codes used to implement the model are given in the Appendix A.5.

6.7 Simulation Output Data Analysis

The performance parameters of the MAC schemes considered were analyze by simulations. A steady-state simulation is applied for investigating the long-run behavior of the MAC schemes. Each simulation was run for a specific long duration of time called *maximum simulated time*. Steady-state results were obtained by running the simulation for a time long enough for all the initial transient effects to settle down. Data collected at the beginning of each simulation were discarded, as non-representive for steady-state behaviour. To collect a satisfactory large sample of data representing steady-state, each simulation was run so long that the initial discarded data constituted always 10% of all data collected during the whole simulation.

The maximum simulated time for each simulated case was chosen by using a graphical method. Namely, a few pilot runs were first executed to determine the necessary maximum simulated time of simulation. In each of these pilot runs different random seeds were assumed to make the runs independent. From the pilot runs, the maximum length of initial transient was determined and then the maximum simulated time necessary for the simulation was estimated as:

$$Maximum\ simulated\ time = 10 \times (\max.\ initial\ transient)$$

An example of this procedure is shown in Figure 6.14a. The figure shows results obtained from three pilot runs made to estimate maximum length of initial transient. In each of these pilot runs, simulation was run for 10 seconds. From the figure, it was found that around 7000 packets were served in each pilot run and about 1500 packets were successfully transmitted during the initial transient phase. From this information, maximum length of initial transient was determined as:

$$(\max.\ initial\,transient) = \frac{(Simulated\ time) \times (Max.\,no.\,of\ \ packets\ successfully\ transmitted\ during\ the\,intial\ phase)}{(total\ no.\,of\ \ packets\,successfully\ transmitted)}$$

$$(\max.\ initial\ transisent) = \frac{10 \times 1500}{7000} = 2.14s$$

Therefore, the maximum simulated time necessary in that case is:

$$Maximum\,simulated\ time = 10 \times 2.14 = 21.4s$$

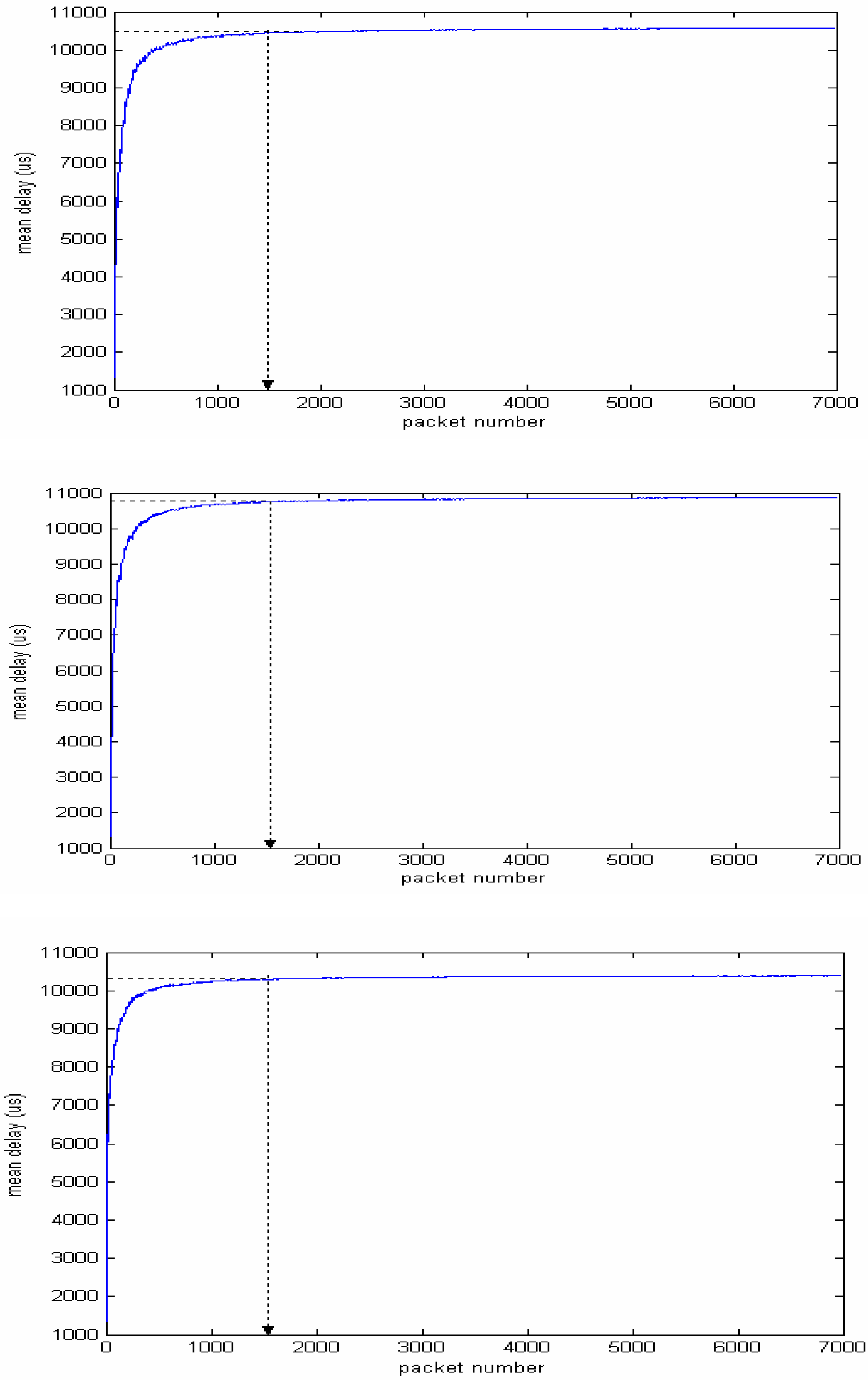


Figure 6.14a Three pilot runs to find the number of initial observations that need to be discarded

From this result, the maximum simulated time was set to 25 seconds for the simulated case. The results obtained for the 25 seconds maximum simulated time are shown in Figure 6.14b. This figure shows that the maximum simulated time was good enough for 10 % of initial data to be discarded.

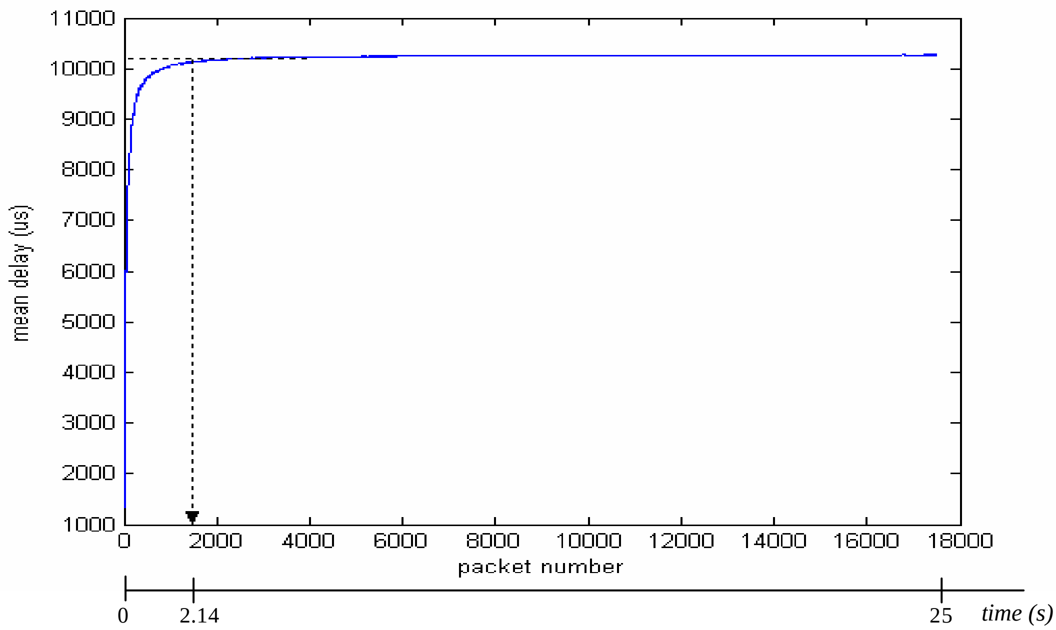


Figure 6.14b The maximum simulated time estimated was good enough for 10 percent of initial data to be discarded

After estimating the maximum simulated time for each case considered, the output data in each case was collected to estimate the performance parameters and their statistical errors using a method called *independent replications*. This is explained in the following section.

6.7.1 Method of Independent Replications

There are several common methods used to analyze steady-state data collected during simulation. The method we used is called *independent replications*. In this method, simulations were repeated a number of times, each time using a different, sequence of independent random numbers, and the mean value of a performance measure from each run was computed [Pawlikowski 1990]. These mean values were used in further statistical analysis as secondary, evidently independent and identically distributed output data items. In MATLAB, independent sequence of random numbers for each run was obtained by repeating the runs without resetting the random number generator and without restarting the MATLAB.

