PERFORMANCE MODELLING AND ENHANCEMENT OF WIRELESS COMMUNICATION PROTOCOLS

PERIKLIS CHATZIMISIOS

A thesis submitted in partial fulfillment of the requirements of Bournemouth University for the degree of Doctor of Philosophy

December 2004

Bournemouth University

PERFORMANCE MODELLING AND ENHANCEMENT OF WIRELESS COMMUNICATION PROTOCOLS

Periklis Chatzimisios

ABSTRACT

In recent years, Wireless Local Area Networks (WLANs) play a key role in the data communications and networking areas, having witnessed significant research and development. WLANs are extremely popular being almost everywhere including business, office and home deployments. In order to deal with the modern wireless connectivity needs, the Institute of Electrical and Electronics Engineers (IEEE) has developed the 802.11 standard family utilizing mainly radio transmission techniques, whereas the Infrared Data Association (IrDA) addressed the requirement for multipoint connectivity with the development of the Advanced Infrared (AIr) protocol stack.

This work studies the collision avoidance procedures of the IEEE 802.11 Distributed Coordination Function (DCF) protocol and suggests certain protocol enhancements aiming at maximising performance. A new, elegant and accurate analysis based on Markov chain modelling is developed for the idealistic assumption of unlimited packet retransmissions as well as for the case of finite packet retry limits. Simple equations are derived for the throughput efficiency, the average packet delay, the probability of a packet being discarded when it reaches the maximum retransmission limit, the average time to drop such a packet and the packet inter-arrival time for both basic access and RTS/CTS medium access schemes. The accuracy of the mathematical model is validated by comparing analytical with OPNET simulation results. An extensive and detailed study is carried out on the influence on performance of physical layer, data rate, packet payload size and several backoff parameters for both medium access mechanisms. The previous mathematical model is extended to take into account transmission errors that can occur either independently with fixed Bit Error Rate (BER) or in bursts. The dependency of the protocol performance on BER and other factors related to independent and burst transmission errors is explored. Furthermore, a simple-to-implement appropriate tuning of the backoff algorithm for maximizing IEEE 802.11 protocol performance is proposed depending on the specific communication requirements. The effectiveness of the RTS/CTS scheme in reducing collision duration at high data rates is studied and an all-purpose expression for the optimal use of the RTS/CTS reservation scheme is derived. Moreover, an easy-to-implement backoff algorithm that significantly enhances performance is introduced and an alternative derivation is developed based on elementary conditional probability arguments rather than bi-dimensional Markov chains. Finally, an additional performance improvement scheme is proposed by employing packet bursting in order to reduce overhead costs such as contention time and RTS/CTS exchanges. Fairness is explored in short-time and long-time scales for both the legacy DCF and packet bursting cases.

AIr protocol employs the RTS/CTS medium reservation scheme to cope with hidden stations and CSMA/CA techniques with linear contention window (CW) adjustment for medium access. A 1-dimensional Markov chain model is constructed instead of the bi-dimensional model in order to obtain simple mathematical equations of the average packet delay. This new approach greatly simplifies previous analyses and can be applied to any CSMA/CA protocol. The derived mathematical model is validated by comparing analytical with simulation results and an extensive AIr packet delay evaluation is carried out by taking into account all the factors and parameters that affect protocol performance. Finally, suitable values for both backoff and protocol parameters are proposed that reduce average packet delay and, thus, maximize performance.

DEDICATION

Αφιερώνεται στην γυναίκα μου Ρένα,
στην μητέρα μου Σοφία-Γιολάντα,
στον πατέρα μου Ιωάννη
και στον αδελφό μου Στέφανο

PUBLICATIONS RESULTING FROM THESIS

I. Journal peer-reviewed publications

[1] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas, "IEEE 802.11 Wireless LANs: Performance Analysis And Protocol Refinement", *EURASIP Journal on Wireless Communications and Networking,* accepted (to appear), 2004.

[2] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas, "Performance analysis of the IEEE 802.11 MAC protocol for Wireless LANs", *Wiley International Journal of Communication Systems,* accepted (to appear), 2004.

[3] P. Chatzimisios and A. C. Boucouvalas, "Packet Delay Analysis of the Advanced Infrared (AIr) CSMA/CA MAC Protocol in Optical Wireless LANs", *Special Issue in Indoor Optical Wireless Communication Systems and Networks, Wiley International Journal of Communication Systems,* accepted (to appear), 2004.

[4] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas, "Effectiveness of the RTS/CTS Handshake in IEEE 802.11a Wireless LANs", *IEE Electronic Letters,* vol. 40, no. 14, pp.915-916, July 2004.

[5] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas, "Influence of BER on the Performance of IEEE 802.11 DCF", *IEE Electronic Letters,* vol. 39, no. 23, pp.1687-1689, Nov. 2003.

[6] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas, "Packet Delay Analysis of the IEEE 802.11 MAC Protocol", *IEE Electronic Letters,* vol.39, no. 18, pp.1358-1359, Sept. 2003.

II. International conference peer-reviewed publications

[7] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas, "Optimisation of RTS/CTS handshake in IEEE 802.11 Wireless LANs for maximum performance", in *Proceedings of the IEEE Global Telecommunications Conference (Globecom 2004),* Dallas Texas, USA, pp. 270-275, 29 Nov.-3 Dec., 2004.

[8] V. Vitsas, P. Chatzimisios, A. C. Boucouvalas, P. Raptis, K. Paparrizos and D. Kleftouris, "Enhancing performance of the IEEE 802.11

Distributed Coordination Function via packet bursting", **invited paper** in *Proceedings of the IEEE Global Telecommunications Conference (Globecom 2004)*, Dallas Texas, USA, pp. 245-252, 29 Nov.- 3 Dec., 2004.

[9] P. Chatzimisios A. C. Boucouvalas and V. Vitsas, "IEEE 802.11 WLANs: Performance analysis in presence of bit errors", *Proceedings of the Communication Systems, Networks & Digital Signal Processing International Symposium (CSNDSP 2004)*, Newcastle, UK, pp. 355-358, July 2004.

[10] P. Chatzimisios and A. C. Boucouvalas, "Improving performance through optimization of the RTS/CTS mechanism in IEEE 802.11 Wireless LANs", *Proceedings of the Communication Systems, Networks & Digital Signal Processing International Symposium (CSNDSP 2004)*, pp. 363-366, Newcastle, UK, July 2004.

[11] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas, "Performance Analysis of IEEE 802.11 DCF in the Presence of Transmission Errors", *Proceedings of the IEEE International Conference on Communications (ICC 2004)*, vol. 7, pp. 3854 - 3858, Paris, France, June 2004.

[12] P. Chatzimisios , A. C. Boucouvalas and V. Vitsas, "IEEE 802.11 Packet Delay - A Finite Retry Limit Analysis", *Proceedings of the IEEE Global Telecommunications Conference (Globecom 2003)*, vol. 2, pp. 950-954, San Francisco, USA, Dec. 2003.

[13] P. Chatzimisios , V. Vitsas and A. C. Boucouvalas, "Throughput and delay analysis of the IEEE 802.11 protocol", *Proceedings of the 5th IEEE International Workshop on Network Appliances (IWNA 2002)*, pp. 168-174, Liverpool, UK, Oct. 2002.

NOTE: Most of the above publications are available at:

http://dec.bournemouth.ac.uk/staff/pchatzimisios/publications.html

http://aetos.it.teithe.gr/~peris

ACKNOWLEDGMENTS

This work was carried out during my stay as a researcher and a part-time lecturer at Bournemouth University from October 2001 until December 2004. It would not have been possible without the help and friendship of several people.

First of all, I would like to thank my supervisor, Prof. Tony Boucouvalas, for his guidance and assistance, throughout my research project. He provided me not only with insightful discussions, but also with the necessary degree of freedom. I have learnt a lot from him.

I am very grateful to my friend Prof. Vassilis Vitsas for his constructive suggestions to improve my work and research performance. I would also like to thank Prof Antony Vafiadis, Prof. Dimitris Kleftouris and Pasxalis Raptis from Technological Education Institution of Thessaloniki for their constant encouragement and support since 1995.

I happily acknowledge my many friends in Thessaloniki and Bournemouth, Pasxalis Kamargiannis, Elsa Paneta, Angeliki Kouvidi, Aleksandra Markezini, Stefanos Georgas and Dimitris Vikeloudas for the wonderful time we spent together. I also want to thank my crazy friends-colleagues in P608, Pi Huang, Peter Voles, Piotr Stepien, Xin Qian and Tom Teng for their friendship.

Although any list of names is inevitably incomplete, I also wish to thank many people supported me in Bournemouth University and who fit into no particular category, especially: Prof. Richard Wynne, Prof. Jim Roach and Prof. Denzil Claremont, Jacqui Holmes, Frank Ball, Peter Barker , Glyn Hadley, Jenny Longster, Carolyn Mair, Rahul Premraj, Marina Miltiadou, Roman Gerodimos, Simon Crowle and Patti Davies.

I am gifted with a father, a mother and a brother who have been believing in me throughout my entire life. I thank them for their continual support and love.

Finally, my love goes to my wife Rena for the 10 years of understanding, love, patience, support and faith. She is the one and only reason for all the success I had up to date and I will have in the future.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

4PPM/VR	4-slot Pulse Position Modulation with Variable Repetition Rate encoding
ACK	Acknowledgement
AIr	Advanced Infrared
AIr MAC	AIr Medium Access Control
AIr LC	AIr Link Control
AIr LM	AIr Link Manager
AP	Access Point
ARQ	Automatic Repeat reQuest
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BSS	Basic Service Set
CA	Collision Avoidance
CAS	Collision Avoidance Slot
CDMA	Code Division Multiple Access
CFP	Contention Free Period
CP	Contention Period
CRC	Cyclic Redundancy Check
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CT	CAS Timer
CTS	Clear To Send
CW	Contention Window
CWmax	Contention Window Maximum
CWmin	Contention Window Minimum
DA	Destination Address
DCF	Distributed Coordination Function
DD	Direct Detection
DIFS	DCF Inter-Frame Space
DS	Distribution System
DSSS	Direct Sequence Spread Spectrum
EDCF	Enhanced DCF
EIFS	Extended Inter-Frame Space
EOB	End Of Burst
EOBC	End Of Burst Confirm
ESS	Extended Service Set
ETSI	European Telecommunication Standards Institute
F-bit	Final bit
FCS	Frame Check Sequence
FER	Frame Error Rate
FH	Frequency Hopping
FHSS	Frequency Hopping Spread Spectrum
FIR	Fast Infrared
FOV	Field Of View

GBN	Go-Back-N
HiperLAN	High Performance Radio LAN
I-frame	Information frame
IrDA	Infrared Data Association
IrLAP	IrDA Link Access Protocol
IM	Intensity Modulation
IR	Infrared
ISM	Industrial, Scientific and Medical
IBSS	Independent Basic Service Set
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
ISO	International Organization for Standardization
LAN	Local Area Network
L-PPM	L-slot Pulse Position Modulation
LLC	Logical Link Control
LOS	Line Of Sight
MAC	Media Access Control
MACA	Multiple Access with Collision Avoidance
MBR	Main Body
MPDU	MAC Protocol Data Unit
MSDU	MAC Service Data Unit
MT	Mobile Terminal
NAV	Network Allocation Vector
NIC	Network Interface Card
OFDM	Orthogonal Frequency Division Multiplexing
P-bit	Poll bit
P/F bit	Poll/Final bit
PA	Preamble
PAN	Personal Area Network
PC	Point Coordinator
PCF	Point Coordination Function
PDA	Personal Digital Assistant
PHY mode	Physical layer mode
PIFS	PCF Inter Frame Space
PSAP	Physical layer Service Access Point
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RH	Robust Header
RR	Repetition Rate
RR S-frame	Receive Ready Supervisory frame
RT	Reservation Time
RTS	Request To Send
S-frame	Supervisory frame
SA	Source Address
SIFS	Shortest Inter Frame Space
SIR	Serial Infrared
SNR	Signal-to-Noise Ratio
SREJ	Selective Reject

STA	Station
SYNC	Synchronisation
SW	Stop-and-Wait
TAT	Turn Around Time
TCP/IP	Transmission Control Protocol/Internet Protocol
TDMA	Time Division Multiple Access
U-frame	Unnumbered frame
VFIR	Very Fast Infrared
VTT	Virtual Transmission Time
WECA	Wireless Ethernet Compatibility Alliance
WEP	Wired Equivalent Privacy
WFCTS	Wait For CTS
WLAN	Wireless LAN
WLANA	Wireless LAN Alliance
WM	Wireless Medium
WPAN	Wireless Personal Area Network
WTT	Window Transmission Time

CHAPTER 1

Introduction

1.1 Motivation

During the past few years, the field of wireless communications has witnessed a massive development and has become one the fastest growing areas in telecommunications and networking [87][112]. Technological and regulatory progress has allowed the issues of high prices, low data rates and licensing requirements to be addressed driving the popularity of wireless devices to grow significantly. With wireless networking, regardless of where end users are, they can have network connectivity being a mouse-click away from key information and applications [100]. Recent advances in wireless technology and mobile communications have provided wireless capabilities to portable devices including palmtop computers, laptops and personal digital assistants (PDAs).

In wireless communications, radio frequencies (RF) and Infrared (IR) optical are competing transmission technologies and are being considered as complementary transmission media [2][71]. Radio is preferred when long-range or omni-directional transmission is required. Radio is also preferable when user mobility is of prime importance. Infrared is preferred when point-to-point or multipoint links of high capacity are necessary and when simple low-cost components and international compatibility are required [3][71]. Infrared links utilize low-cost components with small physical size and low power consumption. In addition, infrared spectrum is unregulated worldwide and can achieve high data rates.

The Institute of Electrical and Electronics Engineers (IEEE) has developed the 802.11 standard family [135][136][137], in order to deal with the modern wireless connectivity needs. Over the years, the IEEE 802.11 protocol has become a mature technology, achieved worldwide acceptance and turned into the dominating standard for Wireless Local Area Networks (WLANs). The IEEE 802.11a standard [137] operating on the 5 GHz radio frequency band and the IEEE 802.11b standard [136] using the 2.4 GHz frequency band, provide up to 54 Mbit/s and 11 Mbit/s data rates, respectively. The IEEE 802.11 standards include detailed specifications for both the Medium Access

Control (MAC) and the Physical Layer (PHY). It employs the contention-based Distributed Coordination Function (DCF) as the essential MAC method. DCF defines two medium access mechanisms to employ packet transmission; the default, two-way handshaking technique called basic access and the optional four-way handshaking RTS/CTS reservation scheme.

The Infrared Data Association (IrDA) was established in 1993 as a 'working group' by major industrial companies aiming to develop a set of protocol standards for infrared wireless connectivity. The resultant IrDA 1.x protocol stack specified point-to-point, short range, directed half-duplex links. IrDA 1.x is widely adopted, fully supported by popular operating systems and millions of devices are shipped every year embedding an infrared port for their wireless transfer needs. IrDA addressed the recognized need for multipoint wireless connectivity, with the development of the Advanced Infrared (AIr) protocol stack. The aim of AIr is to provide a low-cost non-directed ad-hoc IR wireless LAN supporting co-existence with IrDA 1.x point-to-point links. Thus, the AIr proposal preserves the investment in IrDA 1.x upper layer applications by replacing the physical and the link layer of the IrDA 1.x protocol stack. In order to achieve multipoint connectivity a new physical layer, the AIr PHY, is proposed that supports wide-angle infrared links providing a 'broadcast' medium for all devices within range. AIr PHY employs Repetition Rate (*RR*) coding to achieve the increased transmission range required for wireless LAN connectivity at a base data rate of 4 Mbit/s. The transmitter trades speed for range and link quality by repeating the transmitted information *RR* times in order to increase the capture probability at the receiver. With an AIr network, all devices have equal status with no 'master' controller and can join or leave the network at will. IrLAP, the IrDA 1.x link layer is divided into three sub-layers, the AIr Medium Access Control (AIr MAC), the AIr Link Manager (AIr LM) and the AIr Link Control (AIr LC) sub layers. The AIr MAC protocol is a CSMA/CA (Carrier Sensing Multiple Access with Collision Avoidance) protocol. The AIr MAC is responsible for coordinating access to the shared infrared medium and utilizes an RTS/CTS (Request To Send / Clear To Send) reservation scheme to improve performance. Following establishment of medium reservation, a 'burst' of data packets is transmitted.

The performance of wireless links may be measured by the link throughput efficiency (also known as utilization), the average packet delay, the probability of a packet being discarded when it reaches the maximum retransmission limit, the average time to drop a packet and the packet inter-arrival time. Throughput efficiency expresses the time portion of the total time the medium successfully transfers information between stations. The average delay for a successfully transmitted packet is defined to be the time interval from the time the packet is at the head of its MAC queue ready to be transmitted, until an acknowledgement for this packet is received. The drop probability and average drop time are defined respectively as the probability and the average time for a packet to be dropped when its retry limit is reached. The packet inter-arrival time is defined as the time interval between two successful packet receptions at the receiver.

All the performance metrics utilized in this work take into account all the significant factors that affect performance such as (a) the physical layer delays (b) the medium access mechanism, (c) the transmission control passing scheme, (d) the transmission errors introduced by the wireless medium and (e) the acknowledgement delays. Link layer design is very important as it must minimize physical and link layer delays and increase performance for the information transfer scenarios that will utilize the considered radio and infrared WLAN links.

1.2 Statement of the problem

Link layer design must minimize physical and link layer delays such as hardware latency, medium access and delays due to retransmissions. An efficient link layer must minimize performance loss and successfully deliver as much information as possible. In a congested wireless network, if two or more stations simultaneously initiate a transmission, a packet collision occurs and the transmissions must be reattempted, thus affecting the performance of the network. Moreover, when the channel is error-prone (when unsatisfactory channel conditions corrupt the packet at the receiver) performance degradation can be also due to transmission errors. For both the cases, the behavior of the transmitter when a corrupted packet is received at the receiver, is the same as when a packet collision occurs; the transmitter will reattempt the transmission. Therefore, the study as well as the enhancement of performance under congestion and/or transmission errors are of key importance and are addressed in this work. In wireless infrared links, a single transmission error also results in the retransmission of a large amount of information data and performance degradation. A trade-off exists between the desire to reduce the ratio of transmission overhead and the need to reduce the packet error rate in an error-prone channel. Thus, the desire for optimal information amount that simultaneously minimizes retransmission overhead and hardware latency delays makes essential the optimization of the transmission techniques, which is examined in the current thesis. Additionally, in multipoint infrared connectivity, the development of an efficient medium access mechanism that minimizes collisions and channel idle time when many stations wish to utilize the shared medium at the same time is a challenge.

1.3 Outline of research work

This work focuses on the efficient link layer design of WLAN connectivity utilizing the IEEE 802.11 protocol and infrared multipoint links based on IrDA AIr proposals. The following issues are addressed:

a) IEEE 802.11 Wireless LANs

- An elegant and accurate analysis using Markov chain modelling is derived in order to calculate the performance of the Collision Avoidance (CA) procedures of the IEEE 802.11 protocol assuming a finite number of stations and ideal conditions. Simple equations are derived for two models: (a) the ideal IEEE 802.11 MAC throughput model with no packet retry limits and (b) a model that considers packet retry limits and dropped packets as specified in the IEEE 802.11 standard. More specific, in addition to the throughput efficiency, the average packet delay, the packet drop probability, the average time to drop a packet and the packet inter-arrival time are derived for both basic access and RTS/CTS medium access schemes. The accuracy of the derived analysis is verified by means of an OPNET simulator and the improvements in accuracy obtained when retry limits are taken into account are identified. Utilizing the proposed mathematical analysis, an extensive and detailed study is carried out on the influence on protocol performance of the physical layer, network size, data rate, initial *CW* size, maximum *CW* size and packet payload size for both medium access mechanisms.
- The previously developed mathematical model is utilized to study the effectiveness of the RTS/CTS scheme in reducing collision duration at high data rates for both the IEEE 802.11b and 802.11a protocols. An all-purpose expression for the RTS threshold value is derived that actually maximizes

performance by employing the RTS/CTS reservation scheme whenever it is beneficial for both throughput performance and packet delay.

- A new and easy-to-implement backoff algorithm named DIDD (Double Increment Double Decrement) is introduced. An alternative and simpler mathematical analysis is developed based on elementary conditional probability arguments rather than bi-dimensional Markov chains. Results are presented to identify the improvement of DIDD in throughput and packet drop performance comparing to the binary exponential backoff algorithm utilized in the legacy IEEE 802.11.
- Another approach in enhancing performance through reducing overhead costs such as backoff time and RTS/CTS exchanges is proposed by utilizing packet bursting. The concept of transmitting more than one data packets after winning DCF contention can be easily implemented through the fragmentation mechanism of the IEEE 802.11 protocol. Results obtained for different scenarios demonstrate the enhancement of both throughput and packet delay performance. Furthermore, fairness is explored in short-time and long-time scales for both the legacy DCF and packet bursting cases.
- Transmission errors can occur either independently with fixed Bit Error Rate (BER) or in time-variable bursts. Both categories of transmission errors are being modelled by developing an improved mathematical model that predicts very accurately the performance of IEEE 802.11 and DIDD protocols since it considers both packet retry limits and transmission errors. Furthermore, the dependency of the protocol performance on Bit Error Rate and other factors related to burst errors is explored for both IEEE 802.11 and DIDD protocols.

b) **Advanced Infrared (AIr) Wireless LANs**

- Access to shared infrared medium is coordinated by Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) techniques. A station that is not able to hear transmissions originating from another station is called a hidden station. As hidden stations likely appear in infrared wireless LANs, the Request To Send / Clear To Send (RTS/CTS) medium reservation scheme is utilized to cope with the hidden station problem. AIr MAC always terminates medium reservation by an End Of Burst / End Of Burst Confirm (EOB/EOBC) packet exchange to inform all stations that current reservation is over and that the next contention period starts. The RTS and CTS control packets are transmitted using the maximum *RR* value (*RR*=16) in order to increase their transmission range. Thus, the employed CSMA/CA scheme may cause significant utilization degradation if it results in a significant number of collisions or empty collision avoidance slots. The performance of the proposed AIr MAC collision avoidance procedures is analytically studied. A mathematical model is developed based on a 1-dimensional Markov chain model instead of the bi-dimensional model assuming a finite number of stations and error-free transmissions. The significance of the collision avoidance parameters and their effectiveness on utilization is examined.

1.4 Thesis outline

The main scope of the current thesis is to develop algorithms to support high-speed and robust radio and infrared wireless links. It focuses on the data link layer procedures that determine the performance of these links considering WLAN connectivity. This thesis has four parts; chapter 2 discusses radio and infrared connectivity, chapters 3, 4 and 5 study the IEEE 802.11 protocol, propose certain performance improvements and include the consideration of a error-prone channel. Chapter 6 studies infrared multipoint connectivity utilizing the IrDA AIr protocol and chapter 7 presents the conclusions and future research.

Chapter 2 mainly provides background information to the thesis and reviews the research carried out in the area of wireless communications. More specific, after a brief introduction to the general topic of WLANs (including some important properties of wireless media), chapter 2 provides information for radio and infrared transmission media and compares the pros and cons of each technology. It then reviews current standards for wireless links like IEEE 802.11, HiperLAN, AIr and others, focusing on the link layer. Several issues unique to wireless communications are discussed and link layer design challenges are explored when the radio or the infrared medium are utilized at the physical layer. Chapter 2 also presents the two methods, computer simulation and mathematical modelling utilized in the current work to address certain challenges and study the performance of wireless communications. It also discusses the performance metrics that evaluate protocol performance. Finally, chapter 2 critically reviews current research in the area or Wireless Communications and especially work carried out in IEEE 802.11 and IrDA AIr communication protocols.

Chapter 3 introduces the IEEE 802.11 protocol architecture by providing a brief description of its main features and mechanisms. An elegant and intuitive analysis is presented that takes into account packet retry limits and leads to simple equations for additional performance metrics to throughput efficiency such as the average packet delay, the packet drop probability, the average time to drop a packet and the packet inter-arrival time for both basic access and RTS/CTS medium access schemes. The accuracy of the mathematical model is validated by comparing analytical with OPNET simulation results. An extensive and detailed study is carried out on the influence on performance of physical layer, data rate, initial *CW* size, maximum *CW* size and packet payload size on protocol performance. Finally, a simple to implement appropriate tuning of the backoff algorithm for maximising performance is proposed depending on the specific communication requirements.

Chapter 4 develops three different approaches in improving performance for the IEEE 802.11 protocol. Firstly, the mathematical model developed in the chapter 3 is utilized to study the effectiveness of the RTS/CTS scheme in reducing collision duration at high data rates for both IEEE 802.11b and 802.11a protocols. An all-purpose expression for the RTS threshold value is derived that maximizes performance by employing the RTS/CTS reservation scheme whenever it is beneficial for both the packet delay and throughput performance. Secondly, a new easy-to-implement backoff algorithm named DIDD (Double Increment Double Decrement) is introduced. An alternative and simpler mathematical analysis is developed based on elementary conditional probability arguments rather than bi-dimensional Markov chains. Detailed results are presented to identify the improvement of DIDD in throughput and packet drop performance comparing to the binary exponential backoff algorithm utilized in the legacy IEEE 802.11. Finally, a different approach in enhancing performance through reducing overhead costs like backoff time and RTS/CTS exchanges is proposed. The concept of transmitting more than one data packets after winning DCF contention, named packet bursting, can be easily implemented through the fragmentation mechanism of the IEEE 802.11 protocol. The previously mathematical model for the legacy IEEE 802.11 is extended in order to consider packet bursting. Results obtained for different scenarios show that the application of packet bursting significantly enhances throughput and packet delay performance. Furthermore, fairness is explored for both the legacy DCF and packet bursting cases in short-time and long-time scales.

Chapter 5 describes the origin, the effects and the variability of transmission errors. The nature of errors is analyzed and is further categorized to independent with fixed Bit Error Rate (BER) and time-variable burst errors modelled by the two-state Gilbert-Elliot Markov chain model. An improved mathematical model is derived that predicts very accurately the performance of IEEE 802.11 and DIDD protocols since it considers both packet retry limits and transmission errors. The new analytical model is applied to both the cases of independent and burst errors. Furthermore, the dependency of the protocol performance on bit error rate and other factors related to burst errors is explored for both IEEE 802.11 and DIDD protocols.

Chapter 6 presents the AIr protocol stack proposal for wireless LANs and analyses the AIr MAC collision avoidance procedures and transfer schemes, including the Reserved and Unreserved transfer modes. A 1-dimensional Markov chain model is constructed instead of the 2-dimensional model in order to calculate the average packet delay for the AIr protocol by obtaining simple mathematical equations. The derived mathematical model is validated by comparing analytical with simulation results and an extensive AIr packet delay evaluation is carried out by taking into account all the factors and parameters that affect protocol performance. Finally, suitable values for both backoff and protocol parameters are proposed in order to reduce average packet delay and, thus, maximize performance.

Chapter 7 presents the conclusions of this thesis and proposes directions for future research in the field of wireless radio and infrared connectivity.

Appendix A presents a detailed overview of the IEEE 802.11 protocol, emphasizing in details on the MAC layer which is of interest to this work. More specific, information about the IEEE 802.11 architecture and services is provided in conjunction with a brief description of the utilized various physical layers and mechanisms (i.e. PCF and packet fragmentation). Appendix B derives throughput efficiency, average packet delay and packet inter-arrival time utilizing the approach that does not consider packet retry limits. Appendix C presents a detailed proof of the fact that the non-linear system developed in Chapter 3 has a unique solution for the case of finite retry limits.

CHAPTER 2

Background

In this chapter we introduce the technologies that support wireless communications and we classify the proposed technologies using two criteria. First, we distinguish point-to-point connections utilized to form Wireless Personal Area Networks (WPANs) from multipoint connections used to form Wireless Local Area Networks (WLANs). Second, we classify technologies according to the medium they utilize, radio or infrared optical.

WPANs allow mobile devices to function together in ad hoc networks within a personal space. WPANs aim to replace wired connectivity between devices such as still and video cameras, laptops and MP3 players. WLANs provide computer connectivity in a small area such as an office complex, a building or a hallway by extending or replacing a wired LAN. The main attraction in WLANs is the flexibility and mobility; bandwidth considerations are of secondary importance.

IEEE 802.11, HiperLAN, IrDA AIr and HomeRF are some of the wireless technologies that support multipoint WLAN connectivity using radio or infrared. Especially, IEEE 802.11 standard supports multipoint connectivity and offers several choices of physical medium such as radio and infrared transmission capabilities. IrDA AIr protocol proposal utilizes the infrared spectrum to implement wireless LANs.

The outline of this chapter is as follows. Section 2.1 describes wireless connectivity and categorizes radio and infrared communication systems. Section 2.2 compares radio and infrared transmission media for wireless connectivity and section 2.3 presents current standards for WLANs focusing on transmission techniques and medium access procedures. The link layer design challenges arising from both the radio and infrared medium characteristics are discussed in section 2.4. Section 2.5 presents the advantages and disadvantages of computer simulation and mathematical modeling techniques that evaluate the performance of communication systems and section 2.6 presents the performance metrics used to evaluate the system performance. Finally, section 2.7 reviews current research related to link layer design challenges.

2.1 Overview of Wireless Networks

Wireless networks serve many purposes. In certain cases they are used as cable replacements, while in other cases they are used to provide access to corporate data from remote locations. Much of the industry hype surrounds third-generation wide area networks that provide broadband wireless connectivity to users on a national basis. As users carry around multiple devices, a need arises for an easy, effective way for them to communicate; and what is easier than wireless?

Wireless networks are divided in four main categories: wireless personal area networks (WPANs), wireless local area networks (WLANs), wireless wide area networks (WWANs), and satellite networks. For each category, the prevalent technologies and the wireless network protocols as well as the types of applications that these technologies are using, are summarized in Table 2.1. Information such as coverage area, function, relative cost and throughput are some of the main areas where these networks differ. The current work is focusing on WLANs.

Wireless networks can be also divided into two broad segments: short-range and long-range. Short-range wireless pertains to networks that are confined to a limited area. This applies to personal area networks (PANs) where portable computers need to communicate as well as to local area networks (LANs), such as corporate buildings, school campuses, manufacturing plants or homes. These networks typically operate over the unlicensed spectrum reserved for industrial, scientific, medical (ISM) usage. The available frequencies differ from country to country. The most common frequency band is at 2.4 GHz, which is available across most of the globe. Other bands at 5 GHz are also often used. The availability of these frequencies allows users to operate wireless networks without obtaining a license and without any charge.

Long-range networks continue where LANs end. Connectivity is typically provided by companies that sell the wireless connectivity as a service. These networks span large areas such as a metropolitan area, a state or province, or an entire country. The goal of long-range networks is to provide wireless coverage globally. The most common long-range network is wireless wide area network (WWAN). When global coverage is required, satellite networks are also available. Note that in contrast with short-range networks, WWANs and satellite networks often charge either by the minute or by the amount of data transferred.

Type of network	Coverage area	Function	Associated cost	STANDARDS
Wireless personal area network (WPAN)	Personal operating space; typically 10 meters	Cable replacement technology, personal networks	Very low	IrDA, Bluetooth, 802.15
Wireless local area network (WLAN)	In buildings or campuses; typically 100 meters	Extension or alternative to wired LAN	Low-medium	802.11a, b, g, HIPERLAN/2, IrDA AIr, HomeRF
Wireless wide area network (WWAN)	Coverage provided on national basis from multiple carriers	Extension of LAN	Medium-high	GSM, TDMA, CDMA, GPRS, EDGE, WCDMA
Satellite networks	Global coverage	Extension of LAN	Very high	TDMA, CDMA, FDMA

Table 2.1 WPANs, WLANs, WWANs and Satellite wireless network categories

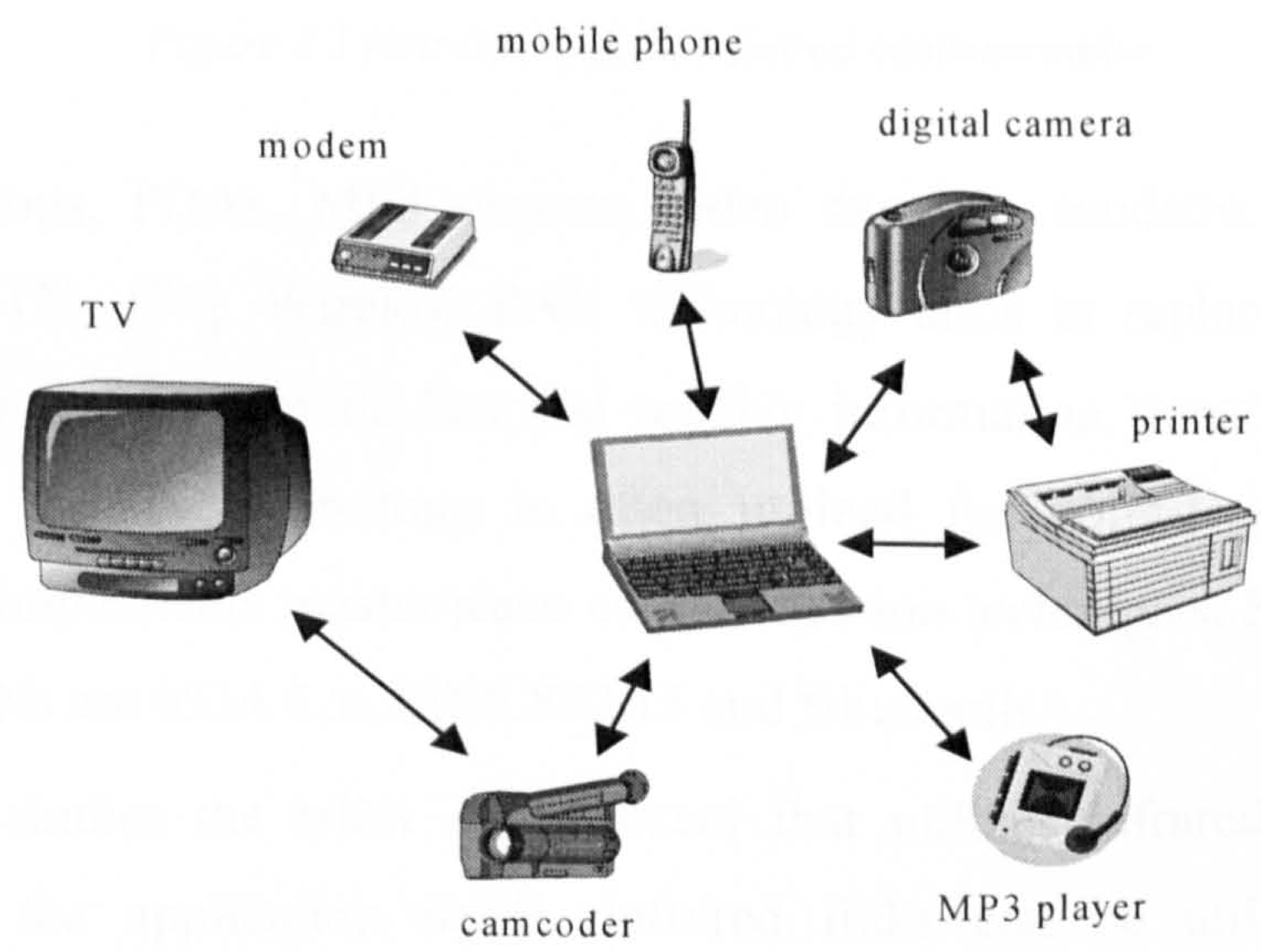


Figure 2.1 WPAN applications

This work considers the information exchange between two or more PCs and/or peripherals. Depending on user applications, two main categories are defined for wireless information exchange:

a) **Wireless PANs.** A wireless PAN (WPAN) (figure 2.1) enables short-range ad hoc connectivity among portable consumer electronics and communications devices, such as laptops, PDAs, MP3 players, video cameras, modems, printers, mobile phones and TVs [74]. Wireless PAN technology aims to replace cables between

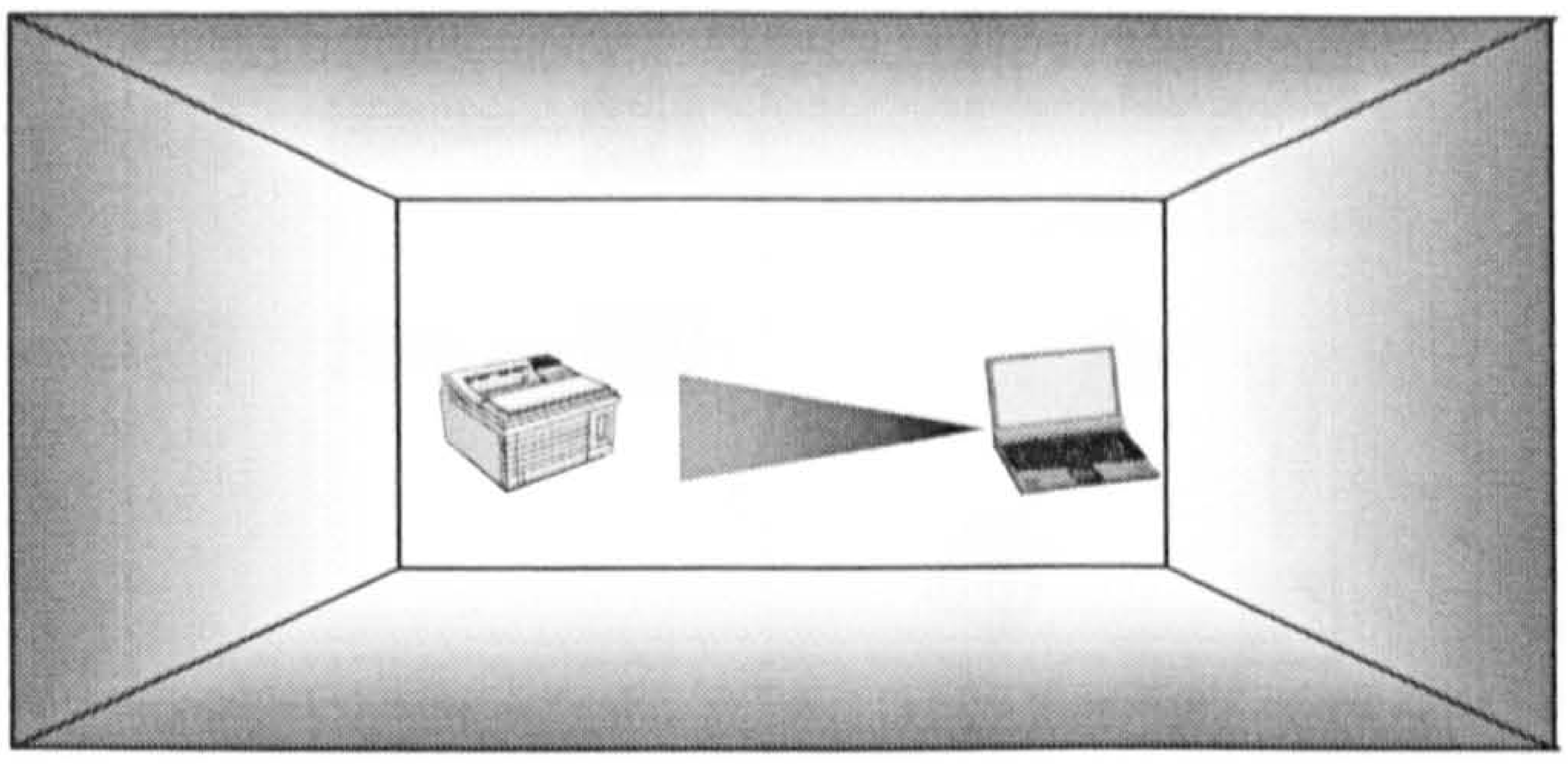


Figure 2.2 Line of sight infrared communication

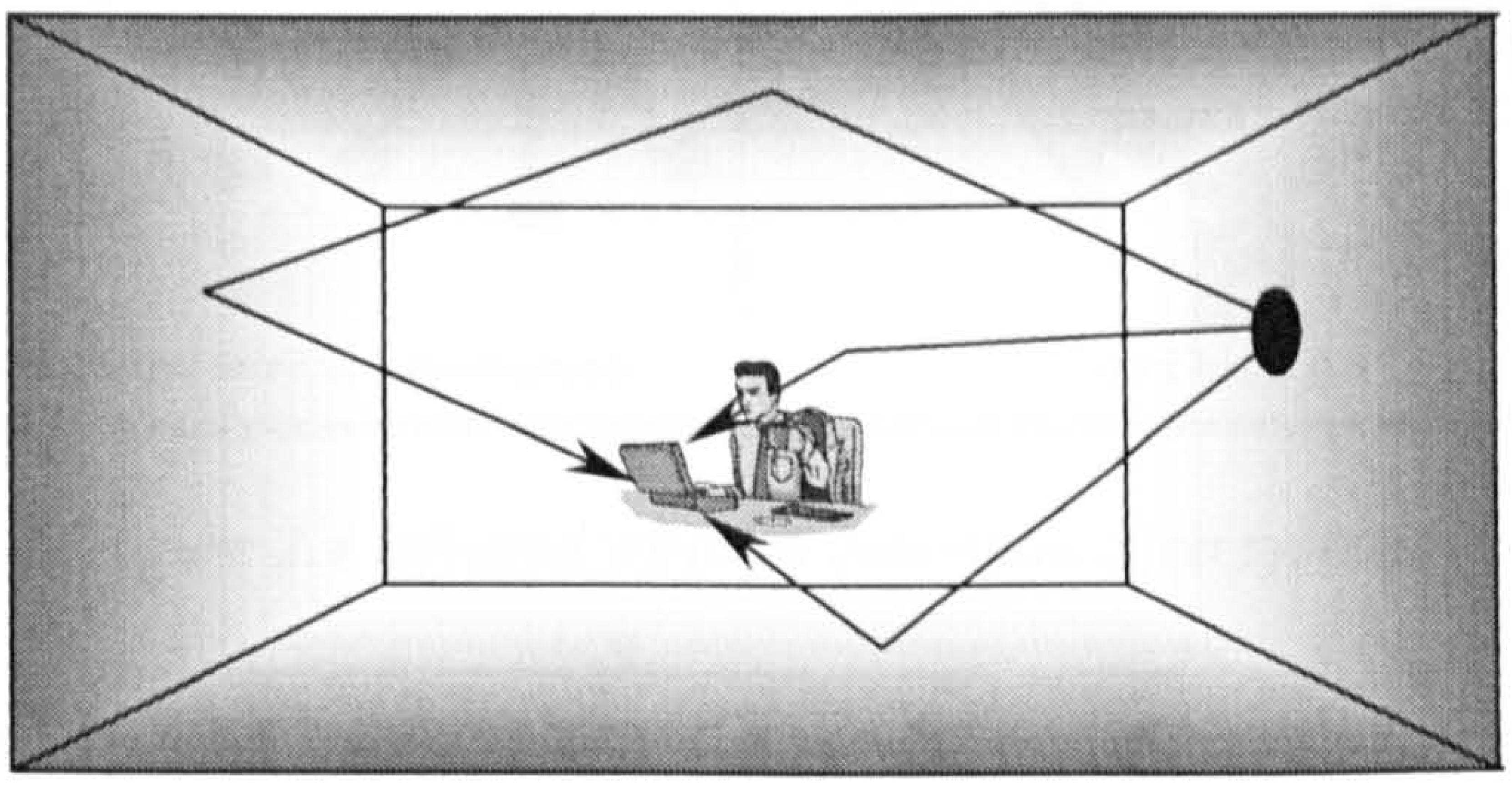


Figure 2.3 Non-line of sight infrared communication

such as laptops, PDAs, MP3 players, video cameras, modems, printers, mobile phones and TVs [74]. Wireless PAN technology aims to replace cables between these devices and to provide fast and reliable information transfer abilities to the single user. WPAN technology is often utilized for point-to-point information transfer and implements master/slave communication techniques. Some examples of Wireless PANs are IrDA 1.x, IEEE 802.15 and Bluetooth.

This work studies the IrDA AIr protocol that utilizes infrared IrDA 1.x links. Depending on the application needs, infrared links can be utilized in different configurations and employ narrow-angle or wide-angle transmitters and receivers. Narrow-angle IR ports have a narrow beam transmission pattern and a narrow reception field of view (FOV). Wide-angle IR ports have a broad beam radiation pattern and a wide FOV [43].

Infrared links are also classified as line-of-sight (LOS) and non-LOS links. In LOS links, there is always an unobstructed line-of-sight direct optical path between the transmitter and receiver. Figure 2.2 presents a narrow-angle LOS infrared communication. Links with transmissions reflected of ceiling and other reflecting surfaces are termed 'non-line-of-sight' (shown in figure 2.3). These links provide a high

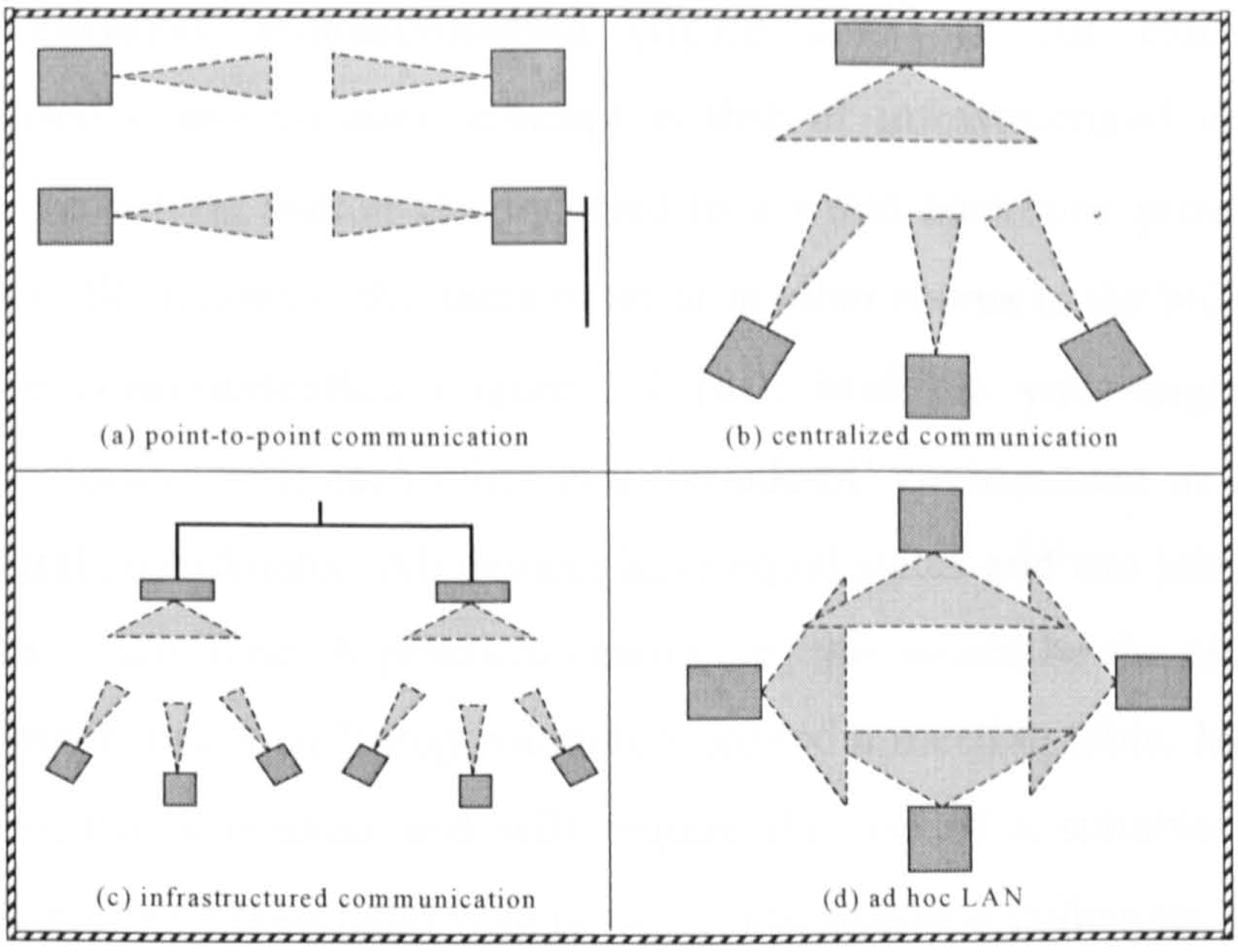


Figure 2.4 Infrared wireless communication systems

level of device mobility (the user does not have to maintain alignment and a LOS path) but a low level of power efficiency and are susceptible to multi-path dispersion which limits the available data rate [71].

Depending on the topology, infrared communications are divided into the following categories:

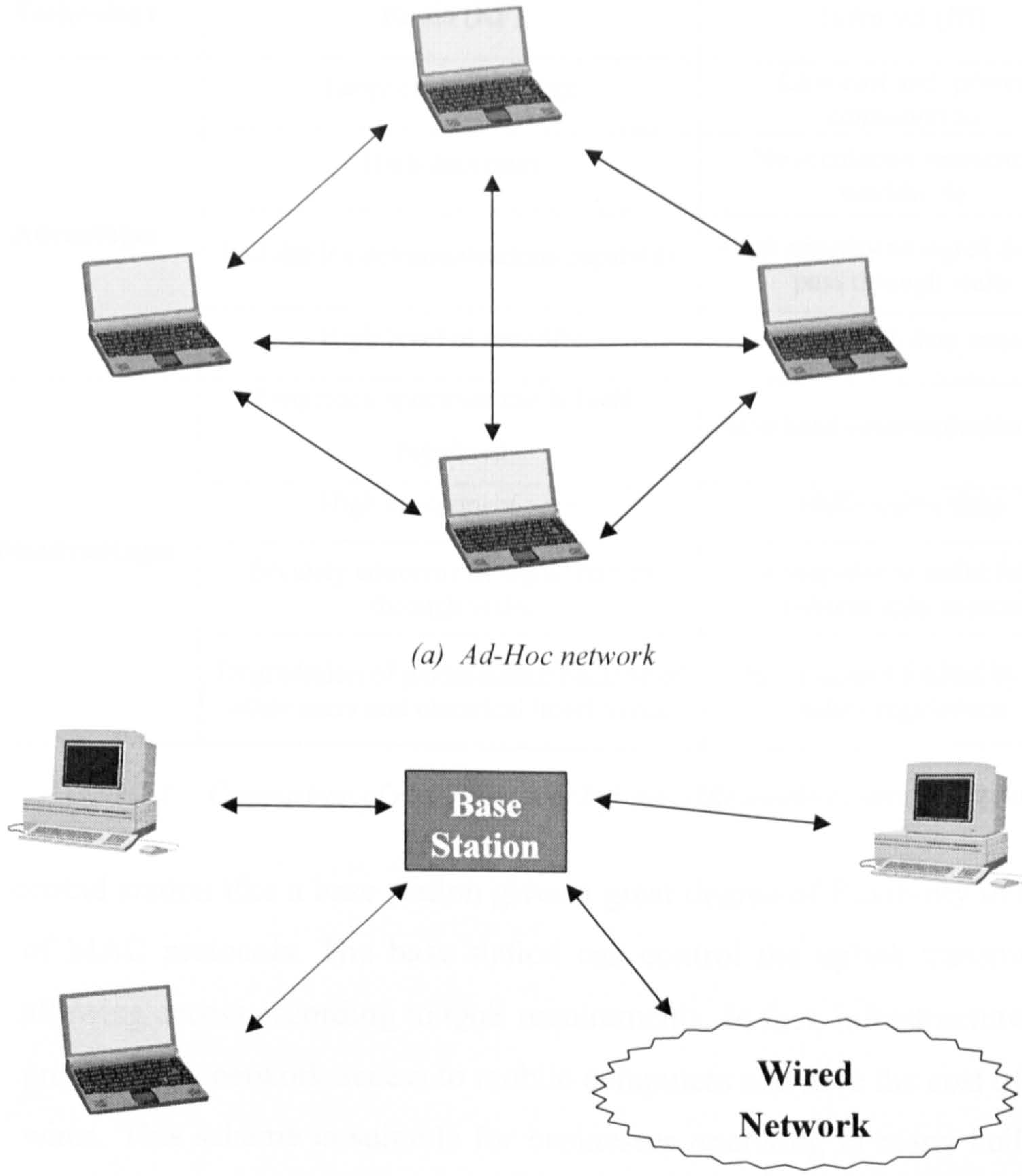
- **point to point communication** (figure 2.4(a)): Two narrow angle infrared devices exclusively communicate with each other. Typical applications are the transfer of files from a mobile computing device to a desktop computer or wireless printing from a mobile device the information and the uploading of music files from a laptop to a portable MP3 player. One of the devices may be fixed and connected to a wired network providing network access to the mobile device. Media access is relatively simple where devices simply exchange periods of transmission with one device as a 'master' controller.
- **centralised communication** (figure 2.4 (b)): Multiple narrow angle devices communicate with a wide-angle central station. All data must pass to and from the central station, i.e. other devices cannot communicate directly between themselves. A laptop computer can be assigned the central station role to form a WPAN. A WLAN is formed if the central station is a hub that echoes the received information to all stations. Media access is generally controlled by the hub device and may involve time division access for the station devices.

- **infrastructured communication** (figure 2.4 (c)): An extension of the centralised communication concept is that of infrastructured communication where the central station is connected to a wired backbone providing network access to IR stations in the same room or in other rooms in the building.
- **ad hoc communication** (figure 2.4 (d)): Multiple wide-angle devices are communicating with each other in a 'broadcast' environment in which there is no central co-ordinator. All devices have equal status and can join and leave the network at any time. A practical example of this would be the establishment of an ad-hoc network of laptop computers around a meeting table. Media access in the scenario is random and will require the use of a suitable media access control protocol to contend with potential transmission collisions.

IrDA 1.x connections may be utilized in the first three categories. In the case of centralized and infrastructure IrDA 1.x communications, the central hub must implement a wide-angle instead of a narrow angle IR port. Infrared devices complying with IrDA AIr consider LOS and non-LOS multipoint infrared communication employing wide-angle IR ports and may be utilized for an ad-hoc WLAN.

b) **Wireless LANs.** A wireless LAN (WLAN) aims to offer wireless stations of the same capabilities that wired LANs provide to stationary stations. WLANs were not widely used due to high prices, low data rates, security issues and license requirements. These drawbacks have been recently addressed and a rapid wireless LAN deployment is expected [113]. Based on the network architecture, wireless LAN connectivity can be logically divided into two classes:

- **Ad-hoc LANs:** Ad-hoc networks, also called distributed wireless networks, are wireless terminals communicating with one another with no pre-existing infrastructure in place; therefore, they are also called infrastructure-less networks (figure 2.5 (a)). Wireless terminals have a wireless interface (RF or infrared) and exchange information between one another in a distributed manner. An ad-hoc network has no central administration, thus ensuring that the network does not collapse when one of the terminals is powered down or moves away. Wireless ad-hoc LANs are suitable for serving an immediate need (e.g. laptop users attending a conference meeting or in a classroom) and communicate without the need of an access point.



(a) Ad-Hoc network

(b) Infrastructure network

Figure 2.5 Wireless LAN configurations

- **Infrastructure LANs:** Infrastructure wireless LANs also known as centralized networks, are extensions to wired networks with wireless in the last section of the network (figure 2.5 (b)). In the infrastructure mode, the wireless network consists of at least one access point (acting as the interface between wireless and wired network infrastructure) and a set of wireless end stations. The access point can control the uplink transmissions by allowing access according to QoS requirements. In centralized networks the downlink transmissions (from base station to wireless stations) are broadcast and can be heard by all the devices on the network. The up link (from wireless terminals to the base station) is shared by all the stations and is therefore a multiple access channel. The existence of a

Technology	Radio (RF)	Infrared (IR)
Advantages	Large coverage range	Low cost and power consumption
	High data rates	No regulation restrictions worldwide
	Full duplex communications capability	High security as signal doesn't pass through walls
	High level of mobility	Very high data rates
Disadvantages	Frequency spectrum use is highly regulated	Restricted communication area
	High component costs	Half-duplex links
	Security concerns as signal passes through walls	Susceptible to noise from ambient light sources
	Degradation of performance because of other users and electrical interference	Power output limited by eye safety regulations

Table 2.2 Comparison of radio (RF) and Infrared (IR) wireless communications

central station like a base station gives a great degree of flexibility in the design of MAC protocols. The base station can control the uplink transmissions by allowing access according to QoS requirements. In fact, infrastructure networks provide easy network access to mobile computers and save the cost of installing wires. This scheme is suitable for businesses operating in many buildings and having a large number of employees with laptop computers. It is also suitable for buildings where wiring is difficult or prohibited (e.g. manufacturing plants, stock exchanges, trading floors and historical buildings).

2.2 Wireless transmission techniques

In wireless communications two transmission techniques are implemented; radio (RF) and infrared (IR). Radio and infrared can be considered as complementary transmission media [3][71]. Radio is preferred when long-range or omni-directional transmission is required [108]. Radio is also suitable when user mobility is of prime importance. Infrared is preferred when point-to-point links of high capacity are necessary and when simple low-cost components and international compatibility are required [3][71]. Furthermore, the IR optical medium provides an attractive alternative in certain applications to RF based communications for short range indoor wireless data communications. Both the radio and infrared wireless technologies have certain strengths and weaknesses which make them more suitable for particular wireless environments and applications (briefly shown in table 2.2).

A comparison of the IR and RF wireless communications is carried out next:

a) **Radio:** Radio transmissions are regulated worldwide and often require government licensing. However, the Industrial / Scientific / Medical (ISM) radio bands are an exception to the licensing rule. The 2.4 GHz ISM band is allocated worldwide but some countries allocate slightly different 900 MHz and 5 GHz ISM bands. RF communications systems can have powerful transmitters with very sensitive receivers providing a large range, with the signal radiated in all directions and passing through walls and objects. RF has therefore become very popular because of the large range and high level of mobility it provides. RF channels have the potential for full-duplex communication (using different frequencies for sending and receiving channels), frequency division multiplexing and spread-spectrum modulation techniques that reduce the effects of interference. However the radio frequency spectrum is heavily congested and tightly controlled by regulation providing a limited bandwidth. Radio communication can achieve high rates but suffers from interference from other radio transmitters. As radio passes through walls, radio links operating in different rooms of the same building must utilize different frequencies from the limited radio spectrum in order to minimize interference. In addition, the same radio spectrum may be utilized from other applications. For example, Bluetooth and IEEE 802.11 operate at the same 2.4 GHz ISM band. When a Bluetooth PAN co-exists in the same room with an IEEE 802.11 WLAN, a serious interference problem arises. RF signals are also susceptible to interference from electrical equipment, multipath fading (from phase difference destructive interference) and dispersion (from multiple reflections). There are also security concerns as the signal passes through walls and safety concerns as radio signals can interfere with safety critical or sensitive electronic equipment. RF components can also be expensive and can have high power consumption.

b) **Infrared (IR):** Infrared waves are suitable for short-range indoor communications having several advantages over radio. Infrared components are cheap, easy to build and the infrared radiation is confined to the room of operation. As a result, no licensing is required. However, infrared connections may require a line of sight (LOS) path between the transmitter and the receiver. The infrared spectrum is unregulated worldwide and offers virtually unlimited bandwidth capable of accommodating high data rates [70][138]. However, to increase infrared data rate requires more expensive components. Infrared wireless communication principally benefits from inexpensive readily available optoelectronic components spawned from the fibre-optics industry. Since the radiation is confined to the room of operation, there is no interference with infrared transmissions in neighbouring rooms. IR links are also inherently immune to electrical interference and will not cause interference to sensitive or safety critical electronics. As the IR optical signal does not pass through walls, good security is provided and security issues are much simplified. Independent narrow-beam directed links can also be established in close proximity without interference. However, IR transmitters have a limited power output for eye safety [9] are directional in nature, and are blocked by opaque objects, thus providing a limited range and less mobility than RF. Infrared receivers are also exposed to high ambient light levels inducing receiver noise. Also, inexpensive links can only be half-duplex (i.e. devices cannot transmit and receive at the same time). For diffuse links, multipath dispersion from wall and ceiling reflections can limit the maximum data rate.