Having found the maximum simulated time for each simulated case, performance measures were estimated by running each simulation n times. In each case, the mean of the performance measure was calculated by ignoring the first 10 % of the collected data. If $\overline{X}_i$ is the mean value obtained from the data collected in the i^{th} replication, after the first 10% of the collected data were discarded, then the estimated mean value, $\overline{\overline{X}}$ of the performance parameter is obtained as:

$$\overline{\overline{X}} = \frac{1}{n}\sum_{i=1}^{n}\overline{X}_i$$

The accuracy with which the estimated mean value $\overline{\overline{X}}$ estimates the actual mean value m_x can be obtained by calculating confidence interval of m_x. This is explained in the section 6.7.2.

6.7.2 Error Analysis

The accuracy with which the estimated mean value $\overline{\overline{X}}$ estimates the unknown actual mean value m_x can be assessed by the probability:

$$P(|\overline{\overline{X}} - m_x| < D_x) = 1 - a$$

where D_x is known as the half width of the confidence interval and $(1-a)$ is the confidence level, $0 < a < 1$. This means that with the probability $1-a$ the interval $(\overline{\overline{X}} - D_x, \overline{\overline{X}} + D_x)$ contains the unknown actual mean value m_x. There

$$D_x = t_{n-1,1-a/2}\hat{s}\left[\overline{\overline{X}}\right],$$

$$\hat{s}^2\left[\overline{\overline{X}}\right] = \sum_{i=1}^{n}\frac{\left\{\overline{X}_i - \overline{\overline{X}}\right\}^2}{n(n-1)}$$

is the unbiased estimator of the variance of $\overline{\overline{X}}$, and $t_{n-1,1-a/2}$ is the upper $(1-a/2)$ critical point of the *t*-distribution with *(n-1)* degrees of freedom.

In our situation, we run each case *n* times and the t-distribution had *n-1* degree of freedom. After each time, we calculated the relative error, defined as:

$$Relative\_error = \frac{D_x}{\overline{\overline{X}}} 100\%$$

If the relative error was less than 10%, at 95% confidence level, the simulation was stopped. Otherwise, the simulation was repeated until the required error level was reached. An example to illustrate the error analysis is given below.

In this example, a simulation was run to obtain the average delay experienced by voice packets with increasing number of stations in a given BSS of the proposed MAC scheme. For each given number of stations in the BSS, a simulation was run n times. In each case, the mean delay of the voice packets were calculated while discarding the initial 10 % of the data collected. After collecting a mean value in each case, the relative error was calculated. When the relative error had reached less than 10%, at 95 % confidence level the simulation was stopped. The final relative errors, mean values, and the confidence intervals obtained are shown in Table 6.4. The results obtained from the calculation are shown in Figure 6.15. The figure shows a plot of the estimated mean values with 95 % confidence interval.

Table 6.4 Example to show confidence interval calculation

no. of stations	Mean	Dx	Relative error (%)	n (required number of replications)
2	2180	60	2.8	3
4	2320	160	6.9	2
6	2420	80	3.3	3
8	2480	150	6.0	2
10	2760	240	8.7	2
12	2840	90	3.2	3
14	3020	120	4.0	3
16	3300	220	6.7	2
18	3640	280	7.7	2
20	4130	300	7.3	5
22	4840	130	2.7	3
24	5520	490	8.9	5
26	7290	640	8.8	4
28	8880	810	9.1	5
30	12760	1030	8.1	4
32	17540	1660	9.5	8
34	27570	2590	9.4	6

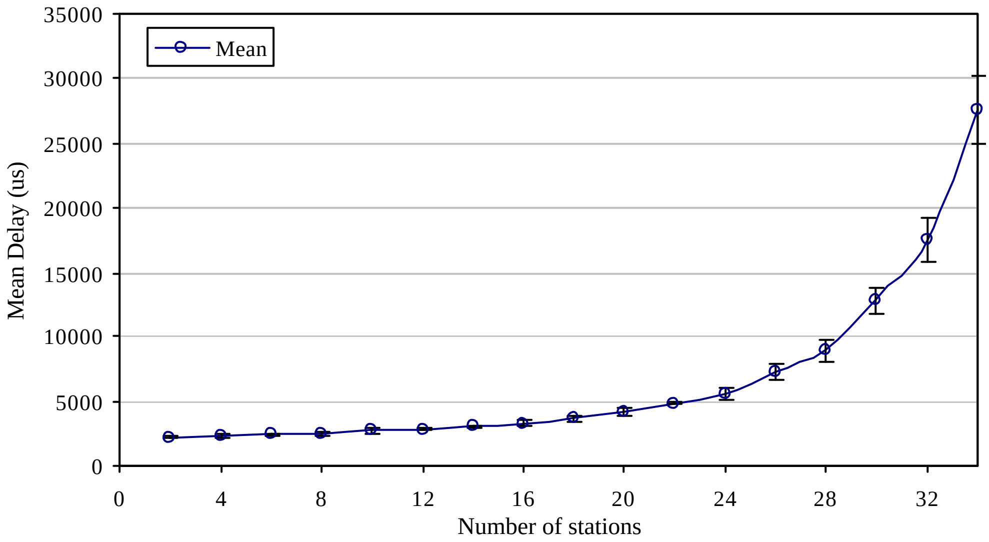


Figure 6.15 Plot of estimated mean values with 95% confidence interval

As shown in the example, for each performance measure, the simulation was run until the relative error was less than 10%, at 95% confidence level. The results obtained for different performance measures are given in the following section with the corresponding confidence intervals.

6.8 Results and Comparisons

This section presents simulation results obtained for the proposed MAC scheme and compares them with the analytical results obtained in chapter 5. We also present and compare simulation results obtained for other MAC schemes described above. For each MAC scheme, the simulation results are obtained for the performance parameters which are explained in section 6.1. The method used to analyze the simulation output data was discussed in section 6.7.

6.8.1 Analytical and simulation results of the proposed MAC scheme

In chapter 5, a mathematical model is derived for the proposed MAC scheme. The model is used to estimate the mean delay of priority one packets when the stations in a given BSS generate priority one traffic only. The model is also used to estimate the delay of the priority one and two packets when the stations generate priority one and two traffic.

Simulation was done under same conditions as given in the mathematical model and the results were compared with the results obtained from the mathematical model. In the first case, each station generates only P1 traffic according to a Poisson process with an average packet generation rate of l packets per second. The delay of P1 traffic was measured by varying the number of stations in the BSS. The results obtained are shown in Figure 6.16. In each simulation, payload of the packets was set to 1280 bits.

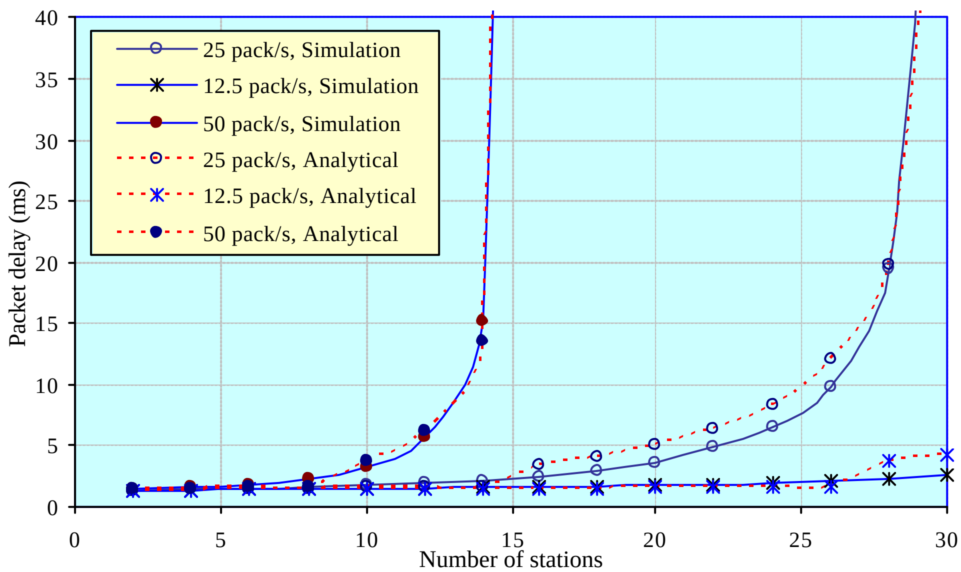


Figure 6.16 Comparison of simulation and analytical results obtained for P1 traffic in the proposed MAC scheme

The figure shows the results obtained for different Poisson arrival rates. From the figure, one can see that the analytical results agree well with the simulation results. The difference between simulation results and analytical results is probably due to rounding errors in the mathematical model and/or relatively large (10%) accepted statistical error of simulation results in this case.

In the second case, each station generates P1 and P2 traffic according to a Poisson process with an average packet generation rate of 12.5 packets per second in each priority class. The delay of P1 and P2 traffic were measured by varying number of stations in the BSS. The results obtained are shown in Figure 6.17. Same as above, payload of the packets was set to 1280 bits. The figure also shows a good agreement between analytical and the simulation results. Furthermore, both simulation and analytical results show that P2 packets had experienced more delay compare to P1 packets, so the proposed MAC scheme supports service differentiation.

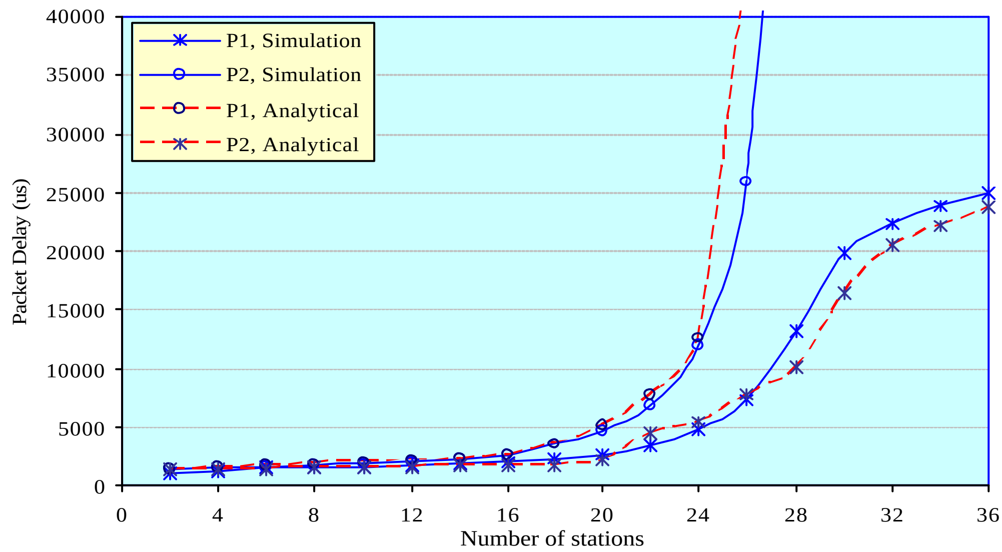


Figure 6.17 Comparison of simulation and analytical results obtained for P1 and P2 traffic in the proposed MAC scheme

An expression for saturation throughput achieved by the proposed MAC scheme has been derived in chapter 5. That expression can be used to determine the saturation throughput achieved by the proposed MAC scheme for a given packet payload for a known number of stations in the BSS. The simulation results obtained for the saturation throughput of the proposed MAC scheme are given in Figure 6.18. In both cases the saturation throughput is almost identical, as that shown in Figure 6.18. Hence, the saturation throughput of the proposed MAC scheme given by equation (5.45) is very accurate.

6.8.2 Comparison of the proposed MAC scheme with other MAC schemes

This section presents simulation results obtained for the different MAC schemes considered. For each MAC scheme, the simulation was used obtain the performance parameters explained in section 6.1. The MAC schemes are also compared here on the basis of simulation results to show the relative performance of each of them. Whenever it is possible 95% confidence intervals of the final estimate are given to show the accuracy of the results.

6.8.2.1 Maximum saturation throughput

The saturation throughput is defined as the throughput achieved by the system at saturation conditions, i.e., when all stations in the BSS have non-empty transmission queues. In our simulation, saturation throughput was measured for the highest priority traffic, P1. Hence each station in the BSS generated P1 traffic only, in such a way that saturation conditions at all stations were achieved. This was achieved by generating CBR traffic faster than the channel can serve, so that each station always had a non-empty transmission queue. In each model CBR traffic was given the highest priority, P1. The simulation was done with perfect channel conditions so that no packet was lost due to transmission errors. The saturation throughput was calculated by using the equation (6.1). Results for the different MAC schemes and for different data packet lengths are shown in Figure 6.18, 6.19 and 6.20.

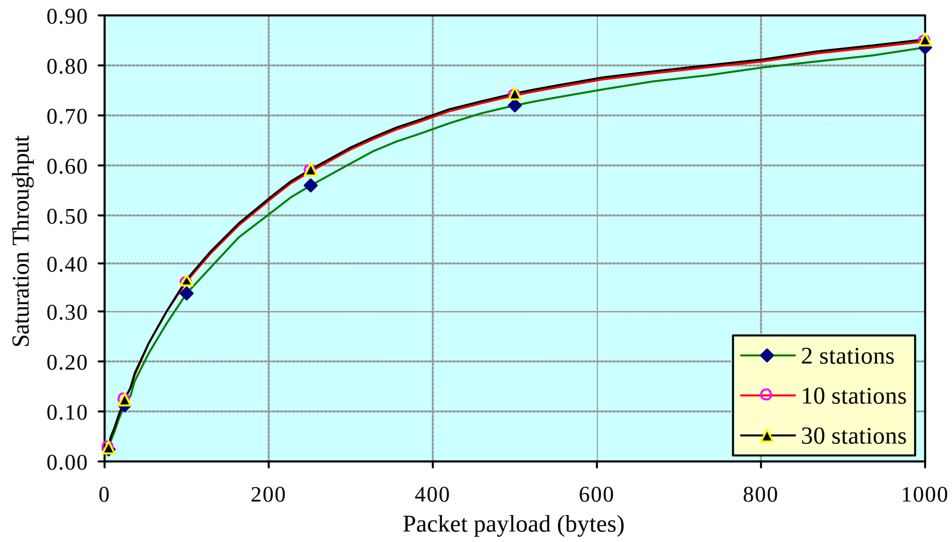


Figure 6.18 Saturation throughput of the proposed MAC scheme

In each of these cases, simulation was run until the relative error was less than 1%, at 95% confidence level. The confidence intervals obtained are not shown in the graphs because the intervals are very small if compared to the scale given. From the figures, it is seen that in each scheme the saturation throughput is very much dependent on the packet payload. However it does not depend much on the number of stations in the channel.

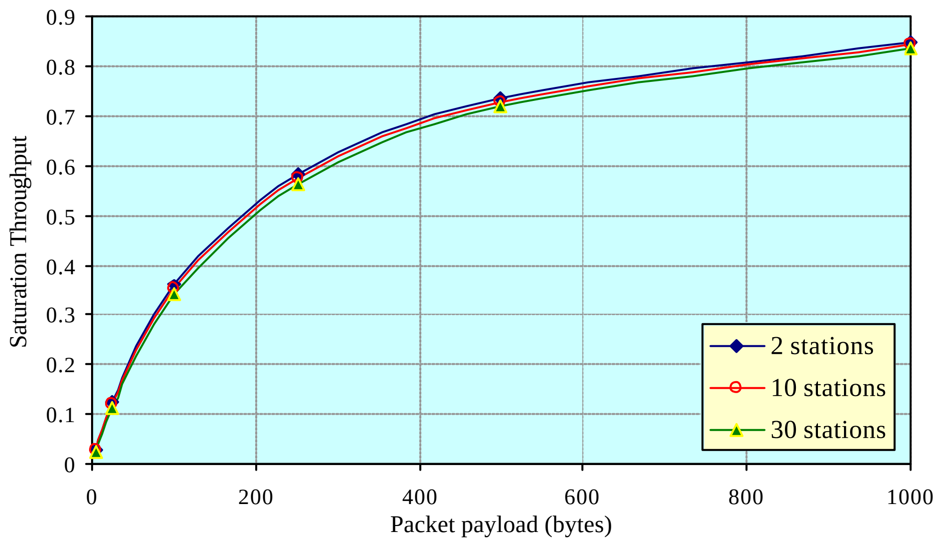


Figure 6.19 Saturation throughput of the IEEE 802.11e EDCF

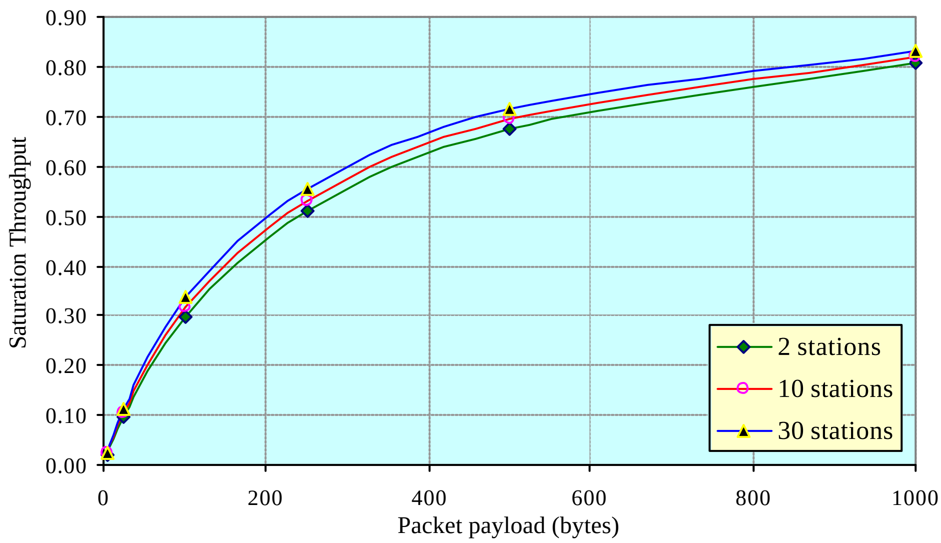


Figure 6.20 Saturation throughput of the IEEE 802.11 DCF

Figure 6.21 shows the results obtained for the three MAC schemes under the same conditions. The figure shows the saturation throughput achieved by each MAC scheme for different number of stations in the channel with data payload of 1000 bytes. The channel error was also assumed to be negligible. In each of these cases, simulation was run until the relative error was less than 1%, at 95% confidence level. The confidence intervals obtained are shown in the figure. As seen from the figure, the saturation throughput offered by the proposed MAC increases with increasing number of stations in the channel and becomes superior in networks with more than 5 stations. The IEEE 802.11e EDCF saturation throughput decreases with increasing number of stations. The main reason is that as the number of stations increases there will be an increase in the number of possible collisions in the channel. The different behavior of the IEEE 802.11 DCF and the IEEE 802.11e EDCF is due to the difference in the contention window (CW) parameter settings. The saturation throughput of the IEEE 802.11 DCF is smaller than in IEEE 802.11e EDCF because CWmin (equal 31) of the DCF is higher compared to EDCF CWmin (equal 7). A different method is also used to decrement the backoff counter.