2.3 Wireless LAN standards

The great range of applications requiring wireless information transfers has led to the development of many communication standards. Devices for wireless LANs follow specifications developed by independent standard bodies or industry consortia. The next section describes current standards for wireless LANs focusing on physical layer and medium access issues.

2.3.1 HomeRF

The Home Radio Frequency Working Group (HRFWG) was launched in 1998 by leading computer companies to interconnect a broad range of electronic consumer products and personal computers anywhere in the home at an affordable price [52][77]. HRFWG developed the HomeRF specification for wireless communications in home deployments for connecting PCs, peripherals, cordless phones and other consumer electronic devices. HomeRF is actually an effort that aims to tackle the interoperability limitations of many wireless networking access devices and products [52][77]. It uses Frequency Hopping Spread Spectrum (FHSS) techniques in the 2.4 GHz ISM band. The data rate of HomeRF is 1.6 Mb/s and the distance range is about to 45 meters [52]. HomeRF supports up to 127 data connections (PCs and peripherals) and four high quality voice connections (cordless telephones) [100].

Meanwhile, many companies are working with the HRFWG to develop the Shared Wireless Access Protocol (SWAP) [37] for radio-based home networks. The SWAP specification aims to define a new, common air interface that supports both wireless voice and LAN data services in the home environment, provide higher data rates and ensure interoperability among various wireless products being developed by PC, communications and consumer electronics vendors for the home market. SWAP supports both a TDMA (Time Division Multiple Access) service to provide delay sensitive services such as voice data [92], as well as a CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance) service for delivery of delay insensitive high-speed data connections. The CSMA/CA scheme is derived from the IEEE 802.11 protocol.

2.3.2 Advanced Infrared (AIr)

Although, IrDA 1.x protocol has been proven very popular and millions of devices are equipped with an IrDA infrared port, IrDA specifications are addressing the 'point and shoot' user model. The significant increase on the number of mobile devices on market today and recent advances in infrared technology have led to the decision to address the communication requirements of a pool of users. IrDA proposed the Advanced Infrared (AIr) standard for WLANs by extending the IrDA 1.x protocol stack relaxing the range as well as viewing angle restrictions posed by the IrDA 1.x physical layer [63].

The AIr protocol specifications are developed for indoor, high-speed, low cost and multipoint wireless communications. The primary goal in developing AIr specifications was to introduce indoor, high-speed, low cost and multipoint connectivity as well as to preserve the investment in upper layer applications by making certain that existing IrDA applications will be able to utilize the proposed extensions in lower layers. A new physical layer, the AIr PHY [62], was introduced and the IrDA IrLAP layer [64] is split into three sub-layers:

- The AIr Medium Access Control (MAC) [65]
- The AIr Link Manager (LM) [66] and
- The AIr Link Control (LC) [67]

AIr MAC sub-layer allows upper layers to cope with the relaxing of restrictions on the angle and range of AIr PHY ports. AIr MAC is responsible for coordinating the access to the infrared medium among AIr and IrDA devices. AIr MAC supports reservation based media access control, reliable and unreliable data transfer, data sequencing and data rate adaptation. AIr MAC coordinates medium access by employing Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) techniques [65]. AIr LM is a 'thin' layer that allows multiplexing of multiple different client protocols. It also provides dynamic addressing, station grouping as well as priority and non-priority data channels [66]. Dynamic addressing is used to cope with MAC address conflicts and station grouping is utilized to enable multicast transmissions. AIr LC supports connections to multiple devices and is a derivative of the widely used HDLC protocol operating at the Asynchronous Balanced Mode of the protocol. AIr LC does not assign primary and secondary roles to communicating devices. It supports error detection and recovery services, address conflict resolution procedures and guaranteed data delivery services.

AIr links support wide-angle ports operating at ±60 to ±75 degrees (compared to narrow-angle ±15 to ±30 degrees for the IrDA 1.x) in order to achieve multipoint connectivity with other devices in range. AIr devices take advantage of line of sight (LOS) propagation paths but they can also communicate relying on infrared signal reflections from the ceiling and walls if the LOS path is obstructed. AIr utilizes one common modulation format defined as the four-slot Pulse Position Modulation with Variable Repetition Rate (RR) encoding (4PPM/VR). AIr data rate is 4Mbit/s but lower data rates (up to 256Kbit/s) can be utilized if the link quality is low due to high link distance, intense background light and/or non-LOS path. The transmission range of AIr depends on the class of the devices that are being used. Standard range (S-class) AIr transceivers are expected to provide a transmission distance from 1m to 2.5m at 4Mbit/s. At 256 Kbit/s, a range of at least 5m is achieved. Long-range (L-class) AIr transceivers accomplish a transmission range from 2.5m to 6m at 4 Mbit/s and a range of at least 5m to at least 12m at 256 Kbit/s [62][63].

2.3.3 HiperLAN and HiperLAN 2

The European Telecommunication Standards Institute (ETSI) proposed the High Performance Radio LAN (HiperLAN) protocol to address the need for high-speed short-range wireless communication [50]. HiperLAN considers a wireless extension of a wired network where Mobile Terminals (MTs), such as laptops and PDAs, establish wireless connections to Access Points (APs) of a wired network. HiperLAN utilizes the 5 GHz ISM band [100], which provides larger frequency bandwidth than the 2.4 GHz band, with a data rate of about 24 Mbit/s. HiperLAN was designed to operate with the IEEE 802.11 family through MAC layer bridging. A major difference between HiperLAN and the IEEE 802.11 PHY layers is that HiperLAN operates at a fixed frequency, with no requirement for spread spectrum operation. Although, HiperLAN provided connection-oriented information exchange, automatic frequency allocation and easy integration, it did not experience any commercial success.

HiperLAN 2 is the next-generation WLAN specification which is equivalent to the IEEE 802.11 standard suite. HIPERLAN 2 has been designed to address various issues present in WLANs; it incorporated quality of service (QoS) support for real-time multimedia communication, efficient power consumption for portable devices, strong security and interoperability with Ethernet, IEEE 1394 (Firewire) and 3G mobile systems. HiperLAN 2 specifications define three basic layers; the physical layer (PHY), the Data Link Control (DLC) layer and the Convergence Layer (CL). HiperLAN 2 physical layer continues to utilize the 5 GHz frequency band, but with Orthogonal Frequency Division Multiplexing (OFDM) technology. The approximate transmission range is up to 100 meters and a maximum data rate of 54 Mb/s can be achieved. HiperLAN 2 physical layer supports several modulation and coding alternatives. The medium access control is achieved by utilizing a centralized controller at the AP with time division duplex (TDD) and dynamic time division multiple access (TDMA) techniques.

The original HiperLAN standard and its successor, HiperLAN 2, are still on the books. Most features of the HiperLAN 2 were either never standardized or left to the vendors to implement. Although there were supporters who marketed this technology for local area networking, in the last few years the development of HiperLAN 2 has stopped and certain features are implemented in the IEEE 802.11 standards.

2.3.4 IEEE 802.11 protocol

The past few years, various wireless communication standards have been developed and used extensively. The IEEE Working Group (WG) proposed the 802.11 family of protocols to deal with the modern wireless connectivity needs. The IEEE 802.11 protocols are a significant development, they are now a mature and the most widely deployed technology for WLANs. They are tested and installed for years in corporate, enterprise, private and public environments (e.g. hot-spot areas), and are high likely to play a major role in multimedia home networks and next generation wireless communications. The main characteristic of the IEEE 802.11 WLAN is its simplicity, scalability and robustness against failures due to its distributed nature.

IEEE 802.11 wireless networks can be configured into two different modes: ad-hoc and infrastructure modes. In ad-hoc mode, all wireless stations within the communication range can communicate directly with each other, whereas in infrastructure mode, an Access Point (AP) is needed to connect all stations to a Distribution System (DS) and each station can communicate with others through the AP. The specifications are detailed and cover both the Medium Access Control (MAC) and the Physical Layer (PHY). They incorporate two medium access methods, Distributed Coordination Function (DCF) and Point Coordination Function (PCF). DCF is an asynchronous data transmission function, which is best suited to delay insensitive data. If time-bounded services are required, the optional PCF is used, which is built on top of the DCF.

The IEEE 802.11 MAC, has to support multiple users on a shared medium. In the wired Ethernet, in order to avoid collisions, the terminal transmits and listens at the same time using Carrier Sense Multiple Access with Collision Detection (CSMA/CD) techniques. In radio systems, however, the terminal is not able to transmit and receive simultaneously, thus it is not able to detect a collision. Thus, IEEE 802.11 uses a MAC protocol is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). The MAC sublayer's most basic ability is to sense a quiet time on the network before transmitting. Once the host has determined that the medium has been idle for a minimum time period, it may transmit a packet. This minimum time period is known as "Distributed Coordination Function inter-frame spacing" or "DIFS". If the medium is not idle, the terminal begins a backoff process and waits for a time interval.

Under DCF, data packets are transferred via two methods. The essential method used in DCF is called basic access method. The IEEE 802.11 standard also provides an alternative way of transmitting data packets, namely the RTS/CTS method. Since collisions in wireless environment cannot be detected, an explicit packet acknowledgment (ACK) is used, which means that an ACK packet is sent by the receiving station to confirm that the correct reception of a data packet. Actually, carrier sensing can be performed on both the physical and MAC layers. On the physical layer, physical carrier sensing is done by detecting any channel activity by other stations. In addition to the physical channel sensing, virtual carrier sensing is achieved by using time fields in the packets, which indicate to other stations the duration of the current transmission. All stations that hear the data or the RTS packet, update their Network Allocation Vector (NAV) field based on the value of the duration field in the received packet which includes the short inter-frame spacing (SIFS) and the ACK packet transmission time following the data packet, before sensing the medium again.

The original standard, known simply as IEEE 802.11, defined three different physical layers utilizing:

a) Frequency Hopping Spread-Spectrum (FHSS) modulation in the 2.4 GHz ISM band
b) Direct Sequence Spread-Spectrum (DSSS) modulation in the 2.4 GHz ISM band
c) Infrared (IR) light using non-directed, line-of-sight and reflected transmissions

All three physical layers support both 1 and 2 Mbit/s data rates. Both radio physical layers operate at the 2.4 GHz band providing a range of up to 100 m indoors and the IR physical layer provides a range of up to 10 m but it is confined to the room of operation. IEEE 802.11 standard considers interference and reliability, security, power saving, human safety and station mobility. It supports access-point oriented and ad hoc networking topologies [84]. The next step after was to publish an enhanced version named IEEE 802.11b that extends the data rate up to 11 Mbit/s at the 2.4 GHz band [136]. A high-speed version at 5 GHz UNII band, i.e. IEEE 802.11a, was also defined [137]. IEEE 802.11a standard can achieve a maximum data rate of up to 54 Mbit/s by using OFDM (Orthogonal Frequency Division Multiplexing) modulation technique at physical layer. A more detailed description of IEEE 802.11b and 802.11a is included in Appendix A.

2.3.5 Other ongoing activities within IEEE 802.11 Working Group

Certain IEEE 802.11 Task Groups are in place to improve upon the existing 802.11x standards. The areas of concentration are security, quality of service, compliance and interoperability. Most of these are still in the Task Group stage of the specification process. We are starting with 802.11b before 802.11a because it has achieved a higher level of commercial adoption. The letter after the name represents the time at which the specification was first proposed, but not necessarily which one was first adopted.

(i) IEEE 802.11b/Wi-Fi

IEEE 802.11b [136] is the most popular standard at the moment in the 802.11x family. The specification was approved at the same time as 802.11a in 1999, but since then has achieved broad market acceptance for wireless networking. 802.11b is based on the DSSS version of 802.11, using the 2.4 GHz spectrum. Since DSSS is easier to implement than orthogonal frequency division multiplexing (OFDM) used in 802.11a, 802.11b products came to market much sooner than their 802.11a counterparts. The 2.4-GHz spectrum is also available globally for WLAN configurations, while the 5 GHz spectrum that 802.11a uses is for limited use in many countries. To help foster interoperability between 802.11b products, the Wi-Fi Alliance (formerly the Wireless Ethernet Compatibility Alliance (WECA)) has set up certification for the Wireless Fidelity, or Wi-Fi. Obtaining Wi-Fi certification ensures that 802.11b products will be able to interoperate with other Wi-Fi products globally. This certification, combined with the release of 802.11b products by leading networking companies has made 802.11b the most commonly used 802.11 standard in commercial WLAN products.

All previously mentioned coding techniques for legacy IEEE 802.11 provide a speed of 1 to 2 Mbit/s, lower than the wired networks that provide data rates of at least 100 Mbit/s. The only technique (with regards to FCC rules) capable of providing higher speed is DSSS, which was selected as a standard physical layer technique. IEEE 802.11b is actually an extension of the IEEE 802.11 DSSS scheme, providing data rates of 1 to 2 Mbit/s and two new speeds of 5.5 and 11 Mbit/s. Each channel requires the same 11-MHz bandwidth as in the case of a DSSS channel. To achieve a higher data rate in the same bandwidth, a new modulation scheme called Complementary Code Keying (CCK) is used.

The use of the 2.4 GHz band for communication has advantages and disadvantages. On the plus side, the 2.4 GHz spectrum is almost universally available for WLAN configurations and 2.4 GHz signals are able to penetrate physical barriers such as walls more effectively than higher frequencies can. The downside of using the 2.4 GHz spectrum is congestion. Since it is unlicensed, meaning anyone can use it without obtaining a special license, other electronic products also use this frequency for communication. Two common examples are cordless phones and microwave ovens. With the widespread use of this spectrum, there is a possibility that it will become overcrowded, resulting in too much interference. Hopefully, this will not be the case since any manufacturer of any 2.4 GHz product is required to take interference into account in its product design.

In typical indoor office configurations, an IEEE 802.11b access point can communicate with devices up to 100 meters away. The further away a terminal is from the access point, the slower the communication will be. Devices within about 30 meters can usually achieve a raw data transfer rate of 11 Mbit/s; beyond 30 meters, the rate drops to 5.5 Mbit/s, to 2 Mbit/s around 65 meters away, and finally, to 1 Mbit/s around the outer edge. These numbers represent the anticipated coverage area and transmission speeds, but the products from each vendor will differ in performance. If you are looking to implement an 802.11b WLAN, it is recommended that you do a site survey to obtain the actual operating range and associated bandwidth for your location.

(ii) IEEE 802.11a

IEEE 802.11a [137] is a very promising high-speed alternative to 802.11b[1], providing wireless data speeds up to 54 Mbit/s in distances up to 50 m, and utilizing the 5 GHz spectrum range, which has less interference than the 2.4 GHz spectrum. Unlike the IEEE 802.11b, IEEE 802.11a uses a multi-carrier system rather than a spread-spectrum scheme based on Orthogonal Frequency-Division Multiplexing (OFDM).

---

[1] A common misconception is that 802.11a came first. IEEE 802.11b does represent the second generation of wireless networking but 802.11a actually represents a third generation.

OFDM uses multiple carrier signals at different frequencies, sending some of the bits on each channel. OFDM, however, dedicates all of the sub-channels to a single data source. OFDM is very efficient in time-varying environments, where the transmitted radio signals are reflected from many points, leading to different propagation times before they eventually reach the receiver. OFDM delivers higher data rates and a high degree of signal recovery, due to its encoding scheme and error correction. IEEE 802.11a can achieve data rates of 6, 9, 12, 18, 24, 36, 48 and 54 Mbit/s.

The move to the 5 GHz band and OFDM modulation provides two important benefits over 802.11b. First, it increases the maximum speed per channel from 11 Mbit/s to 54 Mbit/s. This is a tremendous boost, especially considering that the bandwidth is shared among all the users on an access point. The increased speed is especially useful for wireless multimedia, large file transfers and fast Internet access. Second, the bandwidth available in the 5 GHz range is larger than available at 2.4 GHz, allowing for more simultaneous users without potential conflicts. Additionally, the 5 GHz band is not as congested at the 2.4 GHz band, resulting in less interference.

These advantages come with some downsides. The higher operating frequency equates to a shorter range. This means that to maintain the high data rates, a larger number of 802.11a access points are required to cover the same area, versus 802.11b. While 802.11b access points have a typical range of 100 meters, 802.11a access points are often limited to between 25 and 50 meters. In addition, OFDM requires more power than DSSS, leading to higher power consumption by 802.11a products. This is definitely a disadvantage for mobile devices that have limited battery power. Another downside is that 802.11a and 802.11b products are not compatible. With the large number of 802.11b products on the market, this will have a negative effect on the adoption of 802.11a products. That said, both standards can coexist, and products are now on the market that support both 802.11a and 802.11b in a single chipset. This dual-mode approach is very attractive for users who want the advantages of 802.11a, with the backward compatibility and market penetration of 802.11b.

Due to the increased complexity of 802.11a, the first products did not reach the market until early 2002. Since then other vendors have released 802.11a products, helping 802.11a gain broader market acceptance and interoperability certification. However, there are certain barriers before the worldwide acceptance. First of all, the coverage range is very short. The 5 GHz frequency band is not available worldwide. Japan, for example, permits the use of a smaller band, containing half the channels. In Europe, the standard does not comply with various EU requirements. This actually leaves some doubt as to whether it will become a global standard as 802.11b has. Moreover, IEEE 802.11a does not provide any QoS mechanisms. A step into the direction of wide establishment of IEEE 802.11a is the creation of a multi-vendor interoperability certification for 802.11a products.

(iii) IEEE 802.11d

The IEEE 802.11d Task Group describes a protocol that will allow an IEEE 802.11 device to receive the regulatory information required to configure itself properly to operate anywhere on earth. The IEEE 802.11d standard (referred to as the "global harmonization standard") adds the requirements and definitions necessary to allow IEEE 802.11 WLAN equipment to operate in markets not served by the current standards. This is especially important for operation in the 5 GHz band because the use of those frequencies differ widely from one country to another (especially where the 2.4-GHz band is not available).

(iv) IEEE 802.11e

The IEEE 802.11e Task Group [56] is working to provide quality of service (QoS) characteristics and capabilities within 802.11 wireless LANs. The IEEE 802.11 Working Group realized that the original 802.11 standard and its amendments, a, b, and g, don't provide an effective mechanism to prioritize traffic. Without such a mechanism, there can't be any strong quality of service, which means that Wi-Fi can't optimize the transmission of audio and video.

IEEE 802.11e revises the MAC layer to improve QoS and address MAC enhancement. It accommodates time-scheduled and polled communication during null periods when no other data is moving through the system. In addition, IEEE 802.11e improves polling efficiency and channel robustness by employing a prioritized scheme that can be used to ensure that high priority users get more bandwidth allocation than low priority users. A QoS station is any base station implementing 802.11e. In a QoS station, a hybrid coordination function (HCF) replaces modules for a distributed coordination function (DCF) and point coordination function (PCF).The HCF consists of enhanced distributed-channel access (EDCA) and HCF-controlled channel access (HCCA). EDCA extends the legacy DCF mechanism to include priorities. As with the PCF, HCCA centrally manages medium access, but does so more efficiently and flexibly. These enhancements should provide the necessary quality for services and applications such as voice-over-IP (VoIP), audio and video over 802.11 wireless networks, video conferencing, media stream distribution, enhanced security applications, and mobile as well as nomadic access applications.

Since 802.11e falls within the MAC sub-layer, it will be common to all 802.11 PHYs standards (e.g. 802.11a, b, and g) and be backward compatible with all existing wireless LANs based on the 802.11 series of standards. As a result, the lack of a finalized 802.11e specification shouldn't impact a decision on which Wi-Fi flavour to use when deploying a new WLAN. It should be relatively easy to upgrade any existing access points to comply with 802.11e, once it is ratified, through relatively simple firmware upgrades. Up to now, there have been innumerable delays, thanks to arguments over how many classes of service should be provided and exactly how they should be implemented. However, it appears as if most of the issues have been resolved and that the 802.11e amendment will be ratified and be available very soon.

(v) IEEE 802.11f

IEEE 802.11f [58] addresses interoperability among access points from multiple vendors. Actually, IEEE 802.11f is not a specification; instead, it's a "recommended practice" document, meaning that vendor compliance is completely voluntary. The document was drafted with the goal of improving the handover mechanism in Wi-Fi networks, so that end-users can maintain a connection while roaming between two different switched segments (radio channels), or between access points attached to two different networks. Thus, an access point can function as a bridge that connects two 802.11 LANs across another type of network, such as an Ethernet LAN or a wide area network. In this way, IEEE 802.11f facilitates the roaming of a device from one access point to another while ensuring transmission continuity. This is vital if Wi-Fi networks are to offer the same mobility that cell phone users take for granted. The inclusion of IEEE 802.11f in access point design will open up WLAN design options and add some interoperability assurance when selecting access point vendors.

(vi) IEEE 802.11g

IEEE 802.11g [60] is another important extension of IEEE 802.11b. Just like IEEE 802.11a, IEEE 802.11g extends the OSI Model Physical Layer of 802.11b, by adopting either single-carrier, trellis-coded, eightphase shift keying modulation or OFDM schemes and achieves data rates higher than 22 Mb/s (theoretically up to 54 Mbit/s). However, IEEE 802.11g has two advantages over 802.11a: it operates at the 2.4- GHz band, which is now available worldwide, and it is backwards compatible with the existing installed 802.11b products. In order to achieve the latter, IEEE 802.11g drops the data rate to 11 Mbit/s (or even lower), while the IEEE 802.11a uses the 5 GHz radio frequency and thus it is not interoperable with the 802.11b devices.

IEEE 802.11g brings high-speed wireless communication to the 2.4 GHz band, while maintaining backward compatibility with 802.11b. This is accomplished on two layers. First, 802.11g operates on the same 2.4-GHz frequency band as 802.11b, with the same DSSS modulation types for speeds up to 11 Mbit/s. For 54 Mbit/s, 802.11g uses the more efficient OFDM modulation types, still within the 2.4-GHz band. In practice, an 802.11g network card will be able to work with an 802.11b access point, and 802.11b devices will work with an 802.11g access point. In both of these scenarios, the 802.11b component is the limiting factor, so the maximum speed is 11 Mbit/s. To obtain the 54 Mbit/s speeds, both the network cards and access point have to be 802.11g compliant. In all other aspects, such as network capacity and range, 802.11b and 802.11g are the same. To provide backwards compatibility with 802.11b, the specification supports Complementary Code Keying (CCK) modulation (which 802.11b also uses) and, as an option for faster link rates, it also allows packet binary convolutional coding (PBCC) modulation. Both mandatory and optional aspects are included in the 802.11g standard. The mandatory aspects include the use of OFDM to support higher data rates and support for CCK to ensure backward compatibility with existing 802.11b radios. The optional elements are CCK/OFDM and packet binary convolutional coding (PBCC). Developers may elect to include either optional element or omit both options entirely.

Since 802.11g offers the same speed as 802.11a, comparisons between them are inevitable. And because they both use OFDM modulation, the main differences result from their frequency ranges and corresponding bandwidth. The total available bandwidth at 2.4 GHz remains the same as with 802.11b. This results in lower capacity for 802.11g WLANs when compared to 802.11a. In addition, fewer channels are available, leading to a higher potential of conflicts. When we take into consideration the backward compatibility that 802.11g has with 802.11b, 802.11g becomes an attractive option for companies that have 802.11b installations. In fact, there is a lot of room for debating on many issues about the IEEE 802.11g. In most cases, a 2.4 GHz installation is the way to go for common office applications, since 2.4 GHz products are inexpensive and capable of supporting most application requirements. On the other hand, there will always be situations that can strongly benefit from the use of 5 GHz, e.g. heavily populated environments and networks that support multimedia applications.

(vii) IEEE 802.11h

IEEE 802.11h [57] aims at enhancing the control over transmission power and radio channel selection of IEEE 802.11a in the 5 GHz band in order to make IEEE 802.11a products compliant with European regulatory requirements. IEEE 802.11h covers spectrum and power management. The standard includes a dynamic channel selection mechanism to prevent selection of the frequency band's restricted portion. The standard's transmit-power-control features adjust power to EU requirements. Although European countries, such as the Netherlands and the U.K., currently allow the use of 802.11a under the condition that transmission power control (TPC) and dynamic frequency selection (DFS) must also be present, pan-European approval of the 802.11h standard (along with 802.11e) could be just the ticket to making 802.11a acceptable to many, if not all, local regulatory bodies.

(viii) IEEE 802.11i

Originally focused on 802.11b systems, the IEEE 802.11i Task Group is developing new data security protocols aiming at increasing security and authentication mechanisms for use in all 802.11 systems. The original standard included a wired equivalency protocol (WEP) with two key structures, 40 and 128 bits long. WEP is essentially an encryption technique that incorporates none of the more advanced security techniques known to the networking industry.

Many of the security issues have resulted from companies not using the WEP at all. By implementing additional security mechanisms, corporations can ensure secure wireless communication. In addition, the 802.11i Task Group is working to develop additional security levels for 802.11 WLANs. The developed standard aims at addressing security deficiencies in the WEP algorithm by employing stronger encryption and other security enhancements. Instead of WEP, a new authentication/encryption algorithm based on the Advanced Encryption Standard (AES) is under preparation.

(viii) IEEE 802.11j

IEEE 802.11j [59] is a newly proposed standard. As it now stands, the 802.11j Task Group is mandated to draft a specification that will meet international regulatory requirements, specifically 4.9-5 GHz operation in Japan. Basically, 802.11j is the equivalent of 802.11h, but it is designed for the Japanese regulatory environment.

(ix) IEEE 802.11k

WLAN QoS stands to benefit from another standard proposal, tentatively labelled IEEE 802.11k. The new proposed standard would allow the gathering of detailed information about the communications link between stations and clients. It would standardize the way all 802.11 networks report radio and network performance conditions to other parts of the network stack, to applications, as well as to administrators and operators for the purpose of network management, fault finding and other diagnostics. For example, if a network administrator had all the qualitative information about a station, including its performance capabilities, he or she could then know how to provision it downstream.

The general idea of 802.11k is to strengthen QoS of 802.11e by overlaying 802.11k technology. The 802.11k Task Group only came into existence in early 2003, so its work has just begun. The vision of the 802.11k Task Group is to let higher applications see information about wireless access points and clients, even if they're on different subnets. This is an important step in making an enterprise wireless LAN a unified, consistent system, instead of a loose collection of individual subnets. The goal is to make low-level measurements from the PHY and MAC layers of the wireless LAN available to higher-level applications, which can then make decisions and take actions based on this data. In practice, the protocol elements that will be specified in 802.11k will be MAC and PHY extensions. The standard will also probably deal with protocol, not decision-making or algorithms. For example, a set of measurements may be defined, but there will be no specific rule as to when these measurements should be made, or how the results should be used.

(x) IEEE 802.11m

IEEE 802.11m is proposed as an IEEE 802.11 maintenance Task Group. The group's job is to maintain and correct any errors in any previous amendments to any previously published 802.11 series of specifications like 802.11b, 802.11a, etc.

(xi) IEEE 802.11n

The IEEE 802.11n Task Group is studying various enhancements to the physical and MAC layers to improve throughput. These enhancements include such items as multiple antennas, smart antennas, changes to signal encoding schemes and changes to MAC protocols. The Task Group's current objective is a data rate of at least 100 Mbit/s, as measured at the interface between the 802.11 MAC layer and higher layers. In contrast, the 802.11 physical-layer standards measure data rate at the physical interface to the wireless medium. The motivation for measuring at the upper interface to the MAC layer is that a user can experience a data rate significantly less than that of the physical layer. Overhead includes packet preambles, acknowledgments, contention windows, and various interface spacing parameters. The result is that the data rate coming out of the MAC layer could be about one-half of the physical-layer data rate. In addition to improving throughput, 802.11n addresses other performance-related requirements, including improved range at existing throughputs, increased resistance to interference and more uniform coverage within an area.

2.4 Wireless issues and challenges

The unique properties of the wireless medium make the design of wireless protocols very different and more challenging than wireline networks. Many important issues in the protocol stack design have to be addressed differently if the wireless (either radio or infrared) medium is utilized at the physical layer. Certain properties of wireless systems and their challenges are discussed in detail as follows:

a) **Duplexity:** The duplexing mechanism refers to how the data transmission and the data reception channels are multiplexed. They can be multiplexed in different time slots or different frequency channels. Time division duplex (TDD) refers to multiplexing of the transmission and reception in different time periods in the same frequency band. Using different frequency bands for uplink and downlink is called the frequency division duplex (FDD) mode of operation. In FDD mode it is feasible for the station to transmit and receive data at the same time; this is not possible in TDD. In IR wireless devices, it is very difficult for a station to receive data when it sends data. The reason is that when a station is transmitting data, a large fraction of the signal leaks into the reception circuit (referred to as self-interference). Usually, the power of the transmitted signal is higher by orders of magnitude than the power of the received signal. As a result, the leakage signal has higher power than the received signal, making remote signal detection impossible while transmitting data. However, half-duplex operation degrades the performance of infrared wireless links.

b) **Minimum turn around time:** When a station transmits, the leakage signal blinds its own receiver such that it can not receive remote infrared pulses. After the transmission ends, the receiving circuitry needs a minimum Turn Around Time (TAT) to recover. Thus, a transmitting station is able to receive a TAT time period after its transmission ends. As a result, all participating stations must wait a TAT after a transmission finishes before initiating a new packet transmission to ensure that all stations (including the station that transmitted the previous packet) will be able to receive the new packet. The TAT delay is high in infrared ports and should be taken into account in the design of medium access and retransmission protocols.

c) **Collision avoidance:** Due to hardware constraints, a station can not immediately detect collisions during its transmission. The inability to detect remote transmissions while transmitting results in another implication if many stations compete for medium access; a station can not determine a collision by monitoring channel activity while transmitting, as in Ethernet type protocols. As a result, all stations competing for medium access must implement another collision detection mechanism and employ collision avoidance techniques to minimize the collision probability. Obviously, the more the active stations in the range of a transmitter-receiver pair, the more severe the collisions observed.

d) **Interference and channel errors:** Interference in wireless communications can be caused by simultaneous transmissions (i.e. packet collisions when two or more sources share the same frequency band) or by transmission errors. Packet collisions are typically the result of multiple stations waiting for the channel to become idle and then begin transmission at the same time. Collisions are also caused by the "hidden terminal" problem, where a station, believing the channel is idle, begins transmission without successfully detecting the presence of a transmission already in progress. Interference is also caused by multipath fading, which is characterized by random amplitude and phase fluctuations at the receiver.