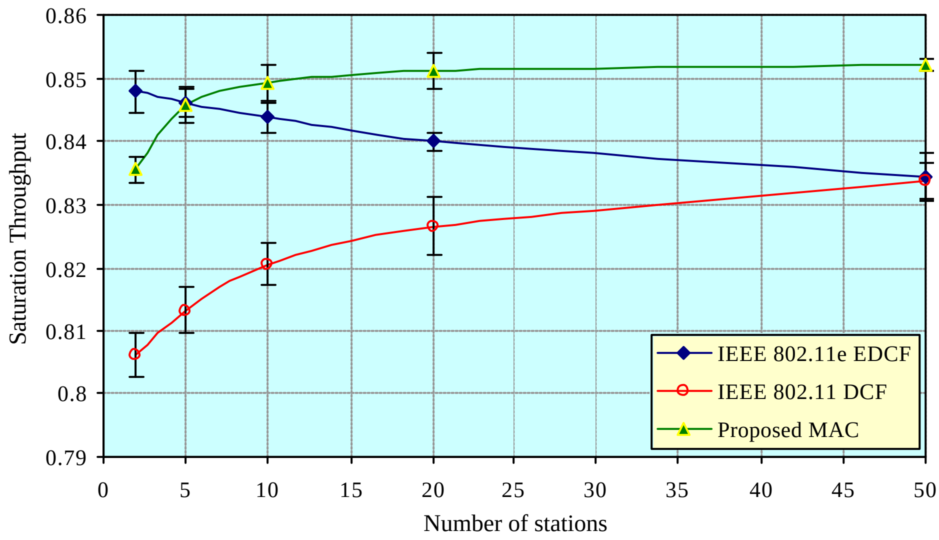


Figure 6.21 Saturation throughputs of the MAC schemes with data payload of 1000 bytes

6.8.2.2 Throughput achieved with channel error

In this case throughputs achieved by each MAC scheme are obtained from the number of packets served within a given period of time. First, each MAC scheme was simulated with perfect channel hence there would be no transmission errors. Then, the MAC schemes were simulated with the presence of channel errors, for $BER_B = 10^{-4}$ and $BER_G = 10^{-10}$. In each case the packet lengths were kept constant, with a payload length of 1000 bytes and the station short retry count (SSRC) and station long retry count (SLRC) kept at 7 and 5, respectively. The number of stations in the channel was also 20 for each MAC scheme. The results obtained are shown in figure 6.22. In each of these cases, simulation was run until the relative error was less than 5%, at 95% confidence level. The confidence intervals obtained are shown in the figure. One can see that when the channel is not perfect, the performance of any MAC scheme goes down significantly, due to packet retransmissions caused by transmission errors. On comparison to the performance of other MAC schemes, it is seen that all the MAC schemes perform equally, though there are small differences.

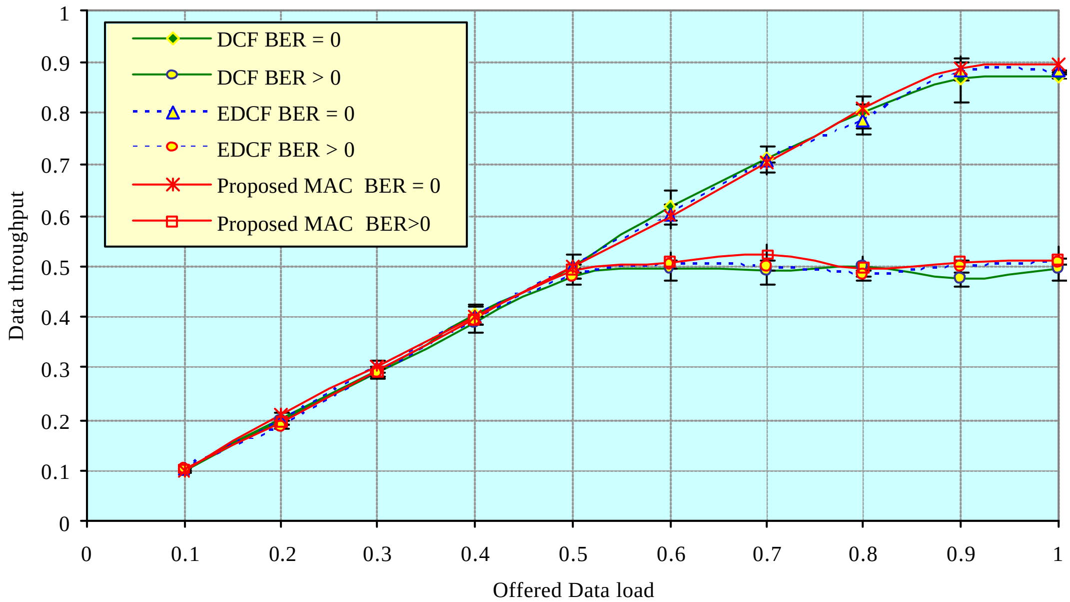


Figure 6.22 Throughput of the MAC schemes with presence of channel error, where BER = 0 means that $BER_G = BER_B = 0$, and BER > 0 means that $BER_G = 10^{-10}$ and $BER_B = 10^{-4}$.

6.8.2.3 Service Differentiation achieved by the MAC schemes

Service differentiation achieved by each MAC scheme was measured assuming that traffic was generated by ON/OFF voice sources. Each station in the channel generated two streams of such traffic. In EDCF and the proposed MAC scheme, one stream of the ON/OFF voice was assigned to P1 and the other to P2. As DCF does not support priorities, the two voice streams were combined according to the arrival times of the packets and served as one stream. The results obtained are shown in Figures 6.23, 6.24, and 6.25. The figures show the average delay accompanied with each voice stream with varying number of stations in the channel. In each of these cases, the simulation was run until the relative error was less than 10%, at 95% confidence level. The confidence intervals obtained are shown in the figures. As seen from the figures average delay of each stream increases with increasing number of stations in the channel. This is quite reasonable as increasing the number of stations introduces more traffic in the channel hence, more collisions and the average delays of the voice streams go up.

Figures 6.23 and 6.24 show that the P1 voice stream of EDCF and the proposed MAC scheme has smaller delay compared to that of P2 voice stream. This confirms that the EDCF and the proposed MAC scheme achieve service differentiation among different priorities of voice streams. However, Figure 6.25 shows that the both voice streams have the same delay so that service differentiation is not achieved in the DCF protocol.

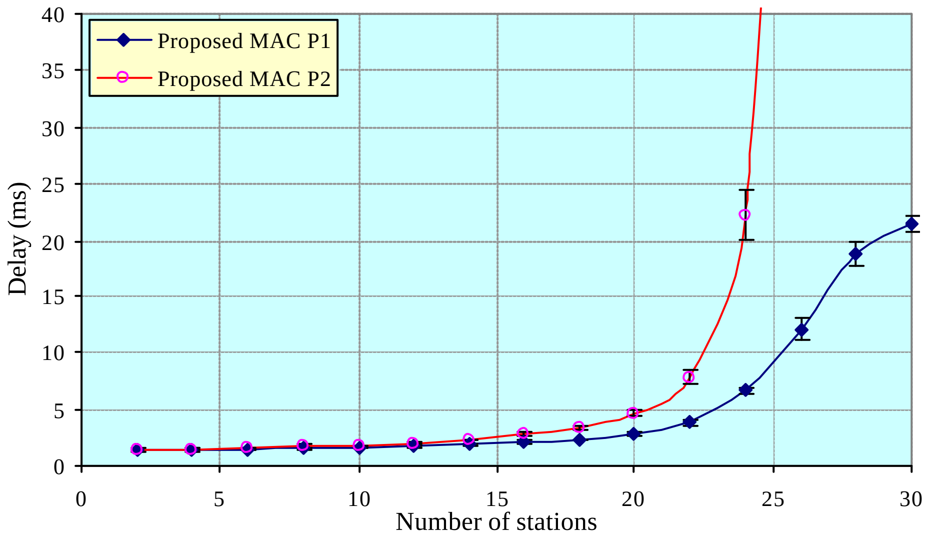


Figure 6.23 Service Differentiation achieved for different priority voice streams in the proposed MAC scheme

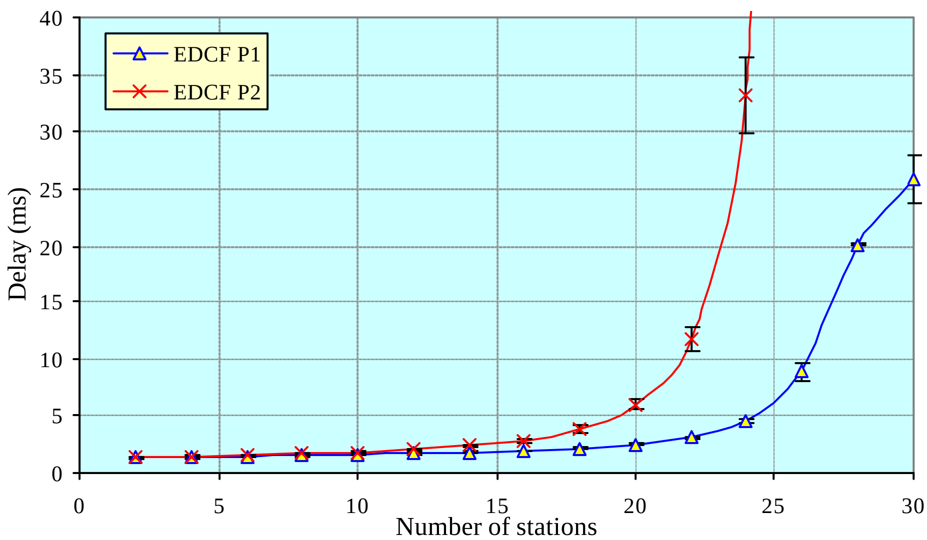


Figure 6.24 Service Differentiation achieved for different priority voice streams IEEE 802.11e EDCF

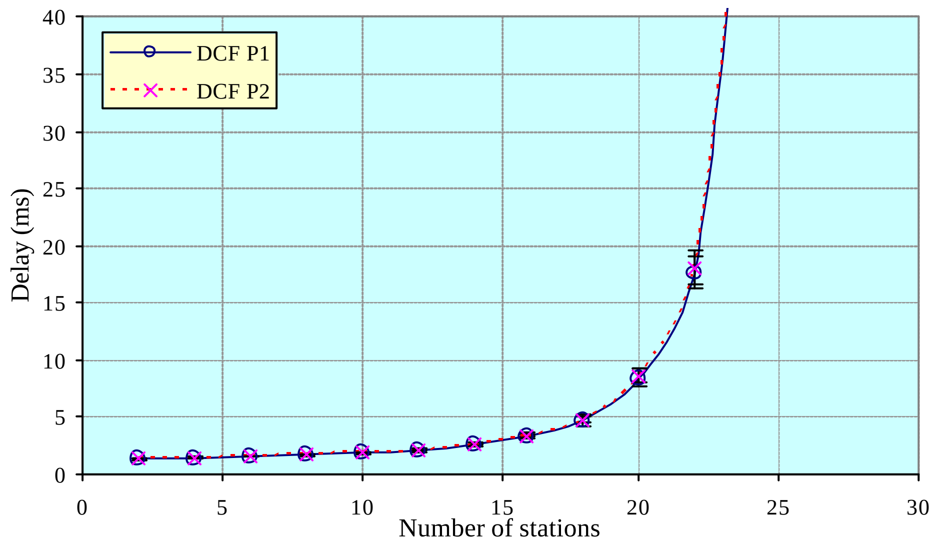


Figure 6.25 Service Differentiation achieved for different priority voice streams IEEE 802.11 DCF

Simulation was also done to show the service differentiation effects on the data throughput of each class. In the simulation each priority (P1 &P2) class generated data traffic according to a given load. In the simulation, number of stations in the channel was set to 20, data payload was set to 1000 bytes, and transmission error was assumed to be negligible. The results obtained are shown in figures 6.26, 6.27, and 6.28.

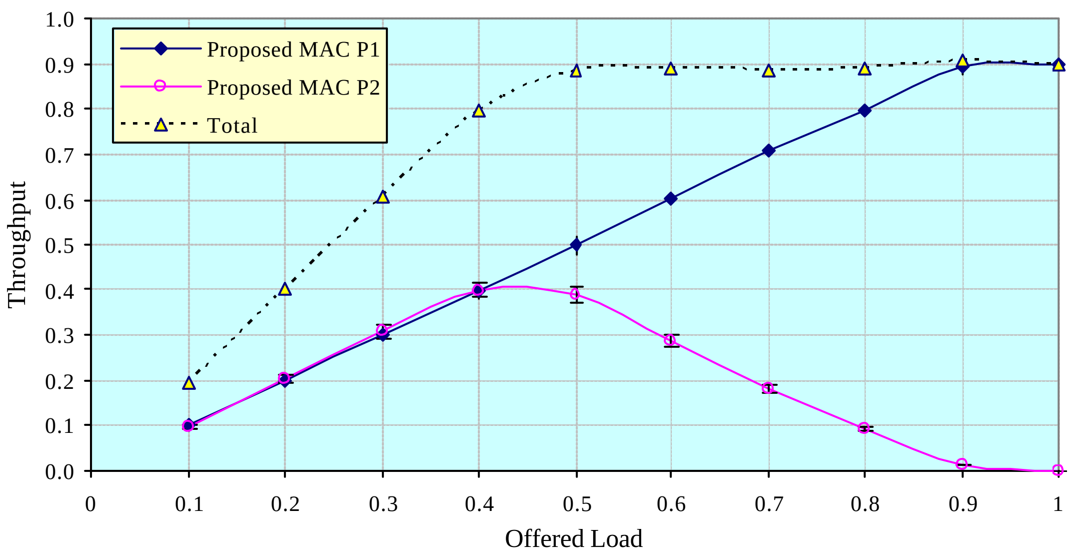


Figure 6.26 Data throughput achieved for each priority class of the proposed MAC scheme

In each of these cases, the simulation was run until the relative error was less than 5%, at 95% confidence level. The confidence intervals obtained are shown in the figures. Figure 6.26 shows that P1 of the proposed MAC scheme obtains the required bandwidth until the channel is saturated and P2 data only takes up any left over bandwidths. This means that if there is any higher priority data that needs to be transmitted, then the lower priority data does not have a chance to access the channel until all higher priority data is served. Hence, higher priority data packets always win compared to lower priority data packets in the proposed MAC scheme.

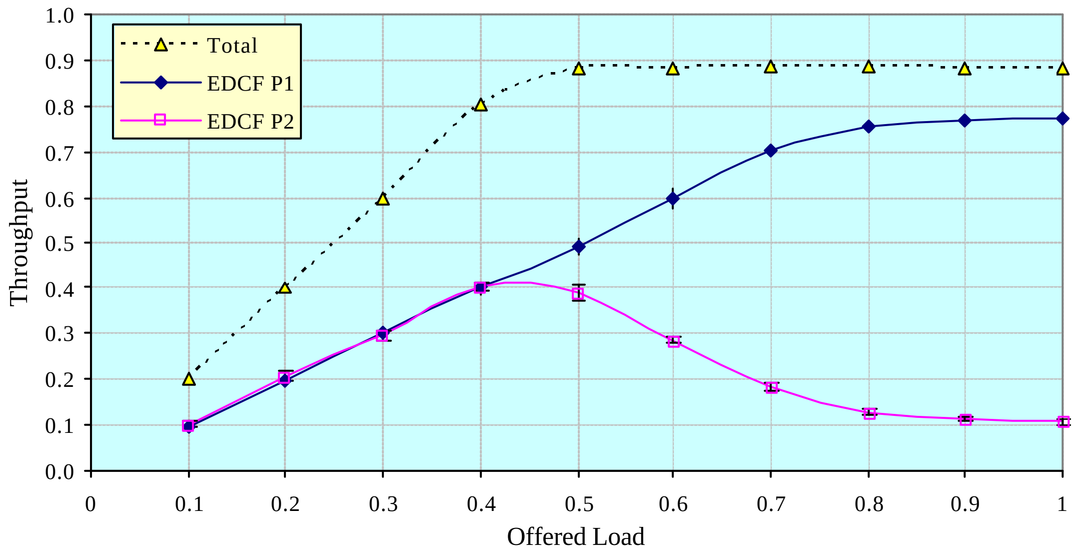


Figure 6.27 Data throughput achieved for each priority class of the IEEE 802.11e EDCF

Figure 6.27 shows that P1 of the IEEE 802.11e EDCF obtains a higher bandwidth compared to the lower priority traffic, P2. The figure also shows that when the channel gets saturated most of the bandwidth is occupied by P1. However, small amounts of bandwidth is still occupied by P2. This means that higher priority traffic has a higher chance to access the channel but lower priority traffic also has a chance to access the channel even if there is some higher priority traffic present in the channel.