The reliability of the communications channel is typically measured by the average bit error rate (*BER*). As a consequence of the time-varying channel and varying signal strength, errors are more likely in wireless transmissions. In wired networks, the probability of errors is very small (*BER* is typically less than 10^{-6}). In contrast, wireless channels may have a *BER* as high as 10^{-3} or higher, resulting in a much higher transmission error probability. Packet loss due to errors can be minimized by using one or more of the following three techniques:

- Smaller packets
- Forward Error Correcting (FEC) codes
- Retransmission methods (i.e. Automatic Repeat Request (ARQ) schemes)

To detect transmission errors, wireless link layer protocols may utilize an immediate acknowledgement (ACK) packet, which follows every data packet transmission. If the ACK packet is not received at the end of a transmission, the transmitter reschedules the data packet for retransmission. ACK packet may result in significant overhead, especially when followed by considerable Turn Around Time (TAT) delays, (i.e. mainly in IR systems). In order to minimize the ACK packet overhead, infrared wireless link layer protocols may choose to acknowledge a number of data packets using a single ACK packet like in IrDA AIr protocol. They may also employ smaller packet sizes to decrease the packet error probability. Another alternative is the implementation of Forward Error Correcting (FEC) codes. Wireless link layer protocols should be efficiently designed to minimize the total delay of data packet retransmissions, ACK packets, packet overheads, TAT delays and FEC.

e) **Human safety:** Research is ongoing to determine whether radio frequency (RF) transmissions from radio and cellular phones are linked to human illness since there are concerns raised, regarding the health risks of wireless use. To date, scientific studies have been unable to attribute adverse health effects to wireless transmissions. Wireless technology should meet stringent government and industry standards for safety and must be designed to minimize the power transmitted by network devices. WLANs should be safe to operate, especially regarding low radiation if used, e.g. in hospitals. For infrared (IR) WLAN systems, optical transmitters must be designed to prevent thermal burns and vision impairment [9].

f) **Security:** In a wired network, the transmission medium can be physically secured, and access to the network is easily controlled. A wireless network using radio transmission techniques[2] is more difficult to secure, since the transmission medium is open to anyone within the geographical range of a transmitter being prone to the dangers of eavesdropping. Wireless access must always include encryption and authentication in order to accomplish data privacy. Efficient and simple-to-use security schemes must be incorporated in wireless designs to minimize the chances of unauthorized access or sabotage. While encryption of wireless traffic can be achieved, it is usually at the expense of increased cost and decreased performance. The insecurity of the wireless links has been identified in literature [46][111] and a number of solutions have been proposed [101][147].

g) **Location dependent carrier sensing:** In the wireless medium, because of multipath propagation, signal strength decays according to a power law with distance. Due to the signal attenuation, data transmission and reception becomes location dependent, function of the position of the receiver relative to the transmitter. Stations far away from the transmitter may not be able to detect the presence of an ongoing transmission. In addition, infrared transmissions are directed; only stations in the reception cone may be able to detect an on-going infrared transmission if adequate reflecting surfaces are not present. In fact, only stations within a specific radius of the transmitter can detect the carrier on the channel. This location dependent carrier sensing results in three possible situations in protocols that use carrier sensing:

- **Hidden Stations:** A hidden station is one that is within the range of the receiver but out of range of the transmitter [4][81]. Let's consider the scenario shown in figure 2.6. Station A transmits to station B. Station C cannot hear the

---

[2] As the IR signal does not penetrate walls, being confined to the room of operation, information exchange between infrared wireless devices is considered particularly secure.

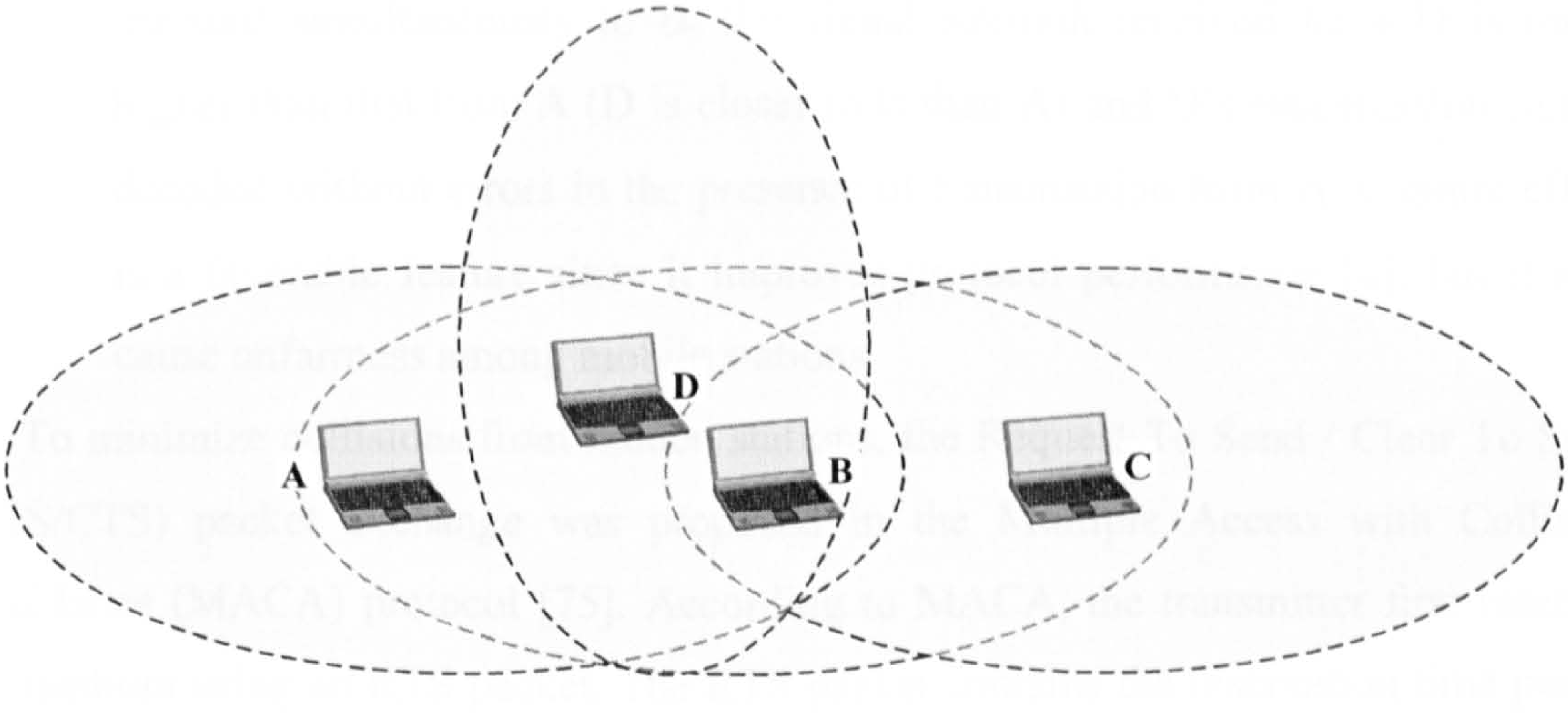


Figure 2.6 The hidden and exposed station problem

on-going transmission from A because it is out of the reception range of station A. If station C wishes to transmit to station B, it listens to the medium and falsely thinks that the channel is idle. Station C initiates transmission and interferes with the transmission from A to B (causing a packet collision). In this case, station C is hidden to station A. Hence, hidden stations may cause packet collisions and, thus, reduce efficiency [78]. Generally, the probability of successful packet transmission decreases as the distance between source and destination increases and/or the traffic load increases [31].

- **Exposed stations**: Exposed stations are complementary to hidden stations. An exposed station is one that is in the range of the transmitter, but out of range of the receiver [4]. e.g. in figure 2.6, consider that B is transmitting a packet to A. C senses the channel busy, and therefore defers transmission of any packet it has, to avoid collisions. However, C could start its transmissions without causing collisions since A is out of range of C (any transmission by station C does not reach station A, and hence does not interfere with data reception at station A). In theory, C can therefore have a parallel conversation with another terminal out of range of B and in range of C. In this case, station C is an exposed station to station B. The link utilization may be significantly impaired due to the unnecessarily deferring stations from transmitting.
- **Capture**: Capture refers to the ability of a receiver to successfully receive a transmission from a given station when multiple stations within range are transmitting simultaneously [4][103]. In figure 2.6, when stations A and D

transmit simultaneously to B, the signal strength received from D is much higher than that from A (D is closer to B than A) and D's transmission can be decoded without errors in the presence of transmission from A. Capture effect is a favorable feature since it improves protocol performance [4], but it may cause unfairness among mobile stations

To minimize collisions from hidden stations, the Request To Send / Clear To Send (RTS/CTS) packet exchange was proposed in the Multiple Access with Collision Avoidance (MACA) protocol [75]. According to MACA, the transmitter first reserves the medium using an RTS packet. The RTS packet contains the reservation time period in a special field. The receiver responds with a CTS packet that echoes the reservation period. Upon receiving the CTS packet, the transmitter proceeds with the data packet transmission. Thus, stations hearing only the RTS or the CTS packet are aware of the medium busy condition and remain silent for the entire data transmission period even if they are not able to hear the data packet [75]. Using the RTS/CTS packet exchange, hidden stations do not result in data packet collisions; collisions can occur only on the short RTS packets if two (or more) stations try to reserve the medium at the same time.

Both IEEE 802.11 and IrDA AIr protocols address the hidden station problem by employing the RTS/CTS control packet exchange. Actually, in IEEE 802.11 the RTS and CTS control packets are transmitted at a lower more robust rate whereas in AIr both RTS and CTS packets are transmitted using the maximum Repetition Rate (*RR*) to increase their transmission range. In order to minimize the RTS/CTS/TAT overhead, in AIr every successful medium reservation may include the transmission of a number of data packets. As the data packets may be transmitted using different RR to match varying channel quality, the reservation time duration is not known when the RTS packet is transmitted. As a result, a reservation is terminated using an End Of Burst / End Of Burst Confirm (EOB/EOBC) control packet exchange.

2.5 Performance modelling of communication systems

As stated in chapter 1, data communications system, including both physical layer and higher protocol layers can be very complex with many factors and system parameters affecting the performance of the whole system. Modelling and analysis of data communications protocols is useful in determining the processes and factors that effect the system performance and optimising parameters to maximise performance.

The principal benefits of modelling are:

- A detailed intuitive understanding of a particular aspect of the system operation and the dominant factors that affect performance can be obtained analyzes protocol operation and leads to protocol design improvements
- A performance evaluation of a particular aspect of a system can be made without physically implement and test a real system
- An evaluation of the effectiveness of all parameter values can be obtained in order to provide optimum performance under specific conditions
- Issues of protocol design that affect performance can be highlighted and possible protocol design improvements can be tested and evaluated evaluates the performance increase of implementing optimum parameter values
- Recommendations can be made to system designers for obtaining optimum system performance

There are two principal methods for performance modelling of communications systems: mathematical analytical modelling and computer simulation [49].

a) **mathematical modelling.** A mathematical model consists of one or more equations that express system performance as a function of protocol parameters, system load and the number of communicating devices. Techniques such as probability theory, statistical mathematics, queuing theory and stochastic process modelling are often used to develop an analytical model for an information exchange system. The mathematical model is used to produce computer graphs that show how system performance changes when one or more system parameters are varied. These graphs are very useful for protocol designers since rapid numerical output results can be easily produced once the mathematical model is developed. The benefits of using mathematical modelling are that relatively simple formulae can be developed to model the behavior of a very specific feature of a system and an intuitive understanding of the dominant factor and relationships that affect performance can be obtained. The disadvantage of analytical modeling is that a number of assumptions and approximations of the system behavior are usually necessary to develop an analytical model. Since simulation modeling accurately predicts system performance (it is explained in detail next), analytical models are usually validated by comparing analytical with simulation results.

b) **Computer simulation modelling.** Computer simulation involves developing models in software that accurately mimic the behavior of a communications system under different situations. With simulation, you can artificially represent any part of the network, such as access points, radio cards, software, the amount of traffic, etc. The computer program actually emulates the behaviour of every station independently and produces very accurate results because it replicates the behavior of a real system. Simulation models usually involve a few or no assumptions. The software model is employed to produce performance results like throughput, delay, collisions, or almost anything else that you want to know when one or more system parameters are varied. This enables designers to determine the results of various configuration settings. The simulation is generally event driven where an event is some time dependent occurrence such as a packet arrival or a timer expiration. Each event in the simulation process has a particular simulation time 'tag' association. This allows events in simultaneous processes in different simulated devices to have the same simulation time although executed sequentially in the computer program. The main advantage of simulation models is that detailed information and output statistics about the performance can be obtained even for very complex communication systems. The main disadvantage of simulation techniques is that, depending on the system complexity and the type of output statistics required, the simulation run (or set of simulation runs) can take a considerable amount of computing time. Also the output results may not give the same level of intuitive reasoning to performance as an analytical model since the dominant factors affecting performance are difficult to determine.

Simulation modelling presented in this thesis uses the OPNET Modeler simulation package. OPNET uses a "process level" to model the behavior of objects and a "node level" that connects the objects to form devices. OPNET also has a "network level" that connects the devices to form actual communication networks. Process models are created using finite-state-machines with C/C++ coded execution blocks using an extensive library of OPNET specific functions in addition to standard C/C++ functions and syntax. Further details can be found in Chapter 3.

2.6 Performance metrics

The metrics that are useful to evaluate the performance of an information exchange system depend on the user applications as well as on the characteristics of the traffic the system is expected to carry. The traffic presented to the system is usually called the offered load. If the offered load contains time insensitive data, such as file transfer, e-mail and web browsing, the communication system must maximize the rate at which data can be sent through the system. If the offered load contains time sensitive data, such as human speech and video, the communication system must minimize the delay of delivering the time sensitive data to the destination; significant variations in the delay of delivering various packets with time sensitive data are often not acceptable.

A brief discussion of the widely accepted performance metrics, which are utilized in the current work, is carried out next[3]:

- **Throughput:** Throughput is the rate at which information data can be sent through the communication system and it is usually calculated in bits per second (bit/s). For time insensitive data, network designers as well as implementers aim to maximize system throughput in order to achieve a better performance; delays in delivering specific data are of secondary importance. Throughput usually expresses the performance of a particular information exchange system and is also referred to as utilization. Throughput is more useful than the data rate because it specifies the actual performance of the system by evaluating all delays introduced by the communication system. It is usually compared to the link data rate to express the performance degradation introduced by the communication technology, such as packet headers, retransmission delays and transmission errors.

This work examines the performance of WLAN communication systems by evaluating the throughput efficiency, which expresses the time portion of the total time that the system delivers offered load to destination at the medium data rate. As an example, if the average packet size is l bits, the average time to transfer a single packet is T secs, and C bit/s is the data rate of the channel, then the throughput efficiency is given by l/TC.

---

[3] Throughput efficiency and average packet delay are considered to be the fundamental quantitative performance metrics of an information exchange system.

• **Delay:** The delay of a system specifies the time needed for information data to travel from the source to the destination station. Furthermore, the average packet delay is defined as the average time spent by a packet from the instant this packet is enqueued until its transmission is completed. Users are particularly interested in the delay in which the system delivers their information data to the destination. Delays are more important on time sensitive data. Types of delays in communication systems are [32]:

i) access delay arises when a transmitted packet is not correctly received at the destination due to a packet collision. Packet collisions are taking place when several stations access the same shared wireless medium. This work analyses the access delay of the collision avoidance procedures of WLANs in chapters 3, 4 and 5 for IEEE 802.11 and DIDD protocols as well as in chapter 6 for the AIr protocol.

ii) retransmission delay due to a transmission error arises when a transmitted packet is not correctly received at the destination because of fading and/or noise. Transmission errors are more likely when a wireless medium is utilized and may significantly degrade performance. This work considers retransmission delays due to transmission errors for IEEE 802.11 and DIDD protocols in chapter 5.

iii) propagation delay arises from the time needed for the signal to travel between two stations. This work considers links that have very small propagation delays, which are safely neglected.

iv) queuing delay occurs in packet switched WANs. When a packet reaches a packet switching device, it may have to wait on a queue if more packets wait for the intended destination. Queuing delay accounts for the time a packet spends on a queue in a packet switching device. This work does not consider queuing delays.

• **Robustness against channel transmission errors:** The wireless channel is time-varying and error-prone [150]. Channel fading and/or noise can significantly degrade performance and make the link between two stations unusable for short periods of time.

• **Fairness:** A MAC protocol is fair if it does not exhibit preference to any single station when multiple stations are trying to access the channel. This results in fair sharing of the bandwidth. This definition can be biased when traffic with different priorities is handled. When multimedia traffic is supported, fairness is defined as being able to distribute bandwidth in proportion to their allocation. Fairness can be either long-term (observed over long time periods) or short-term (the access to the channel should be fair over short time periods).

• **Support for QoS and multimedia traffic**: With the convergence of voice, video and data networks, it is now necessary for MAC protocols to support multimedia traffic. Protocols require mechanisms to treat packets from various applications based on their delay constraints. Two common methods are access priorities and scheduling. Access priorities provide differentiated service by allowing certain stations to get access to the network services with a higher probability than others whereas scheduling can give delay and jitter guarantees. This work considers the probability and the average time for a packet to be discarded as well as the average packet inter-arrival time.

Additional metrics can be used to evaluate the performance of wireless communication protocols such as (however, they are out of scope of the current work):

• **Stability**: Due to overhead in the protocol, the system may be able to handle sustained source loads that are much smaller than the maximum transmission capacity of the channel. A stable system can handle instantaneous loads that are greater than the maximum sustained load when the long-term offered load is less than the maximum.

• **Power Consumption**: Most wireless devices have limited battery power and, hence, it is important to conserve power and provide some power saving features. The need to reduce power consumption is one of the most challenging and interesting topics in wireless engineering, which not only tackles theoretical problems, but also requires complicated physical solutions before any real system can be implemented.

2.7 Research in wireless communication systems

2.7.1 IEEE 802.11 WLANs

Due to the wide acceptance and use of WLANs, extensive research is being carried out to model and study the IEEE 802.11 protocol. Several simulation studies of the 802.11 protocol performance are presented in [33][89] and [132]. Many other papers [5][51] have studied the efficiency of the IEEE 802.11 protocol by investigating the maximum throughput that can be achieved under various network configurations. Protocol fairness is also an interesting issue that has been studied in the literature. In fact, the short-term fairness of a protocol refers to its ability to allocate the channel bandwidth equally to competing stations over short time periods; long-term fairness, in contrast, measures the same ability over longer time periods. The short-term fairness automatically implies long-term fairness, but not vice versa [82]. Chhaya in [31] analyzes the throughput and fairness properties of the asynchronous data transfer methods of the IEEE 802.11 protocol. Fang in [40] reviews a measurement-based backoff algorithm that achieves statistical fair access to the shared medium and models analytically the 802.11 DCF. The analytical model confirms the fairness property of the algorithm and shows the impact of different parameters of the algorithm on the performance of the system. To accurately measure fairness, this work utilizes the average fairness index proposed by Jain [68].

Recently, considerable research activity has concentrated on the performance modelling of DCF by utilizing several analytical techniques. Bianchi in [6] and Wu in [139] employ Markov chain models to analyze DCF operation and calculate the saturated throughput of the IEEE 802.11 protocol. In particular, Bianchi [6] models the idealistic assumption that packet retransmissions are unlimited and a packet is being retransmitted continuously until its successful reception. Wu [139] extends Bianchi's analysis to include the finite packet retry limits as specified in the IEEE 802.11 standard. Nevertheless, neither [6] nor [139] deal with packet delay, packet drop probability or drop time of a transmitted packet utilizing the 802.11 protocol. In [30] we derive the average packet delay for Bianchi's model [6], without considering any packet dropping due to retry limits. Additionally, in [27] we identify the network and traffic conditions for Bianchi's model that render the RTS/CTS mechanism beneficial, achieving lower packet delay with respect to the basic access mechanism. In [19] and

[29] we provide a new performance analysis of the 802.11 protocol, by means of the well-known Markov chain model as developed in [139]. Our work in [19] and [29] considers the effect of packet retry limits and calculates the average packet delay, the packet drop probability and the average packet drop time.

Ziouva in [148] develops a Markov chain model that introduces an additional transition state to the models of [6][19] and [139]. This additional state represents the case that a station transmits a new packet without entering the backoff procedure if it detects that its previous transmitted packet was successfully received and the channel is idle. The model in [148] and subsequent work [141] based on [148], which calculate throughput and packet delay performance, actually allow stations to transmit consecutive packets without activating the backoff procedure. This feature, which is not specified in any IEEE 802.11 standard, causes an unfair use of the medium since stations are not treated in the same way after a successful transmission. In addition, average packet delay calculation in [148] does not consider retry limits and utilizes a very complicated approach; it calculates the average number of packet collisions before a successful reception and the average time a station's backoff timer remains stopped. Moreover, the proposed models in [148] and [141] , lack any validation utilizing simulation results. Calì in [13] makes the assumption that the backoff time is independent of the number of packet retransmissions and sampled from a geometric distribution. Under these assumptions, [13] develops a mathematical model that calculates the DCF throughput and the packet virtual transmission time, which is defined as the time interval between two consecutive successful transmissions from (perhaps different) contending stations. Vishnevsky in [118] extends Bianchi's model [6] and Cali's model [13] by developing a new mathematical model in order to take into account the Seizing Effect. This effect takes place when a station that has just completed successfully its transmission seizes the channel since it has a better chance of winning in the next competition than other stations. This mathematical model, utilizing the geometrically distributed backoff time used in [13], calculates throughput, packet virtual transmission time and seizing probability in order to study the unfairness emerged from the Seizing Effect. However, both [13] and [118] develop complex analytical formulas utilizing several assumptions. In addition, comparison with simulation results in [118] shows that Vishnevsky's model is not very accurate.

In this work, an elegant and accurate analysis using Markov chain modelling is derived in order to calculate the performance of the CA procedures of the IEEE 802.11 protocol, in the absence of hidden stations, transmission errors and assuming a finite number of stations. Simple equations are derived for two models: (a) the ideal IEEE 802.11 MAC throughput model with no packet retry limits and (b) a model that considers packet retry limits and dropped packets as specified in the IEEE 802.11 standard. The derived mathematical analysis calculates in addition to the throughput efficiency, the average packet delay, the packet drop probability, the average time to drop a packet and the packet inter-arrival time for both basic access and RTS/CTS medium access schemes. The accuracy of the derived analysis is verified by comparing analytical with OPNET simulation results. Furthermore, the improvements in accuracy obtained when retry limits are taken into account are also identified. An extensive and detailed study is carried out on the influence on performance of physical layer, data rate, initial *CW* size, maximum *CW* size and packet payload size for both medium access mechanisms. The presented performance results highlight the characteristics of each medium access scheme and give insights on the issues affecting IEEE 802.11 DCF performance.

2.7.2 IEEE 802.11 protocol enhancements

Several other papers in the literature have attempted to improve IEEE 802.11 performance by either modifying the backoff mechanism [90][91][140] or by fine-tuning certain protocol parameters or mechanisms [1][13][14][107][110]. Aad in [1] suggests three different ways to enhance 802.11 performance; by scaling the contention window based on the priority factor of each station or by giving each priority level with a different value of DIFS or different maximum packet length. Cali in [13] proposes a method of estimating the number of active stations via the number of empty slots and exploits the estimated value to tune the *CW* parameter based on a p-persistent version of the IEEE 802.11 protocol. Carvalho in [14] considered the impact of the minimum Contention Window (*CW*) size and the corresponding capacity improvement that is achieved when *CW* increases but not combined with packet retry limits and other protocol parameters. Sadeghi in [107] proposes another approach by transmitting a burst of packets for a single RTS/CTS handshake that considerably improves performance. The concept of transmitting more than one data packets after winning DCF contention is called packet bursting and it is included in the latest 802.11e draft specification [56]. Sheu in [110] suggests concatenating several data packets in a large packet by introducing modifications in certain packet formats.

In this work, three different sets of parameter values for initial contention window size, retry limit and number of backoff stages are proposed. The appropriate adjustment of each proposed set achieves better performance on particular metrics and it could be employed to match specific communication needs. A new easy-to-implement (without any modifications in the packet structure) backoff algorithm named DIDD (Double Increment Double Decrement) is also introduced. An alternative and simpler mathematical analysis is developed based on elementary conditional probability arguments rather than bi-dimensional Markov chains. Detailed results are presented to identify the improvement of DIDD in throughput (higher throughput than the legacy DCF) and packet drop (no packets are discarded) performance comparing to the binary exponential backoff algorithm utilized in the legacy IEEE 802.11 but at the cost of higher packet delay. Furthermore, this work proposes a different approach in enhancing performance through reducing overhead costs like backoff time and RTS/CTS exchanges. The main concept is to transmit more than one data packets after winning DCF contention and can be easily implemented through the fragmentation mechanism of the IEEE 802.11 protocol as discussed in [56]. Results obtained for different scenarios showed that the application of packet bursting significantly enhances throughput and decreases packet delay performance due to the reduction of contention periods and RTS/CTS exchanges. Furthermore, fairness is explored for both the legacy DCF and packet bursting cases in short-time and long-time scales.

There are a number of studies in the literature on the performance of wireless data protocols as well as the RTS/CTS mechanism in IEEE DCF. The authors in [131] and [132] first study the performance of the RTS/CTS mechanism in IEEE 802.11 WLANs through simulations. Although the RTS/CTS scheme is employed to result in a better performance in the presence of hidden stations, research work in [143] and [144] points out that the RTS/CTS handshake does not work as well as expected in dealing with the hidden station problem and reducing interference. Bianchi in [6] proves the superiority of RTS/CTS in most cases by calculating the RTS threshold for throughput maximization but without taking into account packet retry limits. In [27], we evaluate the dependency of the RTS/CTS scheme on network size, but we do not provide any general expression for the RTS threshold. Moreover, in [19] we present a method capable of calculating the average packet delay by taking into consideration retransmission delays with or without packet retry limits. However, [6][27][19] consider the low 1 Mbit/s as being the data and control rate in their presented analysis. Ziouva in [148] demonstrates that for any data rate of IEEE 802.11b (1, 2, 5.5 and 11 Mbit/s) the RTS/CTS scheme always achieves a better throughput and delay performance than the basic access scheme. However, the derived results in [148] did not take into account the fact that the physical header and preamble as well as all the control packets (RTS, CTS and ACK) are always transmitted at either 1 Mbit/s or 2 Mbit/s. Furthermore, the authors in [89][109] perform a simulation study and suggest that the RTS/CTS mechanism must be employed at all times by setting the RTS threshold equal to 0. On the other hand, results in [12] illustrate that the RTS/CTS mechanism provides very limited advantages with respect to the basic access for data rates of 11 Mbit/s when no hidden stations are present.

By utilizing the previously derived mathematical model in [19] and [29] for the throughput and packet delay performance metrics, this work explores the effectiveness of RTS/CTS reservation mechanism in reducing collision duration at high-data rate IEEE 802.11b and IEEE 802.11a WLANs. The study of the impact on performance of using the RTS/CTS scheme is carried out for different data and control transmission rates without the presence of hidden stations. In fact, it is revealed, for the first time, that the overall WLAN performance suffers significantly when the lower rate RTS/CTS exchange is combined with higher transmission data rates and that RTS/CTS effectiveness in improving throughput and packet delay performance is uncertain. Thus, the desire for optimal use of the RTS/CTS reservation scheme makes essential the derivation of an all-purpose expression, which determines when it is beneficial to switch to the RTS/CTS scheme in order to maximize throughput and minimize packet delay performance. The proposed approach allows any station to dynamically adjust its RTS threshold aiming to maximize performance by taking into account the transmission parameters (like data and control rates) in addition to the current congestion level.

2.7.3 Error-prone channels

All previous work assumes the presence of perfect channel conditions; Bianchi in [6] and Wu in [139] developed a mathematical model for the 802.11 DCF throughput performance, utilizing a Markov chain model but without considering the impact of bit errors on performance. Actually, a wireless link is error-prone essentially due to the variation in signal strength of the wireless channel. Crow in [33] and [34] study the effect of errors on performance by means of simulation. Some analytical work has considered the impact of error-prone channels on the throughput [47][85][114][117] [146] and on the energy efficiency of DCF [102]. In particular, Velkov in [117] derives a very complex packet delay analysis, using the Markov chain model of [139] in order to take into account the effects of a fading channel. The authors in [47] and [114] consider transmission errors by means of a Markov chain model but investigated only saturation throughput. A scheme in [85] proposes to optimize the throughput and energy efficiency for a general MAC protocol. Although this optimization provides an insight to the optimization of MAC layer under error-prone environment, it is not specifically for IEEE 802.11 DCF.

Transmission errors are categorized being independent with fixed Bit Error Rate or time-variable burst errors modelled by the two-state Gilbert-Elliot Markov chain model [39]. In fact, by utilizing the Gilbert-Elliot model, a realistic and accurate burst model is produced that in the same time it is not too difficult to implement. An improved mathematical model is derived, which extends the previously derived error-free model that predicts very accurately the performance of IEEE 802.11 and DIDD protocols since it considers both packet retry limits and transmission errors. The new analytical model is applied to both the cases of independent and burst errors. Furthermore, the dependency of the protocol performance on bit error rate and other factors related to independent and burst errors is explored for both IEEE 802.11 and DIDD protocols.

2.7.4 IrDA AIr

Design challenges in IR WLANs have also drawn the attention of the research community. The effectiveness of implementing *RR* on *L*-PPM infrared links is studied in [43][96]. Presented results indicate that RR is suitable on *L*-PPM links as it significantly reduces error rate in hostile medium conditions. Infrared WLANs utilize the RTS/CTS packet exchange to address the hidden station problem. To ensure that the

RTS/CTS scheme operates efficiently, it is essential to maintain reciprocity, which means that the SNR should be symmetric in every pair of stations. The effect of non-reciprocity on various station configurations when the Stop-and-Wait ARQ scheme is implemented at the MAC layer is presented in [16] using AIr PHY and AIr MAC simulators. Results indicate that non-reciprocity depends on physical location and on ambient light level and may result in significant performance degradation.