Figure 6.28 shows the results obtained for IEEE 802.11 DCF. The figure shows that both priority classes of traffic occupy equal amounts of channel space so that no priority has a higher chance to access the channel. This means that service differentiation among priority classes is not supported in the IEEE 802.11 DCF.

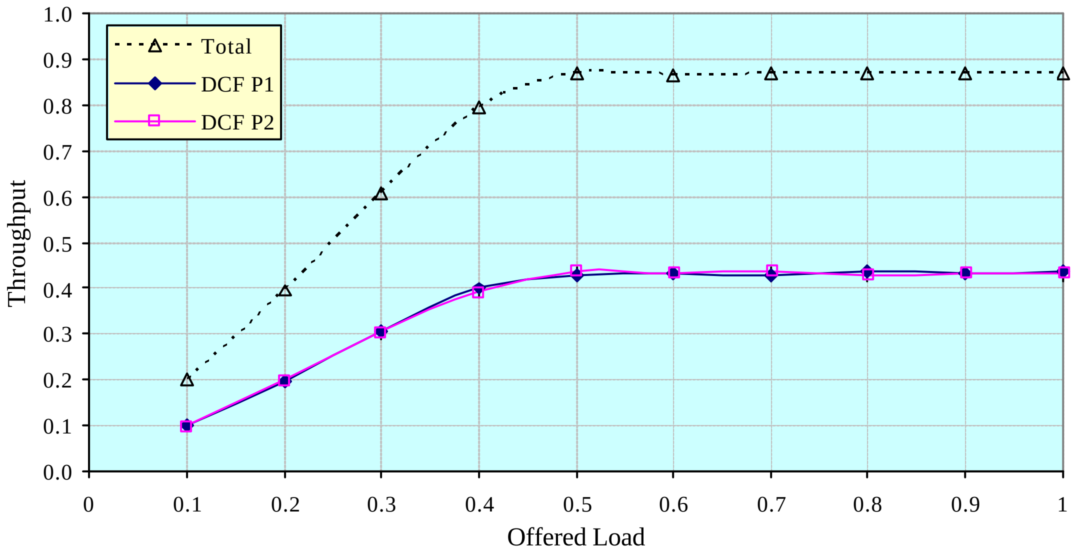


Figure 6.28 Data throughput achieved for each priority class of the IEEE 802.11 DCF

6.8.2.4 Performance of the MAC schemes with voice traffic

In this part, simulation was run assuming two types of voice traffic: Continuous Bit Rate (CBR) and ON-OFF voice traffic. In CBR, each station generates traffic at a rate of 32 kb/s with a packet size of 160 bytes. Thus the inter-packet time is 40 ms and the duration of the voice frame (including MAC and PHY header) is 840 μs. In the ON-OFF type, the voice traffic is modeled using a source with exponentially distributed on and off periods with an average of 300 ms each. As in CBR, traffic is generated during the on periods at a rate of 32 kb/s with a packet size of 160 bytes. Hence as before, the inter-packet time was 40 ms and the duration of the voice frame (including MAC and PHY headers) was 840 μs.

In each case, the simulation was run assuming varying number of stations in the BSS and calculating the delay accompanied with voice traffic. The voice traffic was the only traffic generated by the stations and it was given the highest priority. The results obtained for the MAC schemes are shown in Figure 6.29. In each of these cases, simulation was run until the relative error was less than 10%, at 95% confidence interval. The confidence intervals obtained are shown in the figure.

The average delay shown is the average duration between voice packet generation at the station and the receipt of the packet’s ACK. The figure shows that, as expected, the CBR traffic experiences more time delay in each MAC scheme compared to ON-OFF. This is because in CBR there is more voice packets per station in a given period of time. The difference in the delay of the CBR voice traffic between EDCF and DCF is due to the usage of different CW parameters in each scheme, and also due to the difference in the method used to decrement the backoff counter. The figure also shows that EDCF supports only 28 CBR voice stations with bonded voice delay but the proposed MAC scheme supports about 30 stations. So the performance of the proposed MAC scheme is better compared to EDCF and obviously much better compared to DCF.

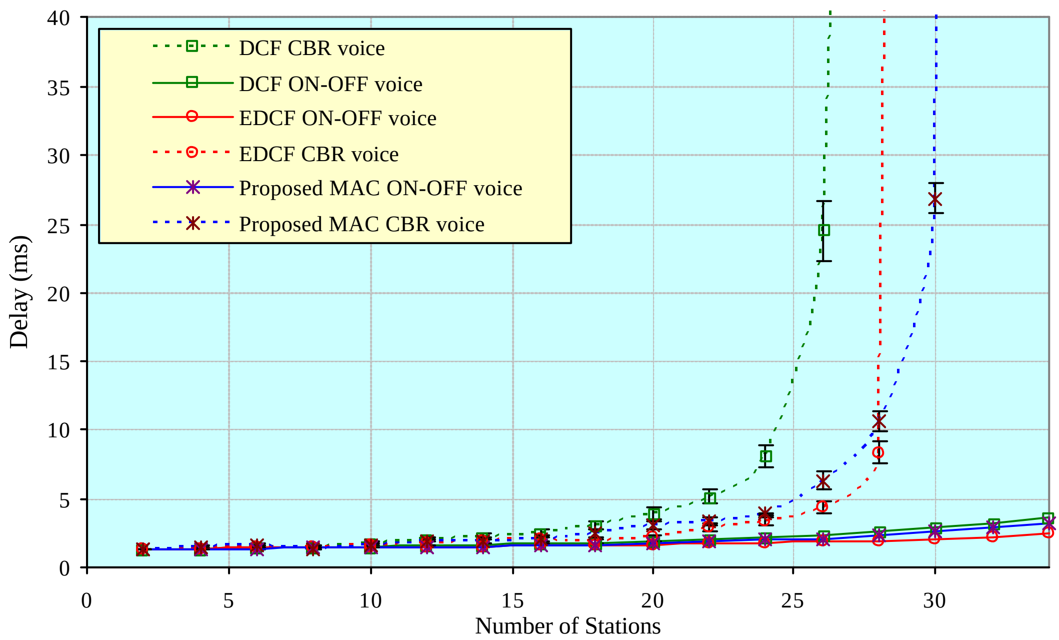


Figure 6.29 Average Delay of the voice packets

Simulation was also done to see the performance of the ON/OFF voice traffic in the presence of background data. In this simulation every station generated ON-OFF with highest priority. In addition, stations also generated background data traffic as in the method given in section 6.2. The data payload was kept constant in every case with 1000 bytes. In the EDCF and the proposed MAC scheme, the data traffic was given priority two so that the data traffic had lower priority than the voice stream. But in DCF, priority is not supported so the voice stream and the data traffic was combined according to the arrival time of packets and served as a single priority.

Simulations were run for 2 different data loads, DL = 0 and DL = 0.3, and the average voice delay is shown in Figure 6.30. In each of these cases, simulation was run until the relative error was less than 10%, at 95% confidence level. The confidence intervals obtained are shown in the figure. As expected, the figure shows that the average voice packet delay increases with an increase in the data load. But DCF has more voice delay compared to other schemes as a result of not supporting service differentiation. The proposed MAC scheme and EDCF have about the same voice delay for given a number of stations. Hence, both the schemes perform equally well.

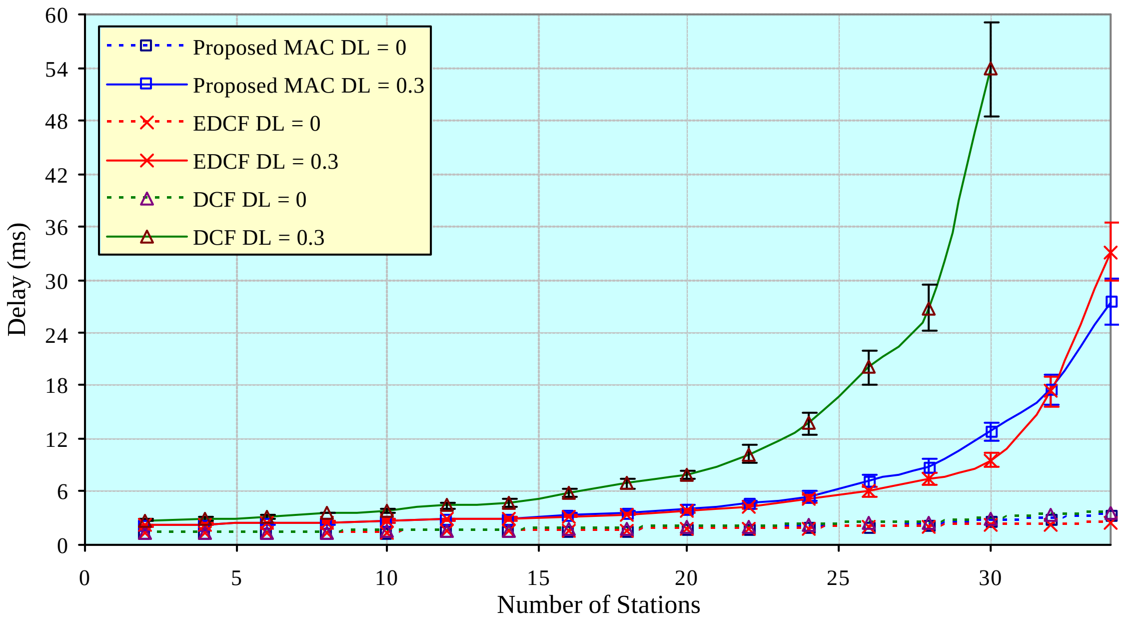


Figure 6.30 Performance of ON/OFF voice with presence of background data traffic

6.9 Conclusion

In this chapter, a performance evaluation study of the proposed MAC scheme as well as the IEEE 802.11 DCF, and IEEE 802.11e EDCF MAC schemes, was presented by means of stochastic simulation. In the chapter, simulation models used for the MAC schemes and the method used to analyze the output data of the simulations were discussed. Then the results were presented.

First, the results obtained from the mathematical model were compared with the results obtained by simulation. It showed that the mathematical model gives results dose to that obtained by simulation.

Next, the simulation results obtained for the different MAC schemes were compared. The results obtained for the saturation throughput showed that in all the schemes saturation throughput increases with an increase in data packet payload. Contrary to that, the number of stations in the BSS does not significantly effect the saturation throughput. The results obtained for service differentiation showed that only the proposed MAC scheme and the EDCF scheme can be used for service differentiation. Overall, the results showed that the proposed MAC scheme and the EDCF perform equally in all the cases and supports QoS in wireless LAN environment but the IEEE 802.11 DCF protocol does not support QoS, while the proposed MAC scheme does.

7.0 Conclusions and Future Work

This chapter provides conclusions to the work done in this thesis. This chapter also includes suggestion for possible future work that could be done to expand the work presented in this thesis.

7.1 Conclusions

Mobile communications is evolving due to the recent technological achievements in wireless networking and portable devices, like notebook computers. Today, wireless networks exist in many forms, providing different types of services in a range of local, wide area and global coverage. Improvements in both wireless and wired network technologies, like the emergence of VoIP, have resulted in the means for providing multimedia services to mobile users. Multimedia applications impose requirements on communication parameters, such as data rate, drop rate, delay and jitter. Combining delay sensitive multimedia services like voice, with streams of data traffic over the same backbone creates lots of problems. Today, network managers are looking for a practical solution to provide the QoS required for these applications.

Although, the IEEE 802.11 standard is the most widely used WLAN standard today, it still has problems with providing QoS required for multimedia services using distributed methods. As a result, distributed methods for implementing QoS in IEEE 802.11 standard via service differentiation have been proposed by several researchers. On the basis of this research, the IEEE 802.11 Task Group E is currently defining enhancements to IEEE 802.11 standard, called IEEE 802.11e, which introduces new methods to support service differentiation among different priority traffic by using distributed method.

As IEEE 802.11e is still not standardized, the main aim of this thesis was to develop a MAC scheme that supports QoS in wireless LANs. In this thesis, a new distributed MAC scheme was proposed to support QoS in wireless LANs. In the scheme, stations use CSMA for channel access, with collisions between stations being resolved by sending a set of beacons in a predefined manner, and *virtual collisions* are resolved by a scheduler inside the stations.

The proposed MAC scheme was analyzed mathematically and results obtained were validated by simulation. In the mathematical model, a maximum of two priority traffic classes were considered to simplify the model. The model estimated the average delay experienced by priority one and two traffic packets under different conditions. It was seen that the results obtained by the mathematical model were in very good agreement with the simulation results obtained.

In addition, a performance evaluation study of the proposed MAC scheme as well as the IEEE 802.11 DCF, and IEEE 802.11e EDCF MAC schemes were done by means of stochastic simulation. In this study, different performance measures of these MAC schemes were evaluated. These performance measures include: the saturation throughput, data throughput of these MAC schemes under different channel conditions, service differentiation achieved, and performance of the MAC schemes with real-time traffic.

Both the simulation and analytical results of the proposed MAC scheme showed that the delay experienced by the higher priority traffic had less average delay compared to the lower priority traffic. This means that the proposed MAC scheme was capable of supporting QoS in wireless LANs, by means of service differentiation. The simulation results obtained for the current IEEE 802.11 DCF scheme showed that the scheme was unable to support QoS by using distributed methods but the upcoming IEEE 802.11e EDCF scheme was capable to support QoS.

The results obtained for saturation throughput of the proposed MAC scheme showed that the saturation throughput was very much dependent on the packet payload. However it did not depend much on the number of stations in the channel. The results also showed that the proposed MAC scheme achieved the same level of saturation throughput compared to the current IEEE 802.11 DCF and the upcoming IEEE 802.11e EDCF schemes.

The simulation results obtained for the data throughput of the considered MAC schemes showed that the MAC schemes achieved the same level of data throughput. However, throughput of these MAC schemes went down under bad channel conditions such as fading.

The simulation done to show the service differentiation effects on the data throughput of the different priority classes of the proposed MAC scheme showed that if any higher priority data needed to be transmitted, then the lower priority data did not have a chance to access the channel until all the higher priority data were served. However in IEEE 802.11e EDCF scheme, lower priority traffic had a small chance to access the channel even if there were some higher priority traffic present in the channel.

When serving delay sensitive real-time data like voice with other non-real-time data, the results showed that the current IEEE 802.11 DCF scheme had more real-time data delay compared to the proposed MAC scheme. This was because the current IEEE 802.11 DCF scheme does not support service differentiation. The results also showed that the upcoming IEEE 802.11e EDCF scheme had almost the same level of real-time data delay compared to the proposed MAC scheme.

Overall, the results showed that the proposed MAC scheme supports QoS by means of service differentiation in the wireless LANs. However the current IEEE 802.11 DCF scheme does not support QoS. In general, the results showed that the proposed MAC scheme performs equally or better compared to the current IEEE 802.11 DCF scheme in every single case considered. It was also found that the proposed MAC scheme performs equally well as the upcoming IEEE 802.11e EDCF scheme, in every single case considered here.

7.2 Future Work

For wireless LANs, well-defined coverage areas simply do not exist. Propagation characteristics are dynamic and unpredictable. Small changes in position or direction may result in dramatic differences in signal strength. As a result, if the BSSs are not physically very far apart, then, two or more BSSs may overlap and cover the same geographical area. In the thesis, it had been assumed that the BSSs are far apart so that there is no interference from neighboring BSSs. It would be more practical to perform an evaluation study of the MAC schemes by considering the interference from neighboring BSSs.

In the proposed MAC scheme, ID numbers of the stations in the BSS play an important role in resolving collisions between stations. The ID numbers are given to the stations when they initially join the BSS. In this thesis, only one method to join the new stations to the BSS is explained. Further research can be done to find other feasible methods to join new stations to the BSS.

The mathematical model derived for the proposed MAC scheme had a few assumptions like infinite storage buffer and errorless channel to simplify the model. However, it would be more practical to consider finite buffer size and transmission errors of the channel in the mathematical model. Hence, a more practical mathematical model could be derived for the proposed MAC scheme by relieving these assumptions.

As noted in the literature review, the hidden and exposed terminal problems have been identified in wireless systems as causing problems in collision detection. In wired networks, this has not been a problem because in wired networks, collisions can be detected by checking the voltage on the wire before transmitting. The hidden and exposed stations in WLANs cause problems and very much decrease the performance of the LANs. In the simulation work, the hidden and exposed stations problems were ignored to simplify the simulation model. It would be very interesting to further investigate the performance of the considered MAC scheme in the presence of hidden and exposed stations in the channel.

It was shown by using simulation that the proposed MAC scheme worked well and achieved service differentiation in WLANs. However, in the simulations several assumptions are made and it is impossible to develop a simulation model that will perfectly represent the real practical situations. Therefore it would be very worthwhile to implement the proposed MAC scheme on hardware to find the real performance of the scheme in a practical environment.

References

[Aad and Castelluccia 2001] - I. Aad and C. Castelluccia, "*Differentiation mechanism for IEEE 802.11*," INFOCOM 2001. Proceedings. IEEE, Volume: 1, 2001, Page(s): 209 –218.

[Abramson 1970] - N. Abramson, " The ALOHA systems – Another Alternative for computer communications," *AFIPS Conf. Proc., Fall Joint Comput. Conf.*, 37, 281-285, 1970.

[Barry et al 2001] - M. Barry, A. T. Campbell, A. Veres, " *Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control*," Selected Areas in Communications, IEEE Journal, Volume: 19, Issue: 10, October 2001, Page(s): 2081–2093.

[Bertsekas and Gallager 1992] - D. Bertsekas, R. Gallager, "*Data Networks*", Prentice Hall, 1992.

[Bhargava et al. 1981] - V. K. Bhargava et al., "Digital Communications by Satellite – Modulation, Multiple Access and Coding," Wiley, New York, 1981.

[BRAN 2000] - ETSI TR 101 683 V.1.1.1, Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; System Overview, 2000.

[Buchholz et al. 1991] - D. Buchholz, P. Odlyzko, M. Taylor, and R. White, "Wireless In-Building Network Architecture and Protocols," *IEEE Network Mag.*, 5, No. 6, 31-38, 1991.