The fairness problem due to hidden stations for AIr LANs and the suitable improvements for the AIr medium access scheme are presented in [97][98]. The effectiveness of implementing the Stop-and-Wait (SW) ARQ scheme at the AIr MAC layer when two stations are communicating in an AIr LAN is presented in [99][120][121]. These results are not complete, as they consider only two ARQ schemes, and incorporate the fixed and significant collision avoidance delays arising when only one station competes for medium access. AIr MAC and LC performance for LANs with many simultaneously transmitting stations has not been studied yet. In addition, the performance of the AIr MAC collision avoidance procedures has not been extensively studied in the literature [119]-[128]. More specific, in [124] an analytical model for the Collision Avoidance scheme of the AIr protocol that computes throughput performance is proposed. Moreover, a simulation model for the proposed IrDA AIr protocol is developed in [125] and the importance of the Collision Avoidance Slot (CAS) window size limits and adjustments is investigated in [126].

This work develops a new modelling approach; by assuming that the probability of collision is constant, a 1-dimensional Markov chain model is constructed instead of the 2-dimensional model. This new approach considerably simplifies previous analyses and is utilized to calculate the average packet delay of the AIr protocol by obtaining simple mathematical equations. The proposed model predicts AIr packet delay performance very accurately since the mathematical analysis is validated using OPNET simulation results. An extensive AIr packet delay evaluation is carried out next by taking into account all the factors that affect protocol performance. This performance evaluation determines the significance of both link layer and physical parameters, such as burst size, minimum CW size value and minimum turnaround time on AIr packet delay performance. Finally, suitable values are proposed for both backoff and protocol parameters that reduce average packet delay and, thus, maximise performance.

CHAPTER 3

IEEE 802.11 Medium Access Control Protocol

This chapter first introduces the IEEE 802.11 protocol architecture by providing a brief description of its main features and mechanisms. Following that, an elegant and intuitive analysis is carried out that calculates in a accurate way the performance of the IEEE 802.11 protocol. The IEEE 802.11 is the de facto technology for WLANs and it has been used widely in most commercial products available in the market. The IEEE 802.11 standards only cover the Medium Access Control (MAC) layer and the physical layer (PHY). The basic tasks of the IEEE 802.11 MAC layer are medium access, acknowledgement and fragmentation of user data. The IEEE 802.11 MAC layer defines two types of medium access procedures; the mandatory contention-based Distributed Coordination Function (DCF) and the optional polling-based Point Coordination Function (PCF). The DCF is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) and must be implemented by all stations. The period during which the wireless LAN operates in the DCF mode is also known as the Contention Period (CP). The Point Coordination Function (PCF) is operated by a logical entity called the Point Coordinator (PC) that controls access to the medium by implementing a polling scheme. Implementation of the polling scheme is not specified by the standard. The period during which the wireless LAN operates in the PCF mode is also known as the Contention Free Period (CFP). At present, only the mandatory DCF is implemented in 802.11-compliant products and more details about PCF are provided in the Appendix.

The IEEE 802.11 PHY layer is the interface between the MAC and the wireless medium, which transmits and receives data packets over the shared wireless medium. It essentially provides wireless transmission mechanisms for the MAC layer, in addition to supporting secondary functions such as assessing the state of the wireless medium and reporting it to the MAC. The IEEE 802.11 physical layers (PHYs) provide multiple data transmission rates by employing different modulation and channel coding schemes. For example, the IEEE 802.11b PHY [136] provides four PHY rates from 1 to 11Mbit/s at the 2.4 GHz band. Another emerging high-speed PHY, the IEEE 802.11a PHY [137], has been developed to extend the IEEE 802.11 in the 5 GHz Unlicensed National Information Infrastructure (U-NII) band and provides eight PHY modes with data transmission rates ranging from 6 Mbit/s up to 54 Mbit/s.

This chapter is outlined as follows. Section 3.1 introduces the IEEE 802.11 protocol architecture by providing a brief description of its main features and mechanisms. Section 3.2 first presents the assumptions and the parameters utilized in the analysis that will follow. In this section an elegant and accurate analytical model is derived in order to calculate the performance of the CA procedures of the IEEE 802.11 protocol assuming a finite number of stations and ideal conditions (no transmission errors or hidden stations are being considered). The model is validated by comparing analytical results with OPNET simulation outcome in section 3.3. Section 3.4 employs the analytical model to evaluate 802.11 MAC performance for both basic access and RTS/CTS access mechanisms. An extensive and detailed study is carried out on the influence on performance of physical layer parameters, data rate, initial *CW* size, maximum *CW* size and packet payload size. Finally, a simple-to-implement tuning of the backoff algorithm is proposed for the basic access scheme (the conclusions are also applicable to RTS/CTS) depending on the specific communication requirements.

3.1 IEEE 802.11 Architecture

The IEEE Working Group (WP) has added the higher data rate 802.11b and 802.11a PHYs. The MAC layer for each of the 802.11 PHYs is the same. Each of the 802.11 PHYs is subdivided in two sublayers (shown in figure 3.1):

- Physical Layer Convergence Procedure (PLCP)
- Physical Medium Dependant (PMD)

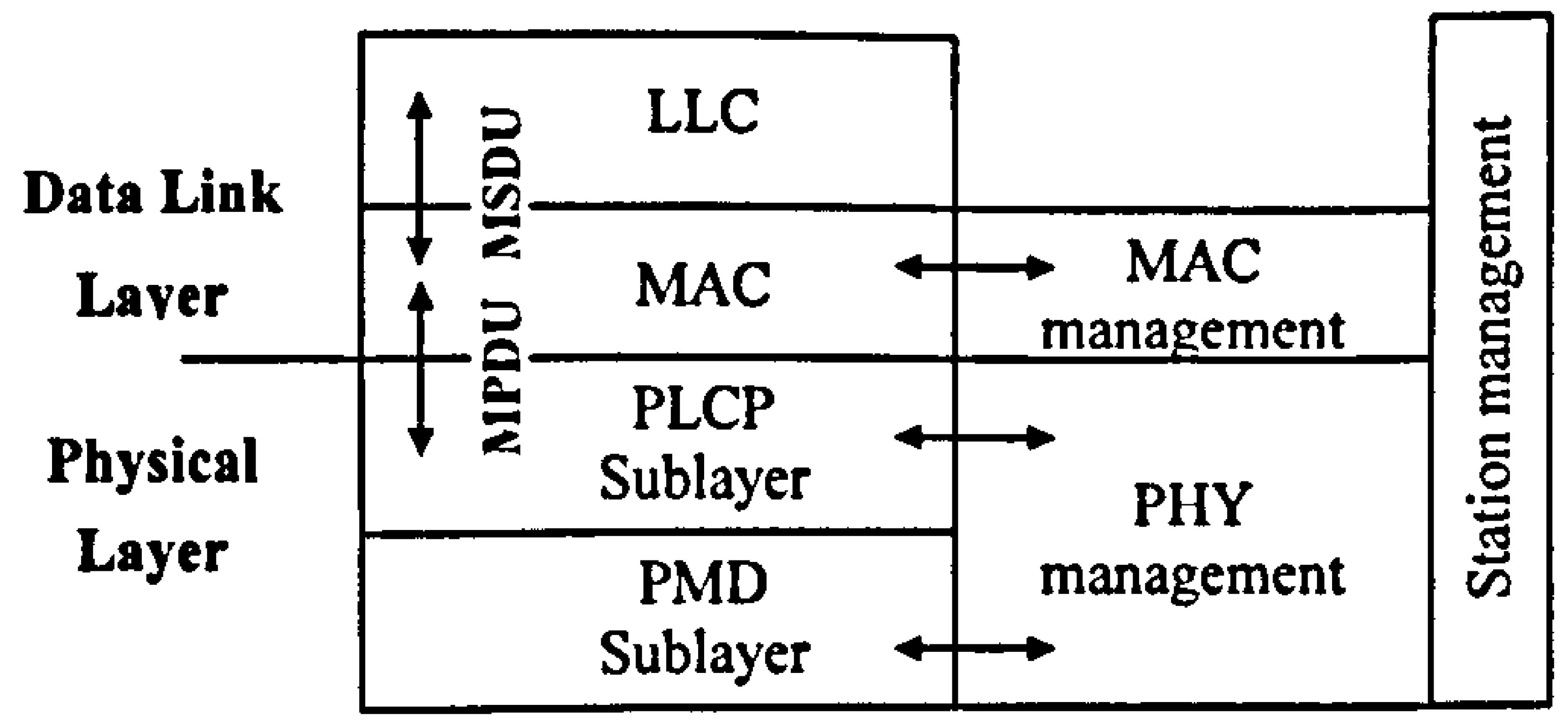


Figure 3.1 PHY sublayers and protocol management

Packages of data delivered to the MAC from the LLC are called MAC service data units (MSDUs). In order to transfer the MSDUs to the PHY, the MAC uses messages (packets) containing functionality related fields. There are three types of MAC packets: control, management and data. One of these messages is called a MAC protocol data unit (MPDU). The MAC passes MPDUs to the PHY layer through the PLCP sublayer. The PLCP sublayer minimizes the dependence of the MAC sublayer on the PMD sublayer by mapping MPDUs into a packet format suitable for transmission over the wireless medium by the PMD, which handles modulation and encoding/decoding of signals. The PMD sublayer actually defines the characteristics and method of transmitting and receiving data through a wireless medium between two or more stations. Moreover, the PLCP sublayer provides a carrier sense signal, called clear channel assessment (CCA). This is needed for the MAC mechanisms controlling the medium access and indicates if the medium is currently idle.

Apart from the protocol sublayers, the 802.11 standards specify MAC, PHY management and station management layers. The MAC management controls authentication mechanisms, encryption and power management. The main tasks of the PHY management includes channel tuning, whereas station management interacts with both management layers and is responsible for higher layer functions .

3.1.1 IEEE 802.11 Physical Layers

The physical layers of IEEE 802.11 have been issued in three stages; the first part was issued in 1997 [135] and two additional parts in 1999 [136][137]. The first part, simply called IEEE 802.11, includes the MAC layer and three physical layer specifications, two in the 2.4 GHz ISM band and one in the infrared, all operating at 1 Mbit/s and 2 Mbit/s. Three physical media were defined in the original 802.11 standard:

- Infrared at a wavelength between 850 and 950 nm, at data rates of 1Mbit/s and 2Mbit/s.
- Frequency Hopping Spread Spectrum (FHSS) operating in the 2.4 GHz ISM band, at data rates of 1 Mbit/s and 2 Mbit/s.
- Direct Sequence Spread Spectrum (DSSS) operating in the 2.4 GHz ISM band, at data rates of 1 Mbit/s and 2 Mbit/s.

The second standardization part was the development of IEEE 802.11b (HR/DSSS) operating in the 2.4 GHz band providing data rates of 5.5 and 11 Mbit/s. The last stage was IEEE 802.11a operating in the 5 GHz band at data rates up to 54 Mbit/s. Figure 3.2 provides the different packet formats used in the IEEE 802.11, 802.11b and 802.11a physical layers. More details about the exact calculation of each field forming a data packet can be found in Appendix A.

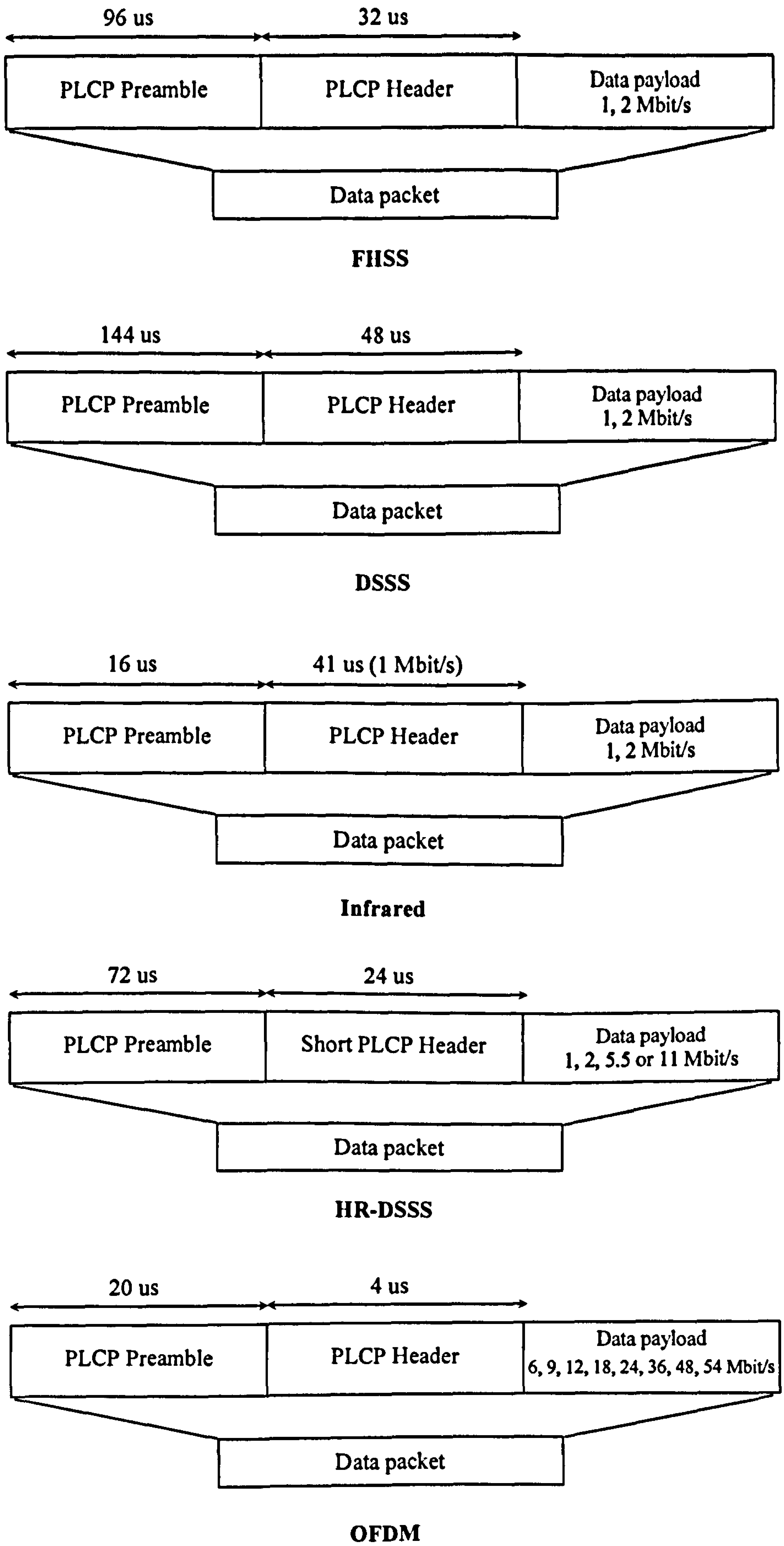


Figure 3.2 Packet formats of all the IEEE 802.11 physical layers (802.11, 802.11b and 802.11a)

3.1.2 IEEE 802.11 Medium Access Control layer

IEEE 802.11 DCF includes carrier-sensing mechanisms in both the physical and MAC layers. On the physical layer, carrier sensing is performed by detecting any channel activity caused by other stations. On the MAC sub-layer, virtual carrier sensing is achieved by using time fields in the data, RTS and CTS packets. These time fields indicate to other stations the duration of an ongoing transmission. All stations that hear any of the data, the RTS or the CTS packets, update their Network Allocation Vector (NAV) according to the value of the duration field in the received packet and do not transmit for the indicated time period. This duration field also incorporates the Short Inter-Frame Space (SIFS) and the ACK packet transmission time period following the data packet, ensuring that the station will sense the medium after the current transmission is over.

In IEEE 802.11 WLANs, priority access to the wireless medium is managed by the use of inter-frame space (IFS) time intervals between the packet transmissions. The IFS time intervals are mandatory periods of idle time on the transmission medium before a station may start transmitting a certain type of packet. Three different IFS intervals have been specified to provide various priority levels for access to the wireless medium; the Short IFS (SIFS), the Point Coordination Function IFS (PIFS) and the Distributed Coordination Function IFS (DIFS). The SIFS is the shortest time interval and is used for the transmission of control packets (RTS, CTS and ACK), which have the highest priority. The time intervals PIFS and DIFS are utilized to separate the PCF and DCF modes, giving a higher priority to the former.

The techniques used for packet transmission in DCF, the basic access and the RTS/CTS reservation scheme, are described next.

A. The basic access method

According to DCF, each station with a new packet ready for transmission monitors the channel activity. If the channel is idle for a time interval equal to DIFS, the station transmits. Otherwise, if the channel is sensed busy (either immediately or during the DIFS), the station persists to monitor the channel until it is determined idle for more than DIFS. The station then initialises its backoff timer and defers transmission for a randomly selected backoff interval in order to minimize collisions. The backoff timer is decremented when the medium is idle, is frozen when the medium is sensed busy and resumes only after the medium has been idle for longer than DIFS. The station whose backoff timer expires first begins transmission and the other stations defer transmission. Once the current station completes transmission, the backoff process repeats again and the remaining stations reactivate their backoff timers (figure 3.3).

A station that receives a data packet, replies by sending a positive acknowledgement (ACK) packet after a SIFS time interval, confirming the successful reception of the data packet. If the source station does not receive an ACK within a specified time, the data packet is assumed to have been lost and a retransmission is scheduled according to the specified backoff rules. This technique may waste a lot of time in case of long packets, keeping the transmission going on while collision is taking place. Moreover, in order to avoid channel capture, a station must wait a random backoff time between two consecutive packet transmissions. After a successful packet transmission, if the station still has packets buffered for transmission, it must execute a new backoff process [135].

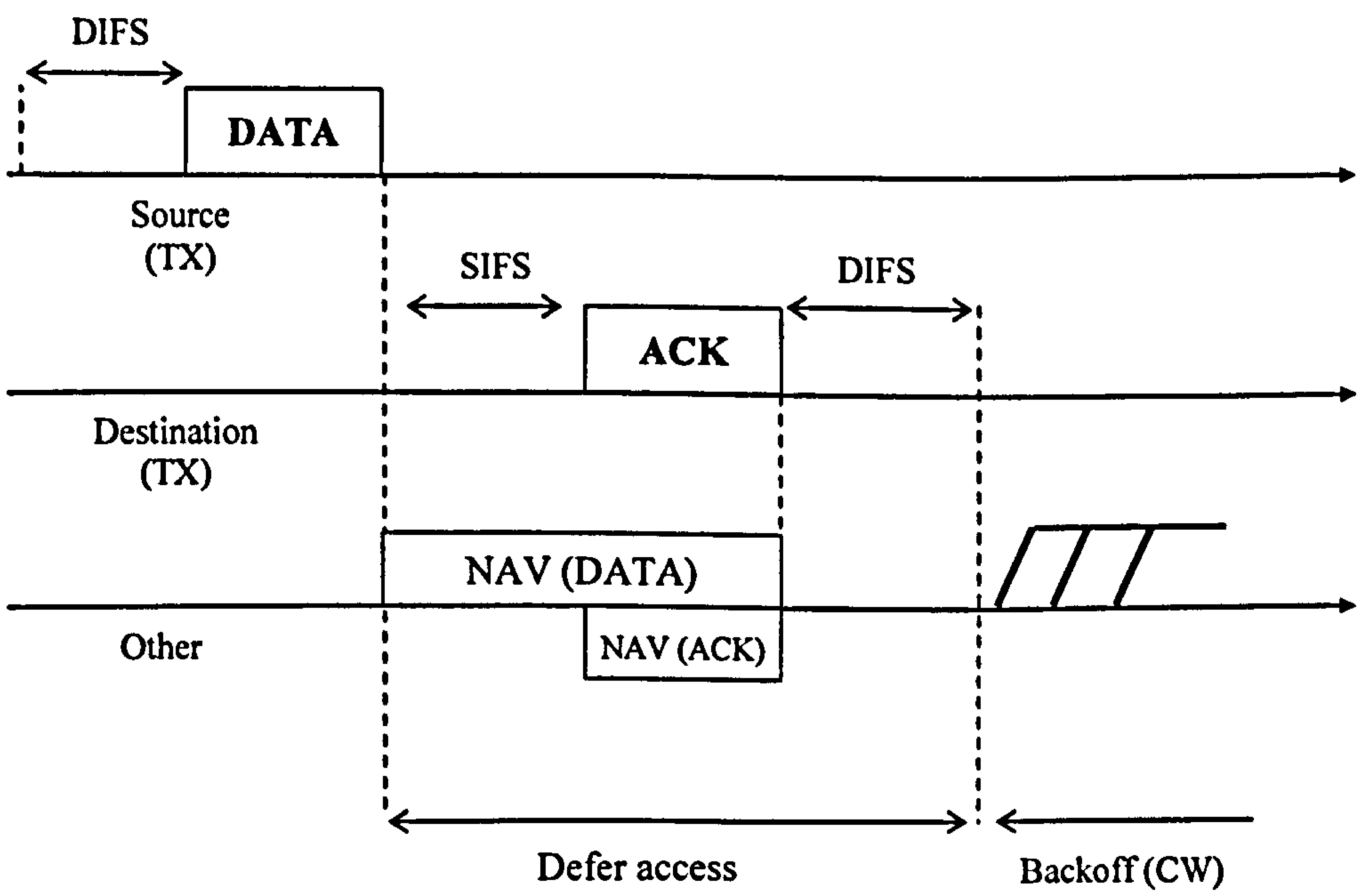


Figure 3.3 Basic access mechanism

B. The RTS/CTS access method

In 802.11, DCF also specifies an optional way of transmitting data packets that involves transmission of special short Request-To-Send (RTS) and Clear-To-Send (CTS) control packets prior to the transmission of the actual data packet. The RTS/CTS scheme is mainly used to minimize the amount of time wasted when a collision occurs and to combat the hidden station problem. Before initiating the transmission of a data packet, the source station sends a RTS packet announcing the duration of the upcoming transmission. When the destination station receives the RTS packet, it replies with a CTS packet after SIFS interval, echoing the duration of the upcoming transmission. After the successful RTS/CTS exchange, the source station transmits the data packet. The receiver responds with an ACK packet to acknowledge successful reception of the data packet (figure 3.4).

Since collisions may occur only on the RTS packets and are detected by the lack of the CTS response, the RTS/CTS scheme results in an increase on system performance by reducing the duration of collisions, especially when long data packets are transmitted. More specifically, if a collision occurs with two or more small RTS packets, the time loss is smaller compared to the collision of long data packets. The RTS/CTS scheme is also employed to result in a better performance in the presence of hidden stations since all the stations are capable of updating their Network Allocation Vectors (NAVs), based on the receipt of either the RTS or the CTS control packets. Thus, if a station is hidden from either the transmitting or the receiving station, by detecting just one packet between the RTS and CTS packets, it can suitably defer transmission, and hence avoid collision.

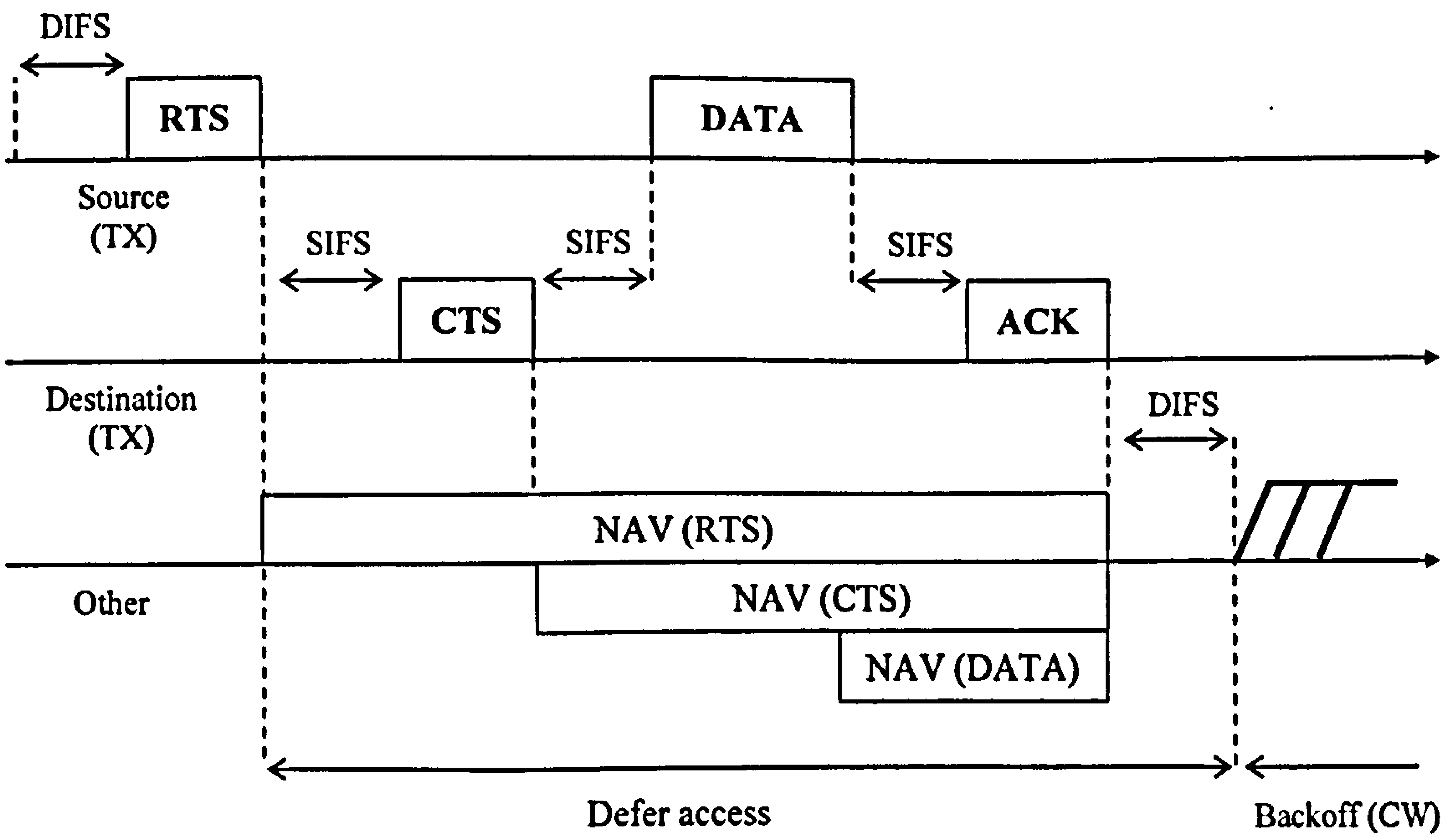


Figure 3.4 RTS/CTS mechanism

However, authors in [143][144] and [145] have reported several potential difficulties in the ability of the RTS/CTS scheme to cope with the hidden station problem. Furthermore, RTS/CTS decreases efficiency since it transmits two additional control packets without any payload. In particular, when short data packets are transmitted, the use of the RTS/CTS scheme might not be advantageous over the basic access. Hence, the standard specifies a manageable object *RTS_Threshold* that indicates the data length under which the data packets should be sent without RTS/CTS. The value of the *RTS_Threshold* is not specified in the standard and has to be set separately by each station. The data packet size is the only parameter used for deciding whether the RTS/CTS reservation scheme should be employed or not. The suitable choice of the *RTS_Threshold* parameter is essential in determining the optimal use of the RTS/CTS mechanism, which can become highly beneficial for the performance of IEEE 802.11 WLANs.

C. The Binary Exponential Backoff (BEB) of DCF

IEEE 802.11 DCF is based on a Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) technique. A contention resolution method, namely binary exponential backoff (BEB), is utilized to randomize moments at which stations are trying to access the wireless medium. By means of this random backoff mechanism, the probability of collisions due to multiple simultaneous transmissions is minimized.

Every station maintains counters that are incremented each time a packet is retransmitted; two different counters are implemented, the station short retry count (SSRC) and the long retry count (SLRC), both of which take an initial value of zero for every new packet. The short retry count indicates the maximum number of retransmission attempts of a RTS packet or of a data packet when the basic access is used. The long retry count indicates the maximum number of retransmission attempts of a data packet when RTS/CTS is used. When either of these limits is reached, retry attempts cease and the packet is discarded. We assume an error free channel, no hidden stations and packets are retransmitted only when they encounter collisions. As a result, the long retry limit is not used in our analysis.

The time following an idle DIFS is slotted and a station is allowed to transmit only at the beginning of each slot. The value of the backoff timer for each station is a uniformly distributed integer number of slots in the interval [0, W_i -1], where W_i is the current contention window (CW) size and i is the backoff stage. The value of W_i depends on the number of failed transmissions of a packet. The backoff timer is decremented when the medium is sensed idle. A station initiates a packet transmission when its backoff timer reaches zero. Figure 3.5 illustrates the CW exponential increase.

At the first transmission attempt of a packet, W_i is set equal to $W_0 = CW_{\min}$, which is called the minimum contention window size. If two or more stations start transmission simultaneously in the same slot, a collision takes place. After a packet collision, the contention window is doubled up to a maximum value, $W_{m'} = CW_{max} = 2^{m'} \cdot W$, where m' is the CW increasing factor. Once W_i reaches CW_{max}, it will remain at the value of CW_{max} until it is reset to CW_{min}. Therefore, the contention window size is given by:

$$\begin{cases} W_i = 2^i \cdot W & i \le m' \\ W_i = 2^{m'} \cdot W & i > m' \end{cases} \tag{3.1}$$

where $i \in [0,m]$ and m represents the station's short retry count. Here m is also the maximum backoff stage. The contention window is reset to CW_{min} in the following cases: (a) after the successful transmission of a data packet, (b) when SSRC reaches the short retry limit (retry attempts shall cease and the packet shall be discarded). The SSRC is reset to 0 whenever a packet is discarded or a CTS is received in response to a RTS or an ACK is received in response to a data packet when RTS/CTS is not used.

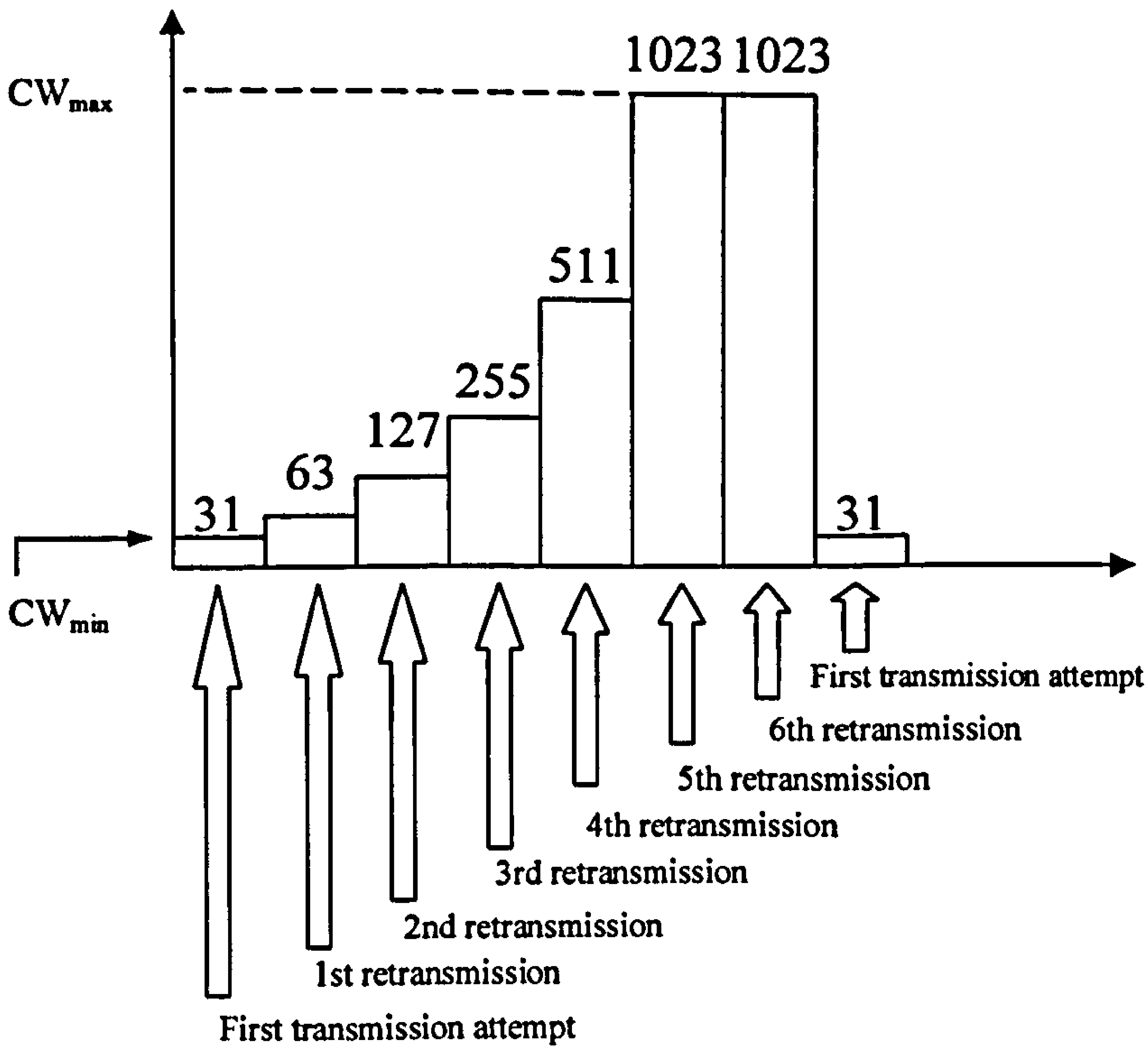


Figure 3.5 The exponential increase of CW

3.2 Mathematical analysis

This section presents a mathematical analysis which is divided into two distinct parts. First, we study the behavior of a single station utilizing a discrete-time Markov chain model and we obtain the stationary probability that the station transmits a packet in a generic (i.e., randomly chosen) slot time. Then, by studying the events that can occur within a generic slot time, we derive the following metrics, which are good indicators of the IEEE 802.11 protocol performance; throughput efficiency, average packet delay, probability of a packet being discarded when it reaches the maximum retransmission limit, the average time to drop a packet and the packet inter-arrival time. The derived performance analysis does not depend on the access mechanism and can be easily applied to both the basic access and RTS/CTS medium access mechanisms.

The mathematical analysis that follows makes use of the same assumptions as in [6] and [139] in order to analyse and study the performance of the IEEE 802.11 protocol. We assume that the network consists of a finite number of n contending stations using the same channel access mechanism in ideal channel conditions (no channel bit errors or hidden stations). We also consider saturation conditions; every station has always a packet ready for transmission (its transmission queue is always non-empty), immediately after every successful packet transmission. The key assumption of our analysis is that the collision probability p of a data packet transmission is constant and independent of the number of collisions the packet has suffered in the past. It is intuitive that the accuracy of this assumption increases as long as W and n become larger. In fact, probability p will be referred to as conditional collision probability, meaning that this is the probability of a collision seen by a packet being transmitted on the channel. Next, we utilize a discrete-time Markov chain model for depicting the backoff procedure followed by each station as in [6].

3.2.1 Calculation of the packet transmission probability

Let $b(t)$ be the stochastic process that represents the backoff timer for a specific station and $s(t)$ be the stochastic process representing the backoff stage $[0,\ldots, m]$ for a given station at time t, where m is the packet retry limit. A discrete integer time scale is adopted; t and $t+1$ correspond to the beginning of two consecutive slot times and the backoff timer of each station decrements at the beginning of each slot time. The process $b(t)$ corresponds to the number of the remaining slot times before a packet transmission and does not represent the remaining time before a transmission attempt. Since a successful packet transmission or a packet collision by other stations may take place between two consecutive slot times, the adopted discrete time scale does not directly relate to system real time. As explained earlier, the backoff timer is "frozen" when the medium is sensed busy and is reactivated again when the medium is sensed idle. For this reason, the time interval between two consecutive slot times for a station may be much longer that the slot time size σ, due to the fact that it could include a packet transmission by another station. Note that with the term slot time we will refer to either the (constant) value σ or the (variable) time interval between two consecutive backoff timer decrements. Since the value of the backoff counter of each station depends on its transmission history (e.g., how many collisions and retransmissions the head-of-line packet has suffered in the past), the stochastic process $b(t)$ is non-Markovian.

Based on the assumption that each packet collides with the same constant probability p regardless of the number of retransmissions the packet has suffered in the past, we utilize the discrete-time Markov chain depicted in figure 3.6 to model the bi-dimensional process $\{s(t), b(t)\}$. Lets assume that a station's bidimentional process $\{s(t), b(t)\}$ is currently at state (i,k), $i=0,1,\ldots,m$ and $k=0,1,\ldots,W_i-1$. As $s(t)=i$, the station's current CW value is W_i (given by equation 3.1) as $b(t)=k$, the station will defer k slots before transmitting a packet. As $b(t)$ decrements at the discrete time scale, a decrement may correspond to an empty idle slot or to a packet collision by other stations or to a successful reservation of another station. After k steps, the station reaches state $(i,0)$ and transmits a packet. As this transmission collides with probability p, the station will transit to state $(i+1,k)$ with probability p/W_{i+1}, where k is randomly selected in the range $[0,W_{i+1}-1]$; as this transmission is successful with probability $(1-p)$, the station will transit to state $(0,k)$ with probability $(1-p)/W_0$, where k is randomly selected in the range $[0,W_0-1]$.

We adopt the same short notation $P\{i_1,k_1 \mid i_0,k_0\} = P\{s(t+1)=i_1, b(t+1)=k_1 \mid s(t)=i_0, b(t)=k_0\}$ used in [6]. The state transition diagram for this Markov chain model has the following non-null one-step transition probabilities:

1. At the beginning of each slot time, the slot time is idle and the backoff counter is decremented by 1.

$$P\{i,k \mid i,k+1\} = 1 \qquad k \in [0, W_i - 2] \quad i \in [0, m]$$

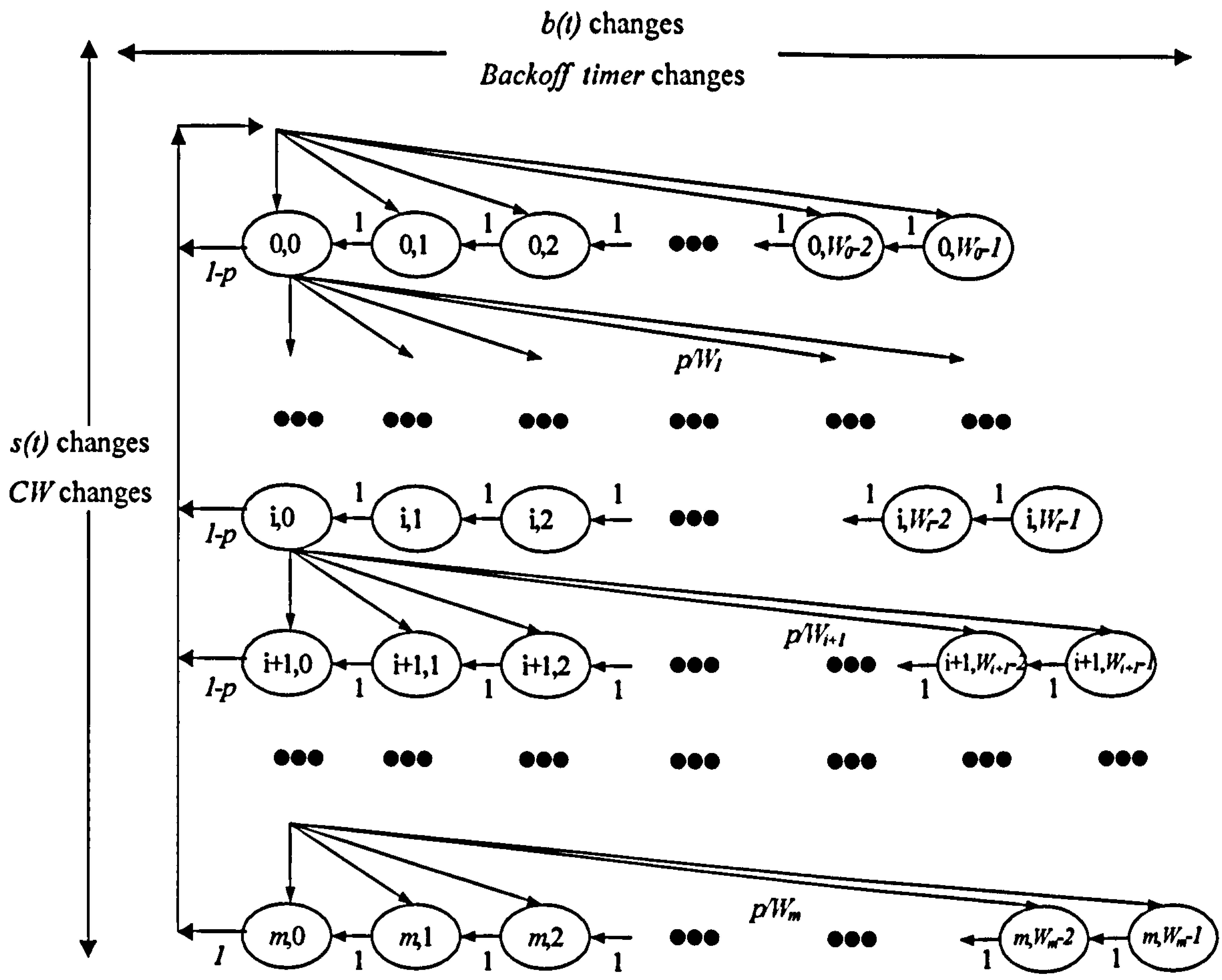


Figure 3.6 Markov chain model

2. After a successful transmission at backoff stage i, the backoff counter of a new packet will reset and start from backoff stage 0. The new value of the backoff timer is uniformly chosen in the interval $[0,W_0\text{-}1]$.

$$P\{0,k \mid i,0\} = (1-p)/W_0 \qquad k \in [0,W_0-1] \quad i \in [0,m-1]$$

3. When an unsuccessful transmission occurs at backoff stage $i\text{-}1$, the backoff increases and the new value of the backoff timer is uniformly chosen in the range $[0,W_i\text{-}1]$.

$$P\{i,k \mid i-1,0\} = p/W_i \qquad k \in [0,W_i-1] \quad i \in [1,m]$$

4. At the maximum backoff stage m, the contention window (CW) is reset to $CW_{min}=W_0$ either after a successful packet transmission or because the retry limit is

reached (the packet will be discarded). In both cases, the backoff mechanism is invoked for a new packet from backoff stage 0.

$$P\{0,k \mid m,0\} = 1/W_0 \qquad k \in [0, W_0 - 1]$$

In order to obtain a closed-form solution for the considered Markov chain, let $b_{i,k} = \lim_{t\to\infty} P\{s(t) = i, b(t) = k\}$ be the stationary distribution of this Markov chain, where $i \in [0,m]$, $k \in [0, W_i - 1]$. Considering that $b_{1,0} = p \cdot b_{0,0}$ and $b_{2,0} = p \cdot b_{1,0} = p^2 \cdot b_{0,0}$, we have the following relations for $b_{i,0}$:

$$b_{i,0} = p \cdot b_{i-1,0} \qquad 0 < i \le m \tag{3.2}$$

$$b_{i,0} = p^i \cdot b_{0,0} \qquad 0 < i \le m \tag{3.3}$$

Owing to chain regularities, the values of $b_{i,k}$ are given by:

$$b_{i,k} = \frac{W_i - k}{W_i} \cdot \begin{cases} (1-p)\cdot\sum_{j=0}^{m-1} b_{j,0} + b_{m,0} & , \quad i = 0 \\ p \cdot b_{i-1,0} & , \quad 0 < i \le m \end{cases} \tag{3.4}$$

By means of equations (3.2) and (3.3) and imposing that $\sum_{j=0}^{m-1} b_{j,0} = b_{0,0} \cdot \frac{1-p^m}{(1-p)}$, equation (3.4) becomes:

$$b_{i,k} = \frac{W_i - k}{W_i} \cdot b_{i,0} \qquad 0 \le i \le m \quad , \quad 0 \le k \le W_i - 1 \tag{4.5}$$

Equations (3.3) and (3.5) express all $b_{i,k}$ values as a function of $b_{0,0}$ and p. Applying the normalization condition for this stationary distribution:

$$1 = \sum_{i=0}^{m} \sum_{k=0}^{W_i - 1} b_{i,k} = \sum_{i=0}^{m} b_{i,0} \cdot \sum_{k=0}^{W_i - 1} \frac{W_i - k}{W_i}$$

$$= \sum_{i=0}^{m} b_{i,0} \cdot \frac{W_i + 1}{2} = \sum_{i=0}^{m} p^i \cdot b_{0,0} \cdot \frac{W_i + 1}{2}$$

$$= \frac{b_{0,0}}{2} \cdot \left(\sum_{i=0}^{m} p^i \cdot W_i + \sum_{i=0}^{m} p^i \right) \tag{3.6}$$

We have to distinguish two different cases according to the values of m and m'.

- When $m > m'$ and by taking into account equation (3.1), equation (3.6) becomes:

$$1=\frac{b_{0,0}}{2}\cdot\left[\sum_{i=0}^{m'}\left((2p)^i\cdot W\right)+\sum_{i=m'+1}^{m}\left(p^i\cdot 2^{m'}\cdot W\right)+\sum_{i=0}^{m}p^i\right]$$

$$=\frac{b_{0,0}}{2}\cdot\left[\frac{1-(2p)^{m'+1}}{1-2p}\cdot W+2^{m'}\cdot W\cdot p^{m'+1}\cdot\frac{1-p^{m-m'}}{1-p}+\frac{1-p^{m+1}}{1-p}\right]$$

from which:

$$b_{0,0}=\frac{2\cdot(1-2p)\cdot(1-p)}{W\cdot(1-(2p)^{m'+1})\cdot(1-p)+(1-2p)\cdot(1-p^{m+1})+W\cdot 2^{m'}\cdot p^{m'+1}\cdot(1-2p)\cdot(1-p^{m-m'})} \tag{3.7}$$

- When $m\leq m'$ and by considering equation (3.1), equation (3.6) turns into:

$$1=\frac{b_{0,0}}{2}\cdot\left[\sum_{i=0}^{m}\left((2p)^i\cdot W\right)+\sum_{i=0}^{m}p^i\right]$$

$$=\frac{b_{0,0}}{2}\cdot\left[\frac{1-(2p)^{m+1}}{1-(2p)}\cdot W+\frac{1-p^{m+1}}{1-p}\right]$$

from which:

$$b_{0,0}=\frac{2\cdot(1-2p)\cdot(1-p)}{W\cdot(1-(2p)^{m+1})\cdot(1-p)+(1-2p)\cdot(1-p^{m+1})} \tag{3.8}$$

Finally, $b_{0,0}$ is given by equation (3.9) and depends on the values of m and m'.

$$b_{0,0}=\begin{cases}\dfrac{2\cdot(1-2p)\cdot(1-p)}{W\cdot(1-(2p)^{m+1})\cdot(1-p)+(1-2p)\cdot(1-p^{m+1})} & ,\ m\leq m'\\[2ex] \dfrac{2\cdot(1-2p)\cdot(1-p)}{W\cdot(1-(2p)^{m'+1})\cdot(1-p)+(1-2p)\cdot(1-p^{m+1})+W\cdot 2^{m'}\cdot p^{m'+1}\cdot(1-2p)\cdot(1-p^{m-m'})} & ,\ m>m'\end{cases} \tag{3.9}$$

Using the previous analysis, we can derive the probability τ that a station transmits a packet in a randomly chosen slot time. Note that a packet transmission occurs when the backoff timer of the transmitting station is equal to zero, regardless of the backoff stage. By utilizing the previous Markov chain model, the probability τ that a station transmits a packet in a randomly chosen slot time is equal to:

$$\tau=\sum_{i=o}^{m}b_{i,0}=\sum_{i=o}^{m}p^i\cdot b_{0,0}=b_{0,0}\cdot\frac{1-p^{m+1}}{(1-p)} \tag{3.10}$$

and $b_{0,0}$ can be acquired from equation (3.9). From equation (3.10) we observe that the transmission probability τ depends on the collision probability p, which is still unknown, and it will be derived next. The probability p that a transmitted packet encounters a collision is the probability that at least one of the $n-1$ remaining stations transmit in the same time slot. If we assume that all stations see the system at steady state and transmit with probability τ, the collision probability p is given by:

$$p = 1 - (1 - \tau)^{n-1} \tag{3.11}$$

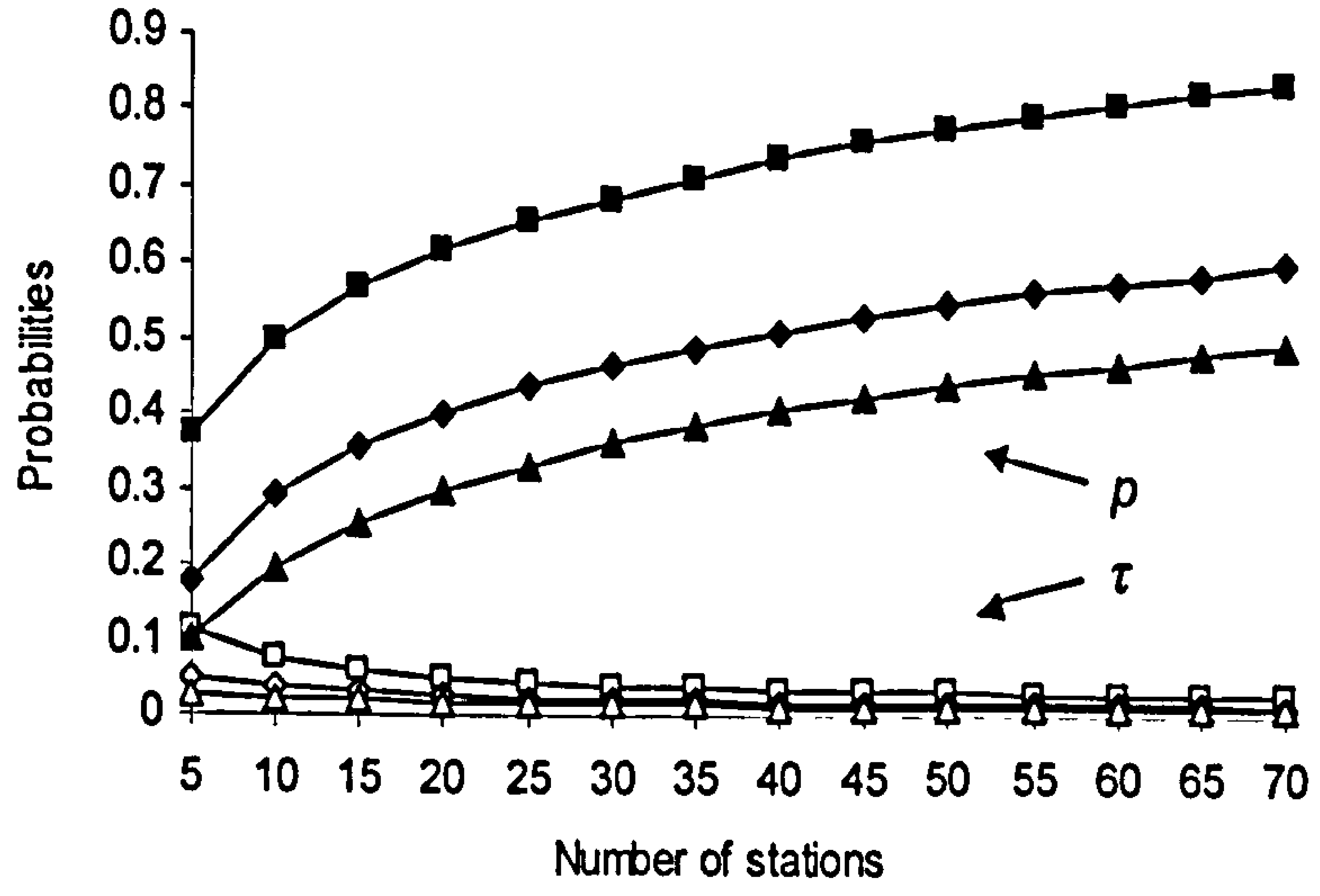
Equations (3.10) and (3.11) represent a non-linear system with two unknowns τ and p, which can be solved utilizing numerical methods. Note that $p \in [0,1]$ and $\tau \in [0,1]$. This non-linear system has a unique solution (detailed proof of the uniqueness is provided in the Appendix C).

Figure 3.7 shows the conditional collision probability p and the transmission probability τ as function of the number of stations and the initial contention window (*CW*) size. We observe that when the network size increases, more packet collisions take place as a result of the higher collision probability. Moreover, the collision probability p is significantly affected by the values of the initial *CW* size; higher W values result in a lower collision probability and, thus, less packet collisions. More contending stations and larger W sizes reduce the transmission probability τ, but ultimately this probability stabilizes to an almost constant value.

3.2.2 Throughput efficiency

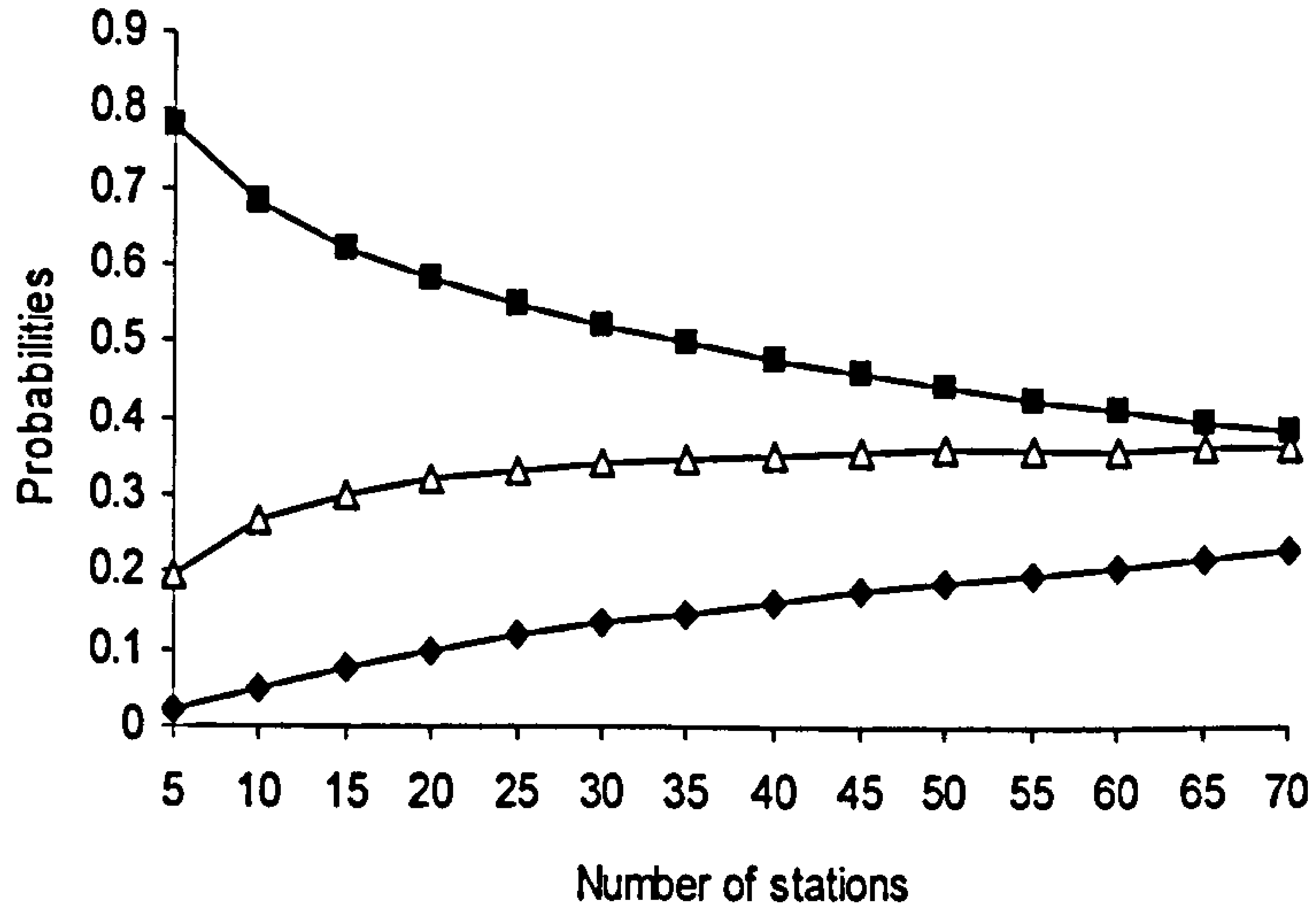
This work utilizes the concept of "saturation throughput" for a finite number of n stations. We assume that a station transmits a data packet of fixed payload size of l bits at a data rate of C Mbit/s. The saturation throughput is defined as the limit reached by the throughput as the offered load increases and represents the maximum load that the system can carry in stable conditions. In particular, as the offered load increases, the throughput grows up to a maximum value, referred to as maximum throughput. However, further increase of the offered load leads to a decrease in the system throughput. More details about the mathematical formulation and interpretation of the unstable behavior of several random access schemes can be found in [6].

Based on the already calculated collision probability p and transmission probability τ, we can now analyse all possible events that can occur in a randomly chosen time slot. Let P_{tr} be the probability that at least one station transmits in the



■ *Collision probability p, W=8*
◆ *Collision probability p, W=32*
▲ *Collision probability p, W=64*
□ *Transmission probability τ, W=8*
◇ *Transmission probability τ, W=32*
△ *Transmission probability τ, W=64*

Figure 3.7 Collision p and transmission τ probabilities versus n



■ *Probability that a randomly selected slot is empty ($1-P_{tr}$)*
△ *Probability that an occurring transmission is successful ($P_s P_{tr}$)*
◆ *Probability that an occurring transmission is not successful ($(1-P_s) P_{tr}$)*

Figure 3.8 Conditional channel probabilities versus n considered slot. Since n stations contend on the channel, each transmitting with probability τ , P_{tr} is given by:

$$P_{tr} = 1-(1-\tau)^n \tag{3.12}$$

A packet collision takes place when two or more contending stations initiate simultaneously a packet transmission in the same time slot. The conditional probability P_s that an occurring packet transmission is successful is given by the probability that exactly one station transmits and the remaining n-1 stations defer transmission, conditioned on the fact that at least one station (out of n stations) transmits:

$$P_s = \frac{n\cdot\tau\cdot(1-\tau)^{n-1}}{P_{tr}} = \frac{n\cdot\tau\cdot(1-\tau)^{n-1}}{1-(1-\tau)^n} \tag{3.13}$$

Figure 3.8 plots the probability that a randomly selected slot is empty (1-P_{tr}), the probability that an occurring transmission is successful ($P_s\ P_{tr}$) and the probability that an occurring transmission is not successful ((1-P_s)P_{tr}) versus the number of stations. We observe that high network sizes result in less empty (idle) slots as well as in an increase of successful and collided transmissions.