[Chen and Yeh 2002] - C. Chen, Jing-Yuan Yeh **"***Support of multimedia services with the IEEE 802.11 MAC protocol*," in Communications, 2002. ICC 2002. IEEE International Conference, vol: 1, 2002, Page(s): 600 –604.

[Cocker et al. 2001] - N. Cocker, T. Moors, M. Veeraraghavan, *"Support of voice services in IEEE 802.11 wireless LANs,"* in INFOCOM 2001. Proceedings. IEEE, vol: 1, 2001, Page(s): 488 –497.

[Crow et al. 1997] - B. P. Crow, I. Widjaja, J. G. Kim and P. T. Sakai, "IEEE 802.11 Wireless Local Area Networks," in IEEE Communication Magazine, Volume: 35 Issue: 9, Sept. 1997, Page(s):116-126.

[Deng and Chang 1999] - D. Deng, R. Chang, "A Priority Scheme for IEEE 802.11 DCF Access Method," IEICE Trans. Commun., E82-B(1), January 1999.

[Freeburg 1991] - T. Freeburg, "A New Technology for High Speed Wireless Local Area Networks," *Proc. IEEE Workshops on Wireless LANs*, Worcester, MA, 127-139, 1991.

[Gilbert 1960] - E. Gilbert, "Capacity of a Burst Noise Channel," Bell Syst. Tech.J., vol. 39, Sept. 1960, pp. 1253-1266.

[Gross and Harris 1974] - D. Gross, C. M. Harris, "*Fundamental of Queueing Theory*", Johen Wiley & Sons, 1974.

[Hammond and O'Reilly 1988] - J. L. Hammond, J. P. O'Reilly, "Performance Analysis of Local Area Networks," Addison-Wesley, 1988.

[Hopper et al. 1986] - A. Hopper, S. Temple, R. Williamson, "Local Area Network Design," Addison-Wesley, 1986.

[IEEE 802.11 1999] - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Standard 802.11, June 1999.

[Kleinrock 1975] - L. Kleinrock, "*Queueing Systems Volume I: Theory*", Johen Wiley & Sons, 1975.

[Kleinrock 1976] - L. Kleinrock, "*Queueing Systems Volume II: Computer Applications*", Johen Wiley & Sons, 1976.

[Kleinrock and Gail 1996] - L. Kleinrock, R. Gail *"Queueing Systems Problems and Solutions"*, Johen Wiley & Sons, Inc, 1996.

[Krishnakumar and Sobrinho 1996] - A.S. Krishnakumar and J.L. Sobrinho, "*Distributed Multiple Access Procedures to Provide Voice Communications over IEEE 802.11 Wireless Networks*," GLOBECOM '96, Volume: 3, 1996, Page(s): 1689 –1694.

[Law and Kelton 2000] - A. M. Law, W. D. Kelton, "Simulation Modeling and Analysis," MacGraw-Hill, New York, 2000.

[Mangold et al. 2002] - S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor, "IEEE 802.11e Wireless LAN for Quality of Service," European Wireless, vol. 1, pp. 32-39, February 2002.

[Marcus 1991] - M. Marcus, "Regulatory Policy Considerations for Radio Local Area Networks," *Proc. IEEE Workshop on Wireless LANs*, Worcester, MA, 42-48, 1991.

[Pahlavan and Levesque 1995] - K. Pahlavan, A. H. Levesque, "Wireless Information Networks," John Wiley and Sons, 1995.

[Pawlikowski 1990] – K. Pawlikowski, "Steady-State Simulation of Queueing processes: A Survey of Problems and Solutions," ACM Computing Surveys 22 (2): 123-170, 1990.

[Petrick 1997] - A. Petrick, "*Voice Services over 802.11 WLAN*," in IIC, Taipei, Conference Proceedings, 1997, Page(s): 61-63.

[Rappaport 1996] – T. S. Rappaport, "Wireless Communications: Principles and Practice," Prentice Hall, 1996.

[Roberts 1975] - L. G. Roberts, " ALOHA Packet System with and without Slots and Capture," ACMSIGCOM Comput. Commun. Rev., 5, 28-42, 1975.

[Rom and Sidi 1990] - R. Rom, M. Sidi, "Multiple Access Protocols Performance and Analysis," Springer-Verlag, 1990.

[Santamaria and Lopez 1993] - S. Santamaria, F. J. Lopez-Hernandez, "Wireless LAN Systems," Artech House, 1993.

[SpectraLink 2000] - *"SpectraLink Voice Priority: Quality of service for voice traffic on wireless LANs*," SpectraLink, 2000.

[Stark and Woods 2002] - H. Stark, J. W. Woods, "*Probability and Random Processes with Application to Signal Processing*", Prentice Hall, 2002.

[Toh 1997] - C-K. Toh, "*Wireless ATM and Ad-Hoc Networks: Protocol and Architectures*," Kluwer Academic Publishers, 1997.

[Tuc 1991] - B. Tuc, "An ISM band Spread Spectrum Local Area Networks: WaveLAN," *Proc. IEEE Workshop on Wireless LANs*, Worcester, MA, 103-111, 1991.

[Visser and Zarki 1995] - M. A. Visser and M. El Zarki, "*Voice and data transmission over an 802.11 wireless networks*," in Proceedings of PIMRC'95, Toronto, Canada, September 1995.

[Walrand 1991] - J. Walrand, *Communication Networks: A First Course.* Aksen Associates & Publishers, Boston, 1991.

[Wesolowski 2002] - K. Wesolowski, "Mobile Communication Systems," John Wiley and Sons, 2002.

Appendix A – MATLAB codes used in the simulation

The appendix A contains the codes used to simulate the proposed MAC, IEEE 802.11 DCF and IEEE 802.11e EDCF. Appendix A is included in the attached CD-ROM as a Microsoft Word document. The CD-ROM also contains the exact MATLAB codes used for simulation of the proposed MAC, IEEE 802.11 DCF and IEEE 802.11e EDCF. The contents of the CD-ROM are shown in Figure A.1 and explained below.

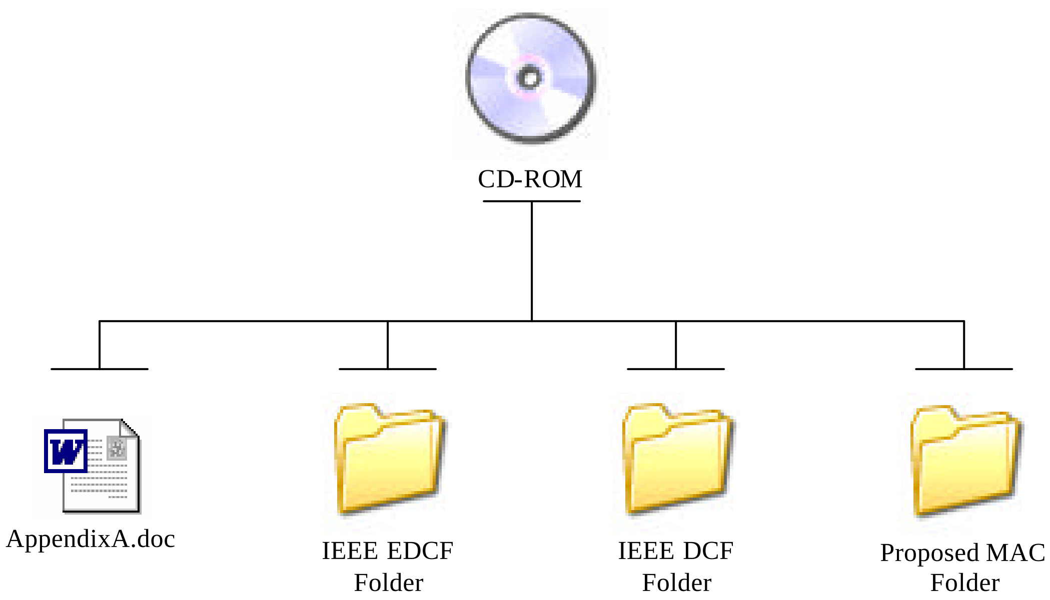


Figure A.1 Contents of the attached CD-ROM

- **Appendix A:** This is a Microsoft Word document containing all the codes used to simulate the above mentioned MAC schemes. This file includes the referred appendices: Appendix A.1 to Appendix A.5.
- **IEEE EDCF Folder:** Contains MATLAB files used to simulate IEEE 802.11e EDCF scheme. The program can be run by calling IEEE_EDCF.m file. The parameters can be set to the appropriate values by changing the codes in the IEEE_EDCF.m file.
- **IEEE DCF Folder:** Contains MATLAB files used to simulate IEEE 802.11 DCF scheme. The program can be run by calling IEEE_DCF.m file. The parameters can be set to appropriate values by changing the codes in the IEEE_DCF.m file.

- **Proposed MAC Folder:** Contains MATLAB files used to simulate the proposed MAC scheme. The program can be run by calling NEWMAC.m file. The parameters can be set to appropriate values by changing the codes in the NEWMAC.m file.