A successful transmission in a randomly selected slot occurs with probability $P_{tr}P_s$ and the time transmitting payload information is given by $P_sP_{tr}\ l/C$, where l is the packet payload data length and C is the data rate. The average slot duration can be evaluated by considering that 1-P_{tr} is the probability that the slot is empty; $P_{tr}P_s$ is the probability that the slot contains a successful transmission and P_{tr}(1-P_s) is the probability that the slot contains a collision. Throughput efficiency S can thus be evaluated as in [6] by dividing the time utilized for transmitting payload information in a slot time by the average duration of a slot time $E[slot]$:

$$S = \frac{P_{tr}P_s\, l/C}{E[slot]} = \frac{P_{tr}P_s\, l/C}{(1-P_{tr})\sigma + P_{tr}P_sT_s + P_{tr}(1-P_s)T_c} \tag{3.14}$$

where σ is the duration of an empty slot time, T_s and T_c are the time durations the medium is sensed busy due to a successful transmission and a collision, respectively. The throughput efficiency can also be expressed as a function of the transmission probability τ and number of contending stations n:

$$S = \frac{n \cdot \tau(1-\tau)^{n-1} \, l/C}{(1-\tau)^n \sigma + n\tau(1-\tau)^{n-1} T_S + [1-(1-\tau)^n - n\tau(1-\tau)^{n-1}] T_C} \tag{3.15}$$

3.2.3 Event timing

The system is in one of the three states in each generic time slot; no transmission (idle), successful transmission (success) or unsuccessful transmission (collision). The durations σ, T_S and T_C of the three respective states depend on the medium access mechanism and on various MAC and PHY parameters. The T_S and T_C values are defined for the basic and the RTS/CTS access mechanisms as follows (see figure 3.9):

$$\begin{cases} T_S^{bas} = DIFS + T_{DATA} + \delta + SIFS + T_{ACK} + \delta \\ T_C^{bas} = DIFS + T_{DATA} + \delta + SIFS + T_{ACK} + \delta \end{cases} \tag{3.16}$$

$$\begin{cases} T_S^{RTS} = DIFS + T_{RTS} + \delta + SIFS + T_{CTS} + \delta + SIFS + T_{DATA} + \delta + SIFS + T_{ACK} + \delta \\ T_C^{RTS} = DIFS + T_{RTS} + \delta + SIFS + T_{CTS} + \delta \end{cases} \tag{3.17}$$

where T_{DATA} , T_{RTS} , T_{CTS} and T_{ACK} is the transmission time for a data, RTS, CTS and acknowledgement packet, respectively and δ is the propagation delay. The duration of these time intervals varies for different physical layers. Next, we consider the IEEE 802.11b and 802.11a physical (PHY) layers.

A. IEEE 802.11b PHY layer

For the IEEE 802.11b physical layer [136], the above time intervals are given by:

$$T_{DATA} = T_{header} + \frac{l}{C} \tag{3.18}$$

$$T_{header} = \frac{MAC_{hdr}}{C} + \frac{PHY_{hdr}}{C_{con}} \tag{3.19}$$

$$T_{RTS} = \frac{l_{RTS}}{C_{con}} \tag{3.20}$$

$$T_{CTS} = \frac{l_{CTS}}{C_{con}} \tag{3.21}$$

$$T_{ACK} = \frac{l_{ACK}}{C_{con}} \tag{3.22}$$

where T_{header} is the time required to transmit the packet payload header, C is the data rate, MAC_{hdr} and PHY_{hdr} is the MAC and the physical header (in bits), C_{con} is the rate

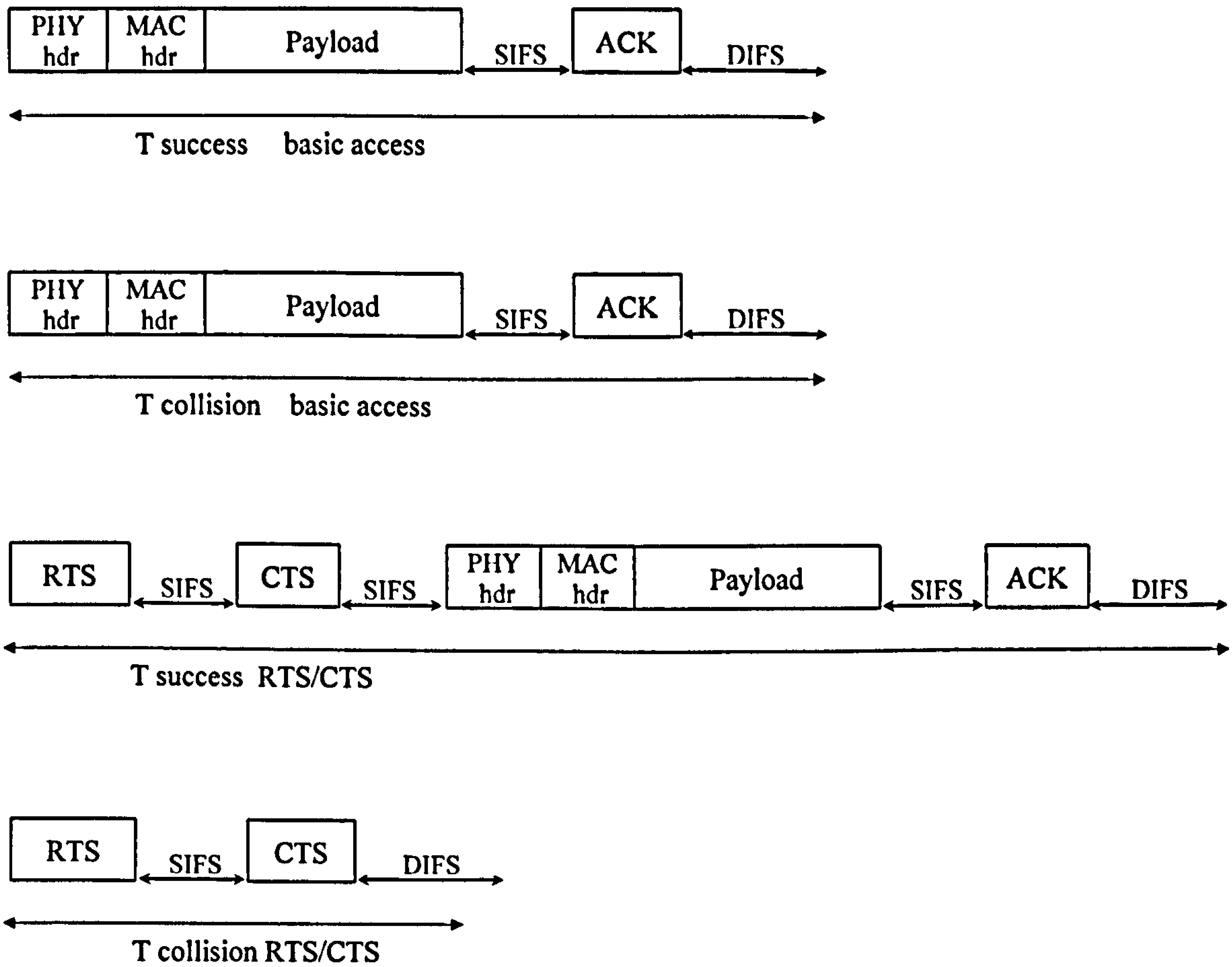


Figure 3.9 Event timing for basic and RTS/CTS access mechanisms

at which the control packets (RTS, CTS and ACK) are transmitted, l_{RTS}, l_{CTS}, and l_{ACK} the length of RTS, CTS and ACK respectively. In order to ensure that the vital information contained in the RTS and CTS packets will be received by all stations within range and to cope with potential hidden stations, control packets are transmitted at a lower data rate which increases the posible reception distance. Note that the data C and the control $C_{control}$ rates may not be the same.

B. IEEE 802.11a PHY layer

For the IEEE 802.11a physical layer [137], the above time intervals are given by:

$$T_{DATA} = T_p + T_{header} + T_{SYM} \times ceiling\left(\frac{16+6+MAC_{hdr}+FCS+l}{4C}\right) = 20us + 4us \times ceiling\left[\frac{294+l}{4C}\right] \quad (3.23)$$

$$T_{RTS} = T_p + T_{header} + T_{SYM} \times ceiling\left(\frac{16+6+160}{4C_{con}}\right) = 20us + 4us \times ceiling\left[\frac{182}{4C_{con}}\right] \quad (3.24)$$

$$T_{CTS} = T_{ACK} = T_p + T_{header} + T_{SYM} \times ceiling\left(\frac{16+6+112}{4C_{con}}\right) = 20us + 4us \times ceiling\left[\frac{134}{4C_{con}}\right] \quad (3.25)$$

where T_P and T_{SYM} is the transmission time for the physical preamble and a symbol, respectively, C is the data rate at which data packets are transmitted (6, 9, 12, 18, 24, 36, 48, and 54 Mbit/s) and C_{con} is the control rate at which the RTS, CTS and ACK control packets are transmitted (6, 12 or 24 Mbit/s). Once more, note that the data and control rate may not be the same. More details on the exact calculation of the duration of the above time intervals can be found in the Appendix A.

The duration of each delay component is determined from the standards [135][136][137]. Certain delay components (i.e. DIFS, SIFS, T_{RTS} , T_{CTS} and T_{ACK}) vary with the PHY layer technology but not with the data rate. The transmission of an MPDU depends on its size and data rate. Each station has a (*data rate, control rate*) pair. As explained earlier, control packets such as RTS, CTS and ACK are transmitted at the control rate, which may not be the same to the data rate. Tables 3.1 and 3.2 list the constant and varying delay components for basic access and RTS/CTS schemes, for all the different PHY layers specified in the IEEE 802.11 standards.

Scheme	Constant and varying delay components (in μs)					
	DIFS	SIFS	T_{RTS}	T_{CTS}	T_{ACK}	T_{DATA}
Basic access						
Infrared-1 (1,1)	26	10	N/A	N/A	169	57 + (34 + MSDU)/1
Infrared-2 (2,1)	26	10	N/A	N/A	169	57 + (34 + MSDU)/2
FHSS-1 (1,1)	128	28	N/A	N/A	240	128 + (34 + MSDU)/1
FHSS-2 (2,1)	128	28	N/A	N/A	240	128 + (34 + MSDU)/2
DSSS-1 (1,1)	50	10	N/A	N/A	304	192 + (34 + MSDU)/1
DSSS-2 (2,1)	50	10	N/A	N/A	304	192 + (34 + MSDU)/2
HR-5.5 (5.5, 2)	50	10	N/A	N/A	152	57 + (34 + MSDU)/5.5
HR-11 (11, 2)	50	10	N/A	N/A	152	57 + (34 + MSDU)/11
OFDM-6 (6,6)	34	9	N/A	N/A	44	20+ 4 x [(16+6 +(34+MSDU))/24]
OFDM-12 (12,12)	34	9	N/A	N/A	32	20+ 4 x [(16+6 +(34+MSDU))/48]
OFDM-24 (24,24)	34	9	N/A	N/A	28	20+ 4 x [(16+6 +(34+MSDU))/96]
OFDM-54 (54,24)	34	9	N/A	N/A	24	20+ 4 x [(16+6 +(34+MSDU))/216]

Table 3.1 Delay components for Basic access scheme and different IEEE 802.11 PHY layers

Scheme	Constant and varying delay components (in μs)					
	DIFS	SIFS	T_{RTS}	T_{CTS}	T_{ACK}	T_{DATA}
RTS/CTS						
Infrared -1 (1,1)	26	10 x 3	217	169	169	57 + (34 + MSDU)/1
Infrared -2 (2,1)	26	10 x 3	217	169	169	57 + (34 + MSDU)/2
FHSS-1 (1,1)	128	28 x 3	288	240	240	128 + (34 + MSDU)/1
FHSS-2 (2,1)	128	28 x 3	288	240	240	128 + (34 + MSDU)/2
DSSS-1 (1,1)	50	10 x 3	352	304	304	192 + (34 + MSDU)/1
DSSS-2 (2,1)	50	10 x 3	352	304	304	192 + (34 + MSDU)/2
HR-5.5 (5.5, 2)	50	10 x 3	176	152	152	57 + (34 + MSDU)/5.5
HR-11 (11, 2)	50	10 x 3	176	152	152	57 + (34 + MSDU)/11
OFDM-6 (6, 6)	34	9 x 3	52	44	44	20+ 4 x [(16+6 +(34+MSDU))/24]
OFDM-12 (12, 12)	34	9 x 3	36	32	32	20+ 4 x [(16+6 +(34+MSDU))/48]
OFDM-24 (24, 24)	34	9 x 3	28	28	28	20+ 4 x [(16+6 +(34+MSDU))/96]
OFDM-54 (54, 24)	34	9 x 3	24	24	24	20+ 4 x [(16+6 +(34+MSDU))/216]

Table 3.2 Delay components for RTS/CTS access scheme and different IEEE 802.11 PHY layers

3.2.4. Relative comparison of events duration

The previous analytical model allows measurement of the time portion utilized on all events affecting 802.11 performance. Such an evaluation reveals the impact of physical and link layer parameters on performance. Considering that a randomly selected slot is empty with probability $1\text{-}P_{tr}$ and that the empty slot duration is σ, then the relative time utilized in empty slots compared to the expected time slot is given by:

$$U_{empty} = \frac{(1-P_{tr})\sigma}{(1-P_{tr})\sigma + P_{tr}P_S T_S + P_{tr}(1-P_S)T_C} \tag{3.26}$$

A randomly selected slot could be in a collision event with probability $P_{tr}(1\text{-}P_s)$ and the relative time duration utilized on collisions when two or more stations are simultaneously trying to transmit is:

$$U_{coll} = \frac{P_{tr}(1-P_s)T_C}{(1-P_{tr})\sigma + P_{tr}P_S T_S + P_{tr}(1-P_S)T_C} \tag{3.27}$$

The relative time duration utilized on transmitting data packet overheads, reservation control packets (RTS, CTS, and ACK) is:

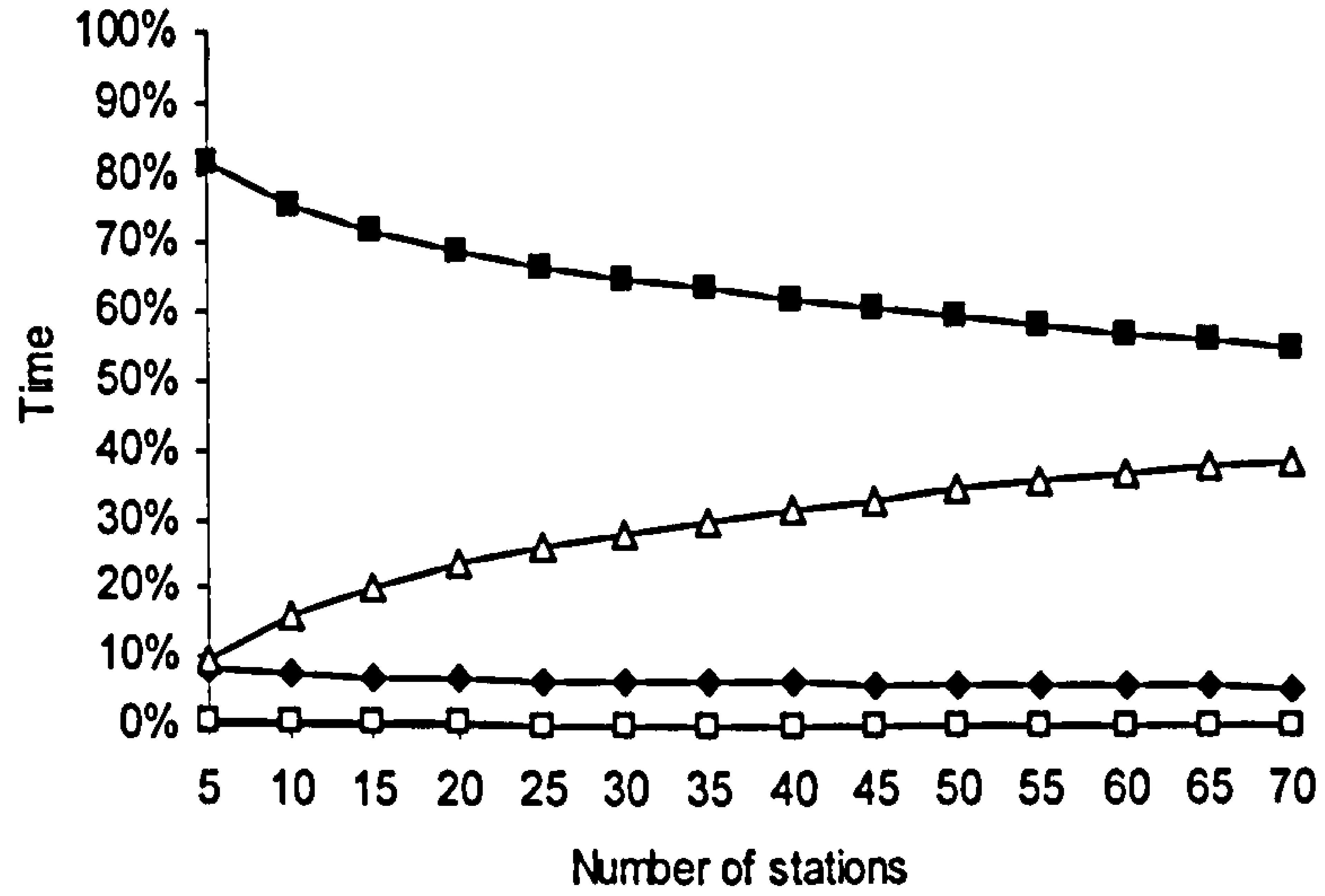
$$U_{over} = \frac{P_{tr}P_s\left(T_s - \frac{l}{C}\right)}{(1-P_{tr})\sigma + P_{tr}P_S T_S + P_{tr}(1-P_S)T_C} \tag{3.28}$$

As all component events that affect IEEE 802.11 performance are considered, the following equation always holds true:

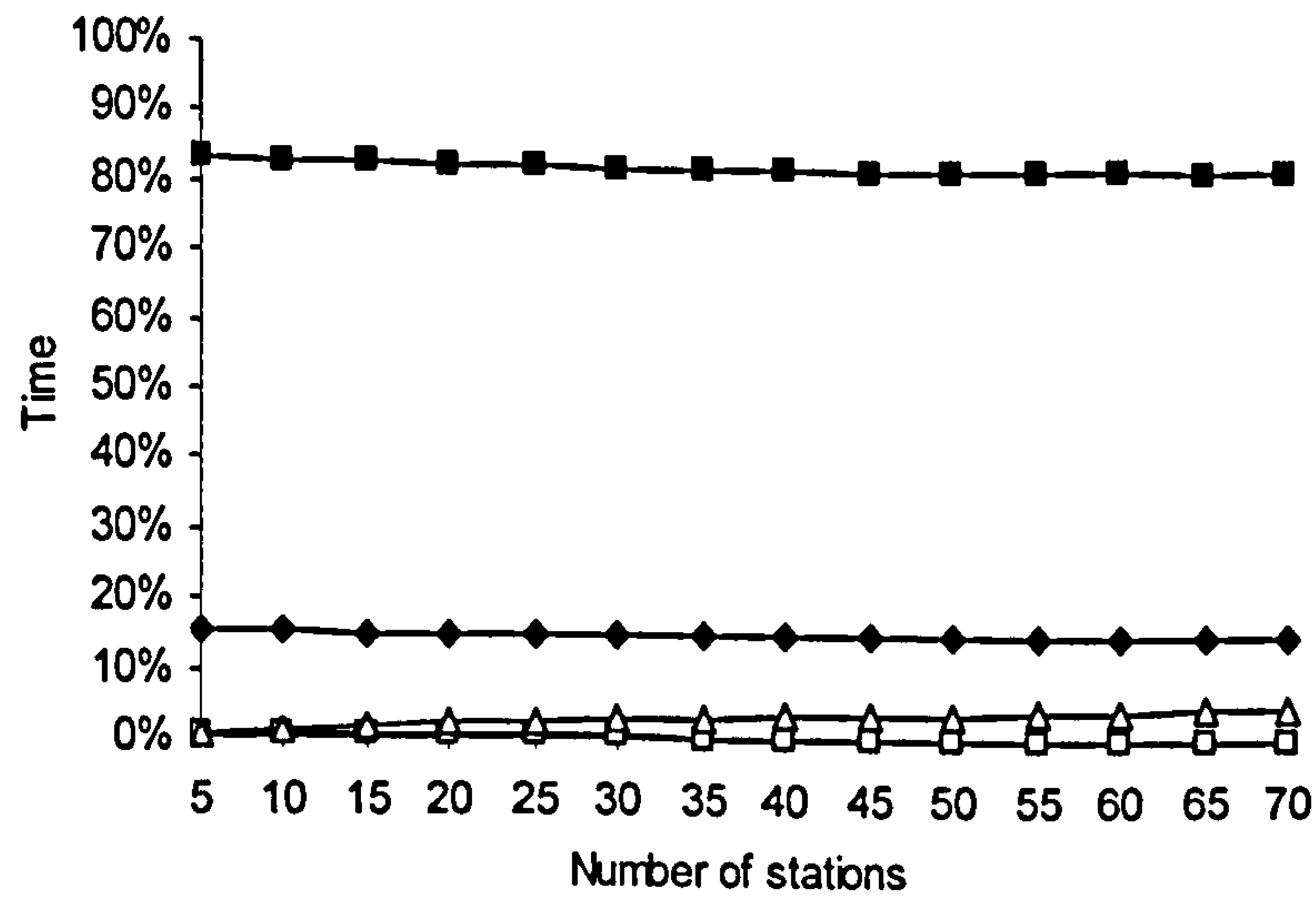
$$U_{empty} + U_{coll} + U_{over} + S = 1 \tag{3.29}$$

where S is the throughput efficiency (useful payload data transmission) given in equation (3.14). Equation (3.29) can be easily verified from equations (3.26), (3.27) and (3.28).

Figures 3.10 (a) and 3.10 (b) display the relative % time for each of the event components of the protocol versus network size for basic access and RTS/CTS schemes, respectively. The time component utilized by empty slot times, which represents the amount of time the channel is idle, does not highly depend on the employed access mechanism. The figures reveal that for both medium access schemes this time component is always minimal and the main factors affecting performance are the time portions utilized in collisions and transmitting overheads. In fact, when the basic access scheme is employed throughput efficiency significantly degrades for large network scenarios. The situation is explained by considering that the percentage time due to collisions highly increases when n increases. On the contrary, the RTS/CTS mechanism appears to be almost insensitive on the network size since it significantly reduces the time percentage for collisions (according to equations (3.16)-(3.17)) relative to the basic access mechanism. This reduction is extremely effective for larger network sizes. In this case, the additional amount of time due to collisions is extremely large for basic access compared to the RTS/CTS mechanism regardless the network size. Finally, the drawback of RTS/CTS due to the additional overhead introduced by the exchange of the RTS and CTS control packets, becomes noticeable in figure 3.9. As expected, the RTS/CTS mechanism takes considerably longer time for transmitting overheads compared to the basic access. This will eventually turn out to be a significant shortcoming of the RTS/CTS effectiveness at high data rates and is studied later on chapter 4.



(a) *Basic access*



(b) *RTS/CTS scheme*

■ *Useful data transmission (throughput efficiency)* △ *Collisions*

◆ *Transmitting overheads* □ *Empty (idle) slots*

Figure 3.10 Time allocation of various 802.11 tasks versus n for basic access and RTS/CTS

3.2.5. Packet drop probability

The packet drop probability is defined as the probability of a packet being dropped when the retry limit is reached. A packet is found in the last backoff stage m, if it encounters m collisions in the previous stages and it will be discarded if it experiences another collision. Therefore, packet drop probability is independent of the employed access mechanism (basic access or RTS/CTS) and can be expressed as a function of the last backoff stage (by means of equation (3.1)) and the collision probability p as:

$$p_{drop} = \frac{b_{m,0}}{b_{0,0}} p = p^m p = p^{m+1} \tag{3.30}$$

3.2.6. Average packet delay

Next, we calculate the delay D for a successfully transmitted packet, which is defined to be the time interval from the instance a head-of-queue packet is ready for transmission, until an acknowledgement for this packet is received (until its successful reception). When retry limits are considered, packet delay cannot be simply obtained from throughput (as shown in Appendix B); the calculation of the average number of slot times needed for a successful packet transmission is necessary. If a packet is dropped because it has reached the specified retry limit, the time delay for this packet will not be included in the calculation of the average packet delay since this packet is not successfully received. The average packet delay $E[D]$ is given by:

$$E[D] = E[X] \cdot E[slot] \tag{3.31}$$

where $E[X]$ is the average number of time slots required for a successful packet transmission and $E[slot]$ the average slot time (can be found in equation (3.14)). $E[X]$ can be found by multiplying the number of slot times d_i the packet is delayed in each backoff stage by the probability q_i that a packet that is not dropped, arrives at the i backoff stage:

$$E[X] = \sum_{i=0}^{m} d_i \cdot q_i \tag{3.32}$$

The average number of time slots a station utilises in stage i (including the transmission slot) d_i is given by:

$$d_i = \frac{W_i + 1}{2} \quad , \ i \in [0, m] \tag{3.33}$$

The probability q_i that a packet reaches the i backoff stage, provided that this packet is not discarded, is given by:

$$q_i = \frac{(p^i - p^{m+1})}{1 - p^{m+1}} \quad , \ i \in [0, m] \tag{3.34}$$

since packets that are not dropped (with probability $1-p^{m+1}$) arrive at the i stage with probability $\left(p^{i}-p^{m+1}\right)$ (we have to deduct the probability p^{m+1} of dropped packets from the probability p^{i} of the total number of packets arriving at the i stage).

Combining equations (3.32), (3.33) and (3.34), $E[X]$ is given by:

$$E[X]=\sum_{i=0}^{m}\left[\frac{(p^{i}-p^{m+1})\frac{W_i+1}{2}}{1-p^{m+1}}\right] \tag{3.35}$$

After some algebra, equation (3.35) becomes:

$$E[X]=\begin{cases}\dfrac{W\cdot(1-(2p)^{m+1})\cdot(1-p)+(1-2p)\cdot(1-p^{m+1})}{2\cdot(1-2p)\cdot(1-p)\cdot(1-p^{m+1})}-\dfrac{p^{m+1}}{1-p^{m+1}}\cdot E[T_{drop}] & , m\leq m' \\[2ex] \dfrac{W\cdot(1-(2p)^{m'+1})\cdot(1-p)+W\cdot 2^{m'}\cdot p^{m'+1}\cdot(1-p^{m-m'})\cdot(1-2p)+(1-2p)\cdot(1-p^{m+1})}{2\cdot(1-2p)\cdot(1-p)\cdot(1-p^{m+1})}-\dfrac{p^{m+1}}{1-p^{m+1}}\cdot E[T_{drop}] & , m>m'\end{cases} \tag{3.36}$$

Finally, if we substitute equation (3.36) and $E[slot]$ into equation (3.31), the average packet delay $E[D]$ can be easily calculated.

3.2.7. Average time to drop a packet

A packet is dropped when it reaches the last backoff stage and experiences another collision. Using the same approach like for the derivation of average packet delay, the average time to drop a packet $E[D_{drop}]$ because its retry limit is reached is equal to:

$$E[D_{drop}]=E[T_{drop}]\,E[slot] \tag{3.37}$$

where $E[T_{drop}]$ is the average number of time slots required for a packet to experience m+1 collisions in the (0,1,...m) stages and $E[slot]$ is the average slot time. Given that the average number of time slots a station defers in the i stage is (W_i+1)/2 and since a packet utilizes all backoff stages the conditional probability q_i is equal to 1, $E[T_{drop}]$ is calculated as:

$$E[T_{drop}]=\sum_{i=0}^{m}d_i=\sum_{i=0}^{m}\frac{W_i+1}{2}=\begin{cases}\dfrac{W\cdot(2^{m+1}-1)+(m+1)}{2} & , m\leq m' \\[2ex] \dfrac{W\cdot(2^{m'+1}-1)+W\cdot 2^{m'}\cdot(m-m')+(m+1)}{2} & , m>m'\end{cases} \tag{3.38}$$

3.2.8. Packet inter-arrival time

The packet inter-arrival time is defined as the time interval between two successful packet receptions at the receiver and can be simply obtained from throughput:

$$E[D_{inter}] = \frac{l}{S/n} \tag{3.39}$$

Using the same reasoning with equation (3.31), the average packet inter-arrival time $E[D_{inter}]$ is also given by:

$$E[D_{inter}] = \left(\sum_{j=0}^{\infty} \sum_{i=0}^{m} p^{j(m+1)} p^{i} \frac{W_i + 1}{2} \right) E[slot] \tag{3.40}$$

which after some algebra, reaches equation (3.39).

Intuitively, the average packet delay, inter-arrival time and drop time are related by:

$$E[D] = E[D_{inter}] - \frac{p_{drop}}{1 - p_{drop}} E[D_{drop}] \tag{3.41}$$

where $E[D_{inter}]$ is given by either (3.39) or (3.40), $E[D_{drop}]$ is given by (3.37) and $E[D]$ by (3.31). The expression $p_{drop}/1 - p_{drop} = p^{m+1}/1 - p^{m+1}$ represents the average number of dropped packets before for a successful transmission. The expression in (3.41) is of key importance since it gives insights into the delay characteristics of the backoff mechanism of IEEE 802.11 and it relates the average packet delay with the packet inter-arrival time, the packet drop probability and the average time to drop a packet.

3.3 Model validation

The mathematical analysis presented in this work is validated by comparing analytical with simulation results from our IEEE 802.11 simulator. This IEEE 802.11 simulator is developed using the OPNET Modeler communication networks modeling and simulation software package from OPNET Technologies (formerly MIL3 Inc). OPNET Modeler is an event-driven simulator and provides a powerful graphical tool to display simulation statistics. OPNET uses hierarchically linked domains to denote a network design and stations are defined in the network domain, which is the top-level domain. Each station has a set of processes and each process can represent a layer in the protocol stack. A process can be defined by a finite state machine. The transmission of packets across network links is controlled by pipeline-stage C/C++ coded routines. The user can produce and add code to be executed when entering and exiting each state. Finally, the code is accumulated and compiled.

The OPNET 802.11 simulator emulates the real operation of a wireless station as closely as possible, by implementing the collision avoidance procedures and all parameters such as packet transmission times, propagation delays, turnaround times, etc. The simulator closely follows all timer values and packet element transmission times defined by IEEE 802.11 specifications. Furthermore, we have suitably modified the standard library of the OPNET 802.11 simulation package in order to implement a LAN of n stations operating at saturation conditions, i.e. each station always has a packet ready for transmission. A set of simulation runs was taken to examine the performance of the IEEE 802.11 protocol under ideal channel conditions; an error free medium is assumed and no hidden stations are considered.

The Markov chain analysis, presented in the previous section, is independent of physical layer parameters and can be applied to any IEEE 802.11 PHY standard. Unless otherwise specified, the presented analytical and simulation performance results in the following figures have been obtained using the system parameters in table 3.3 specified for the Direct Spread Sequence Spectrum (DSSS) physical layer utilized in IEEE 802.11b.[1]

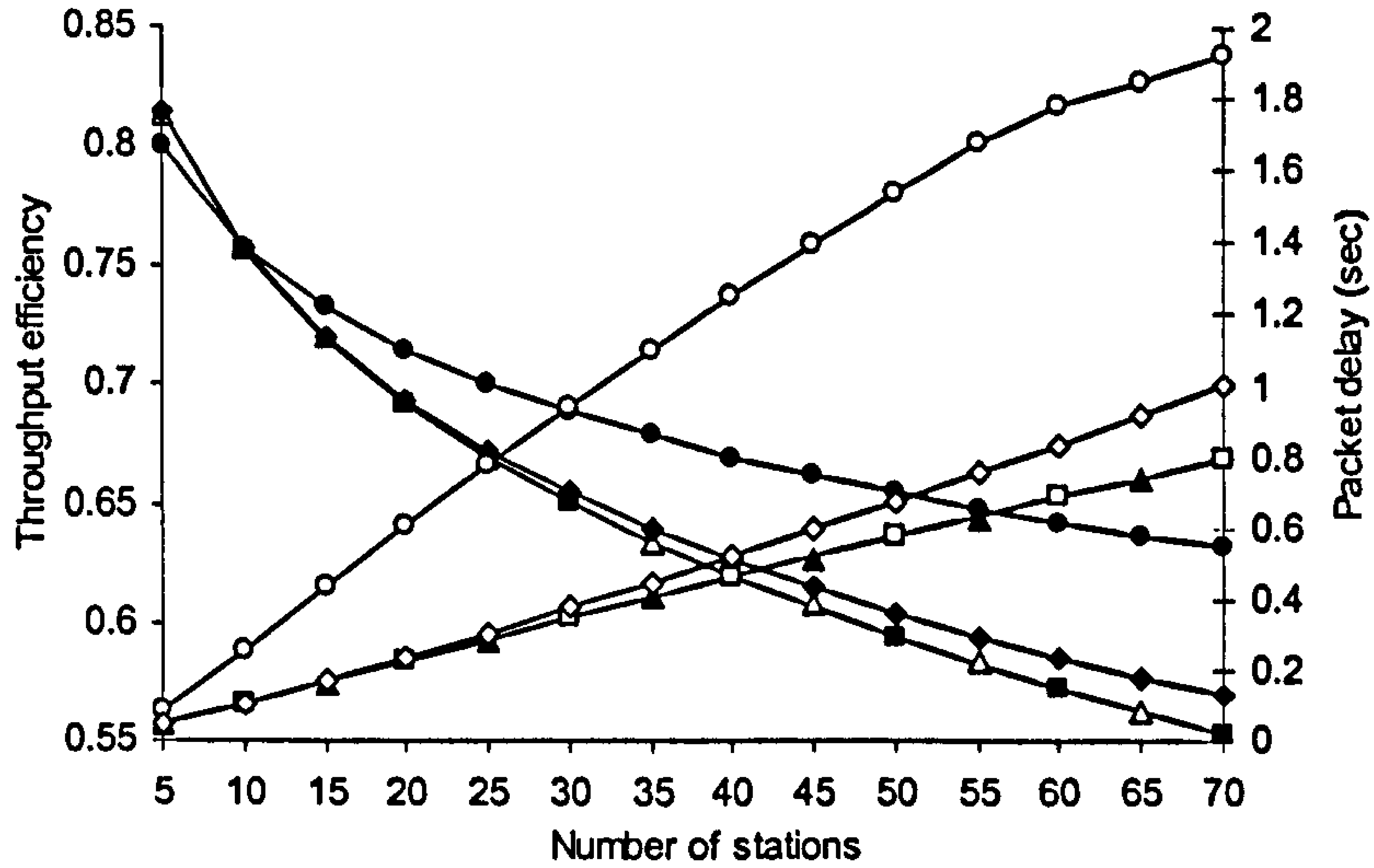
Figures 3.11 and 3.12 confirm the accuracy of the considered modeling assumptions by comparing results obtained from our mathematical analysis and simulation outcome utilizing the IEEE 802.11 simulator developed with the OPNET™ simulation package.

[1] In certain cases, whenever it is of key importance, performance results are also derived for the Orthogonal Frequency Division Multiplexing (OFDM) physical layer utilized in IEEE 802.11a.

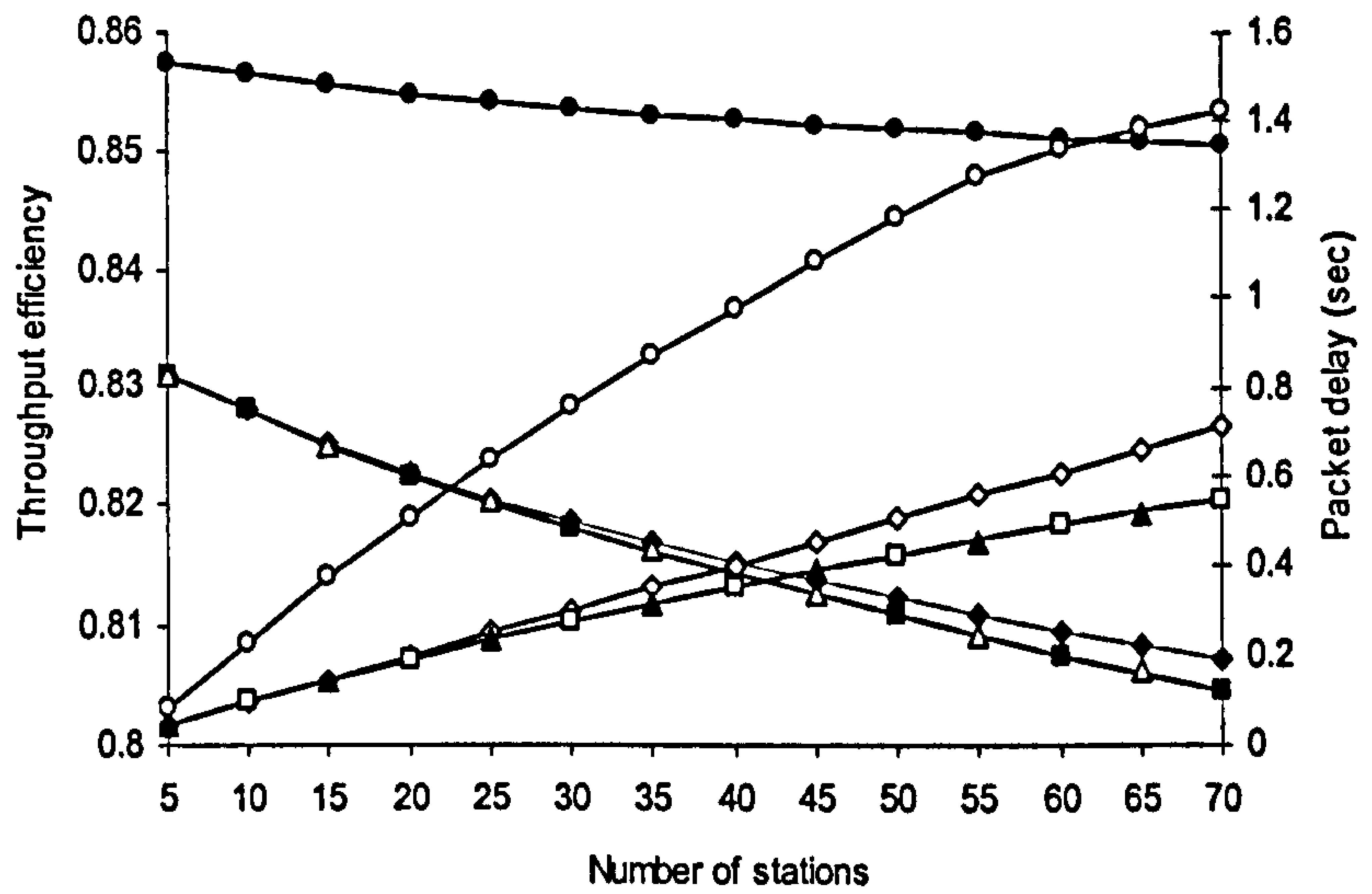
Parameter	802.11 and 802.11b (DSSS)	802.11a (OFDM)
Slot time, σ	20 μs	9 μs
SIFS	10 μs	16 μs
DIFS	50 μs	34 μs
Propagation delay, δ	1 μs	<< 1 μs
Transmission time of physical (PHY) preamble, T_P	144 μs (long, 802.11) 72 μs (short, 802.11b)	16 μs
Transmission time of physical (PHY) header, T_{hdr}	48 μs (long, 802.11) 24 μs (short, 802.11b)	4 μs
Transmission time for a symbol, T_{SYM}	N/A	4 μs
MAC header, MAC_{hdr}	240 bits	240 bits
Frame Check Sequence, FCS	32 bits	32 bits
Channel data rate, C	1, 2, 5.5, 11 Mbit/s	6, 9, 12, 18, 24, 36, 48, 54 Mbit/s
Control (base) rate, C_{con}	1, 2 Mbit/s	6, 12, 24 Mbit/s
Minimum CW, CW_{min}	32	16
Number of CW sizes, m'	5	6
Maximum CW, CW_{max}	1024	1024
Short retry limit	6	6

Table 3.3 Parameter values of IEEE 802.11, 802.11b and 802.11a

The figures provide performance results (throughput efficiency, packet delay, packet drop time and packet drop probability) versus the number of contending stations for the basic access and RTS/CTS mechanisms. Figures 3.11(a) and 3.11(b) plot throughput efficiency against the number of stations, for basic access and RTS/CTS schemes respectively for Bianchi's [6], Wu's [139] and Ziouva's [148] models. As stated previously, Bianchi does not take into account retry limits, Wu introduces packet retry limits and Ziouva utilizes an additional transition state to the models of [6] and [139] that allows stations to transmit consecutive packets without activating the backoff procedure as it was explained in section 2.7. Note that in figure 3.11(a), the vertical axis scale is different to that in figure 3.11(b). The comparison of the analytical models with OPNET simulation results reveals that the more realistic analytical model that considers retry limits predicts very accurately DCF throughput performance, a conclusion not clearly drawn in [139] which added retry limits in the analytical model of [6]. Note that simulation results are acquired with a 95% confidence interval lower than 0.002. Throughput is overestimated from both Bianchi and Ziouva since when a packet



(a) Basic access



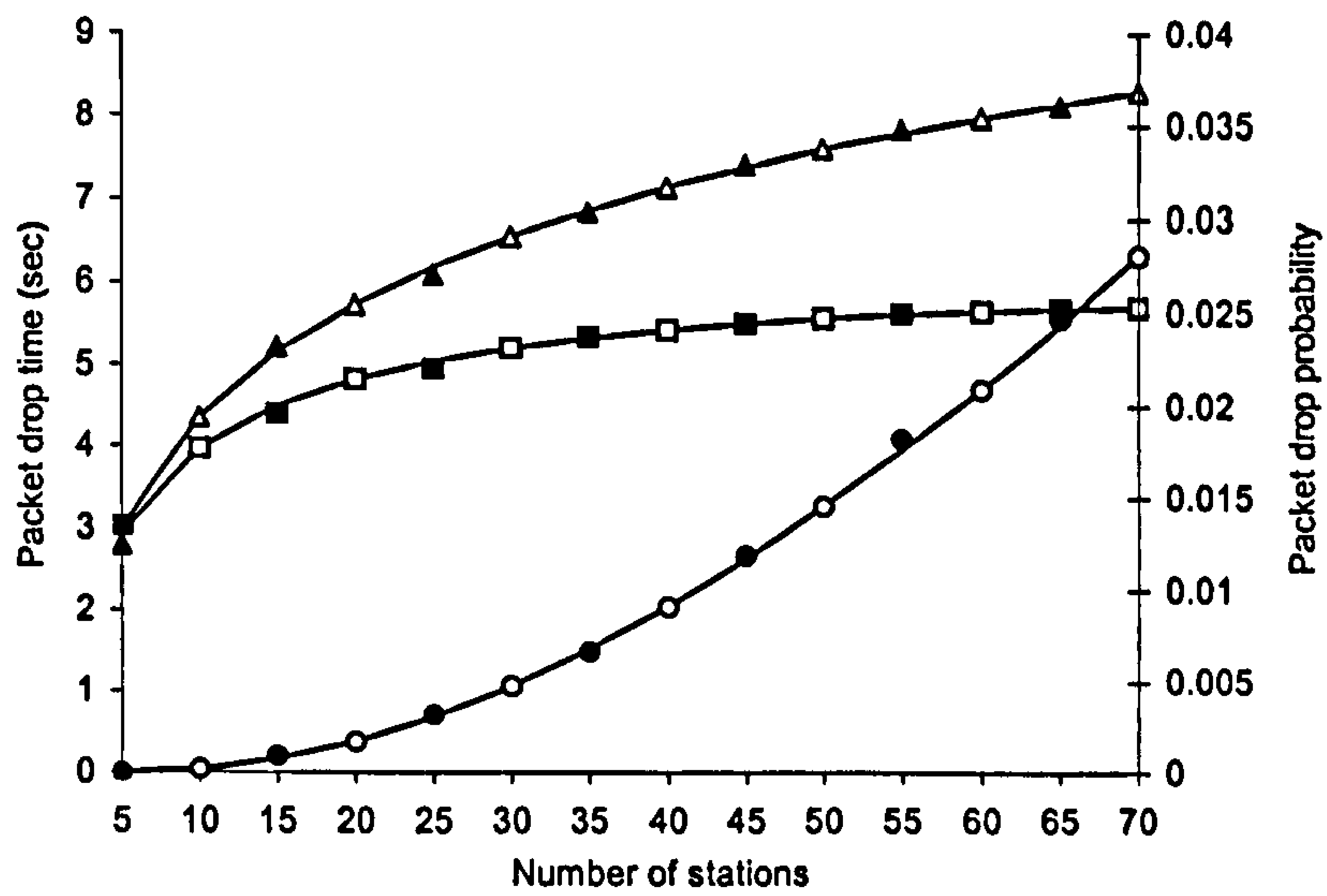
(b) RTS/CTS scheme

◆ *Throughput efficiency, no retry limits (Bianchi)* ◇ *Packet delay, no retry limits (Bianchi)*
● *Throughput efficiency, (Ziouva in[148])* ○ *Packet delay, (Ziouva in[148])*
■ *Throughput efficiency, m=6 (Wu)* □ *Packet delay, m=6 (Wu)*
△ *Throughput efficiency, (OPNET simulation)* ▲ *Packet delay, (OPNET simulation)*

Figure 3.11 Throughput efficiency and packet delay for basic access and RTS/CTS: Analysis versus OPNET simulation

is being retransmitted with no retry limit it reaches higher backoff stages causing decrease of collision probability.

Figures 3.11(a) and 3.11(b) also plot packet delay calculated utilizing our delay analysis, as well as for Bianchi's and Ziouva's models, against OPNET simulation results. Simulation results are again calculated with a 95% confidence interval lower than 0.002. The performance comparison shows that our packet delay analysis gives



▲ *Packet drop time, basic access (simulation)* △ *Packet drop time, basic access (analysis)*

■ *Packet drop time, RTS/CTS (simulation)* □ *Packet drop time, RTS/CTS (analysis)*

● *Packet drop probability, (simulation)* ○ *Packet drop probability (analysis)*

Figure 3.12 Packet drop time and packet drop probability for basic access and RTS/CTS: Analysis versus OPNET simulation

results in high agreement with OPNET simulations. In fact, the assumptions of Bianchi's model lead to an overestimated packet delay since it includes the long time delay of unlimited retransmissions of packets that should have been discarded as specified in the 802.11 standard. Furthermore, we can clearly observe that Ziouva's model, which is less conformant to the IEEE 802.11 standard than our model, causes a high overestimation of packet delay due to the adoption of the arbitrary additional transition state and the absence of packet retry limits. Thus, the results derived from the model in [148] and subsequent work [141], which is based on [148], show that the utilized DCF operation model overestimates delay performance and leads to ambiguous conclusions for the performance of IEEE 802.11 protocol.

Figure 3.12 validates our analysis for the other two considered performance metrics, packet drop time and packet drop probability, since analysis (lines) coincides with simulation results (symbols). Moreover, figures 3.11 and 3.12 show that the RTS/CTS reservation scheme achieves higher throughput, lower packet delay as well as lower packet drop time comparing to basic access, for the specific large packet size, as a result of shorter collision duration. Moreover, an interesting observation is that packet drop probability is the same for both basic access and RTS/CTS since it is independent of the medium access scheme that is employed.

3.4 Performance evaluation

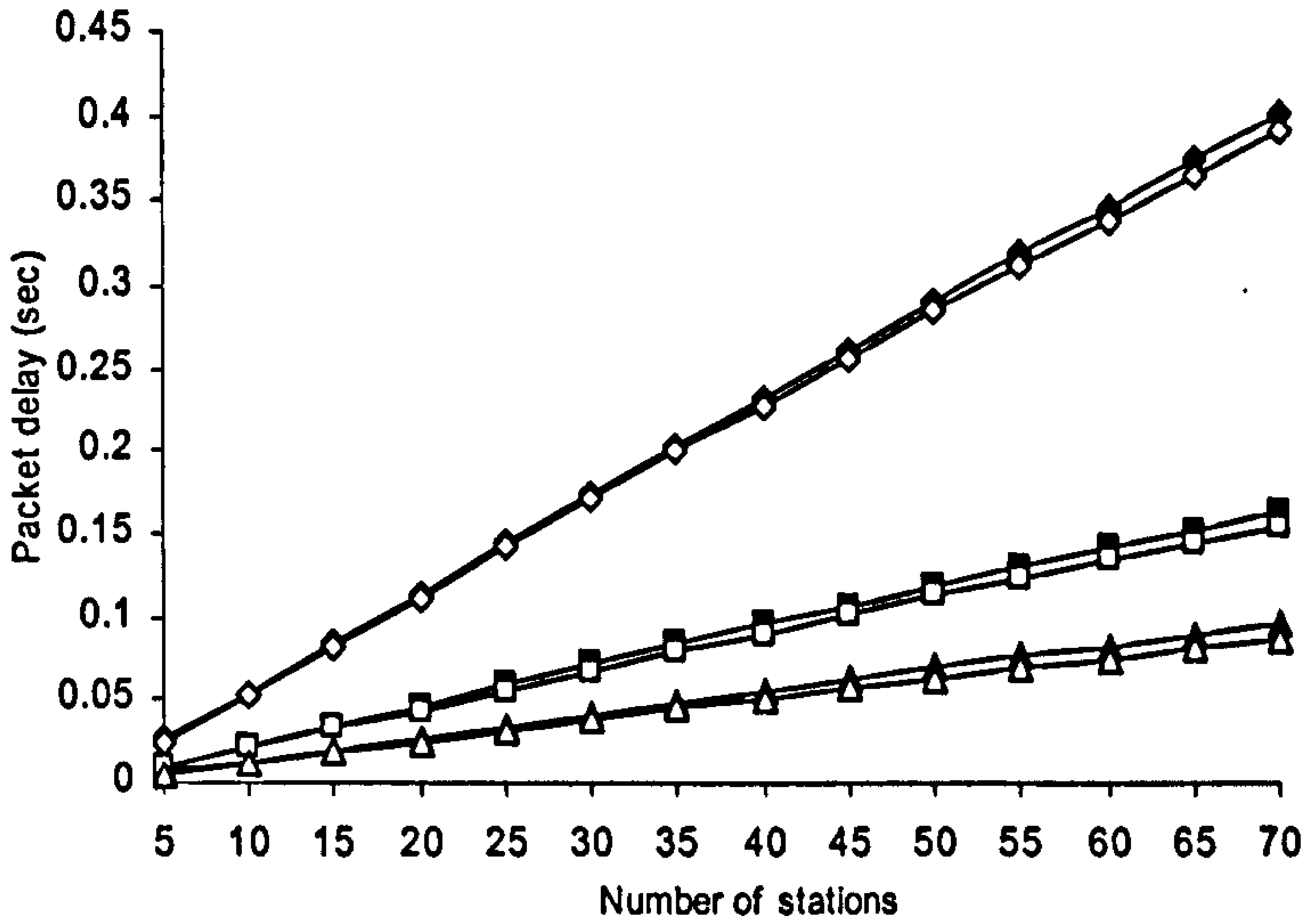
This section presents a performance evaluation of the IEEE 802.11 protocol by employing the analytical model developed in section 3.2. Most of the results presented in the current section are derived for the Direct Spread Sequence Spectrum (DSSS) physical layer utilized in IEEE 802.11b. Only in cases that is necessary, performance results will be also derived for the Orthogonal Frequency Division Multiplexing (OFDM) physical layer utilized in IEEE 802.11a. In all cases, the presented analytical performance results are presented in the following figures have been obtained using the system parameters in table 3.3.

3.4.1 The effect of physical layer and high data rates

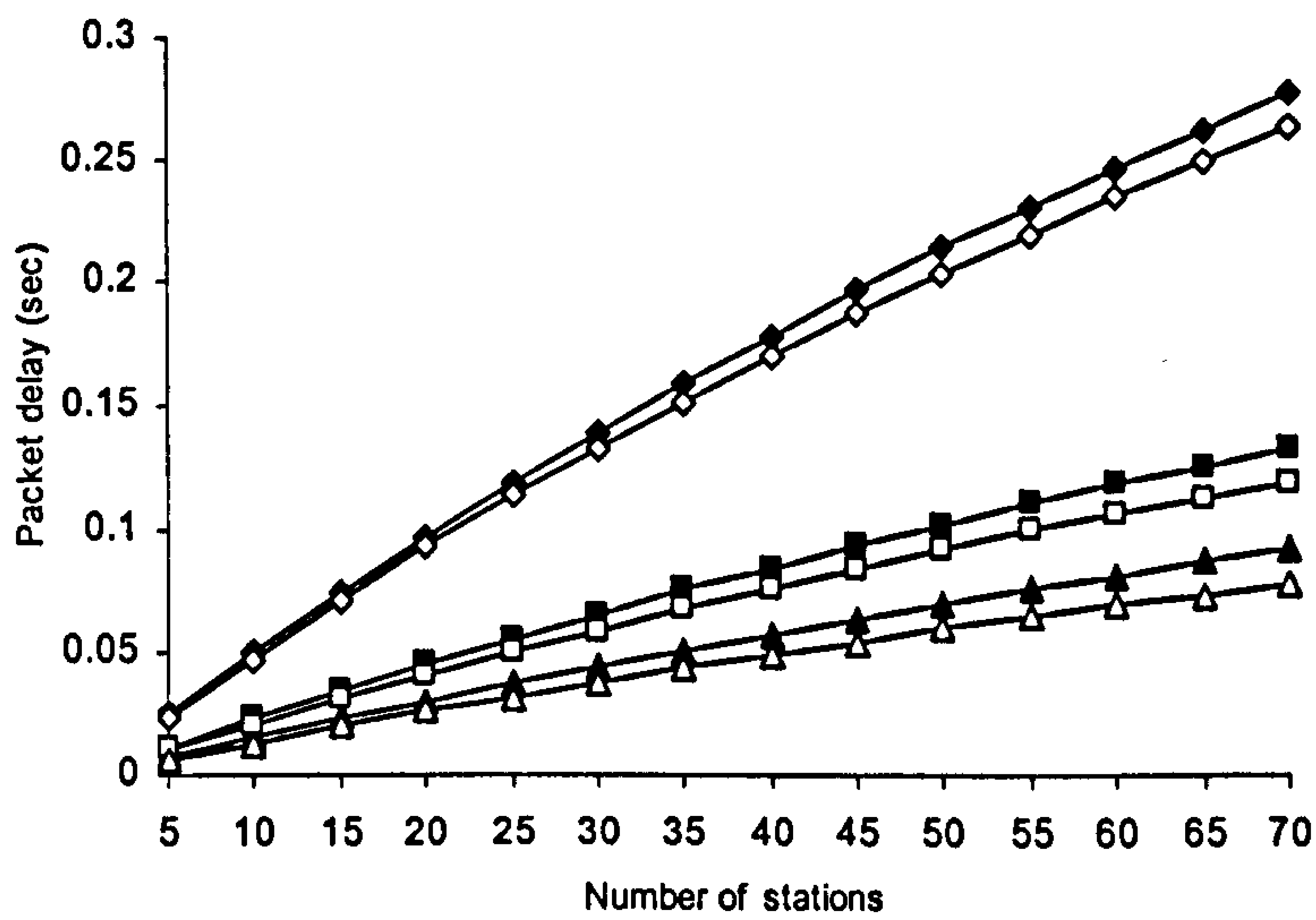
Since the IEEE 802.11 standards specify various physical layers and data rates, it is interesting to study how performance is influenced in each different case. The IEEE 802.11b protocol supports data rates of 1, 2, 5.5 and 11 Mbit/s. The standard defines two different formats for the preamble and header (PHY_{hdr}): the mandatory supported Long PLCP PHY_{hdr} which interoperates with the 1 Mbit/s and 2 Mbit/s data rates and an optional Short PLCP PHY_{hdr}. The Short PLCP PHY_{hdr} allows performance at the high rates (2, 5.5 and 11 Mbit/s) to be significantly increased. In fact, the Short PLCP PHY_{hdr} is intended for applications where maximum performance is desired and interoperability with legacy is not a consideration. The format of both the Long and Short PLCP PHY_{hdr} of a data packet are shown in figure 3.2.

Figure 3.13 plots packet delay versus network size for three data rates (C = 2, 5.5 and 11 Mbit/s) as well as for a short and long PHY packet overhead in IEEE 802.11b. The results show that packet delay is highly dependent on the data rate. When the data rate increases, packet delay values significantly drop off since packet transmission time is considerably reduced. Moreover, the use of a short PHY header, which results in a lower transmission time comparing to the long PHY header's transmission time, considerably decreases packet delay.

Figure 3.14 illustrates the effect of data rate on throughput efficiency for both basic access and RTS/CTS access schemes. When data rate increases, throughput efficiency decreases. The situation is explained considering that the time spent on packet transmission is reduced but the duration of DIFS, SIFS and the slot time is independent of medium data rate and remains the same. Thus, the time spent on DIFS, SIFS and backoff delay increases in relation to packet transmission time, resulting in throughput efficiency degradation.



(a) *Basic access*

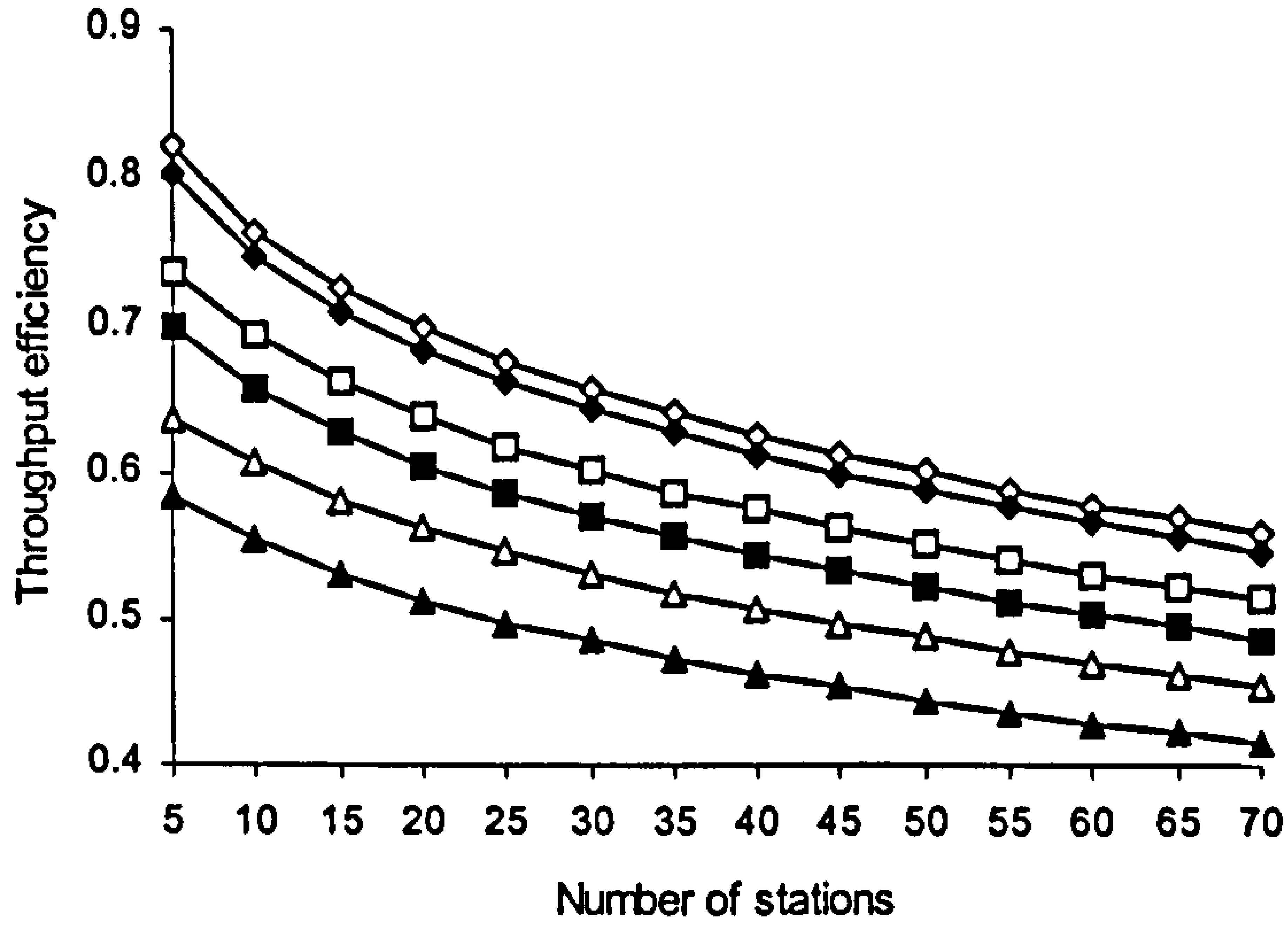


(b) *RTS/CTS*

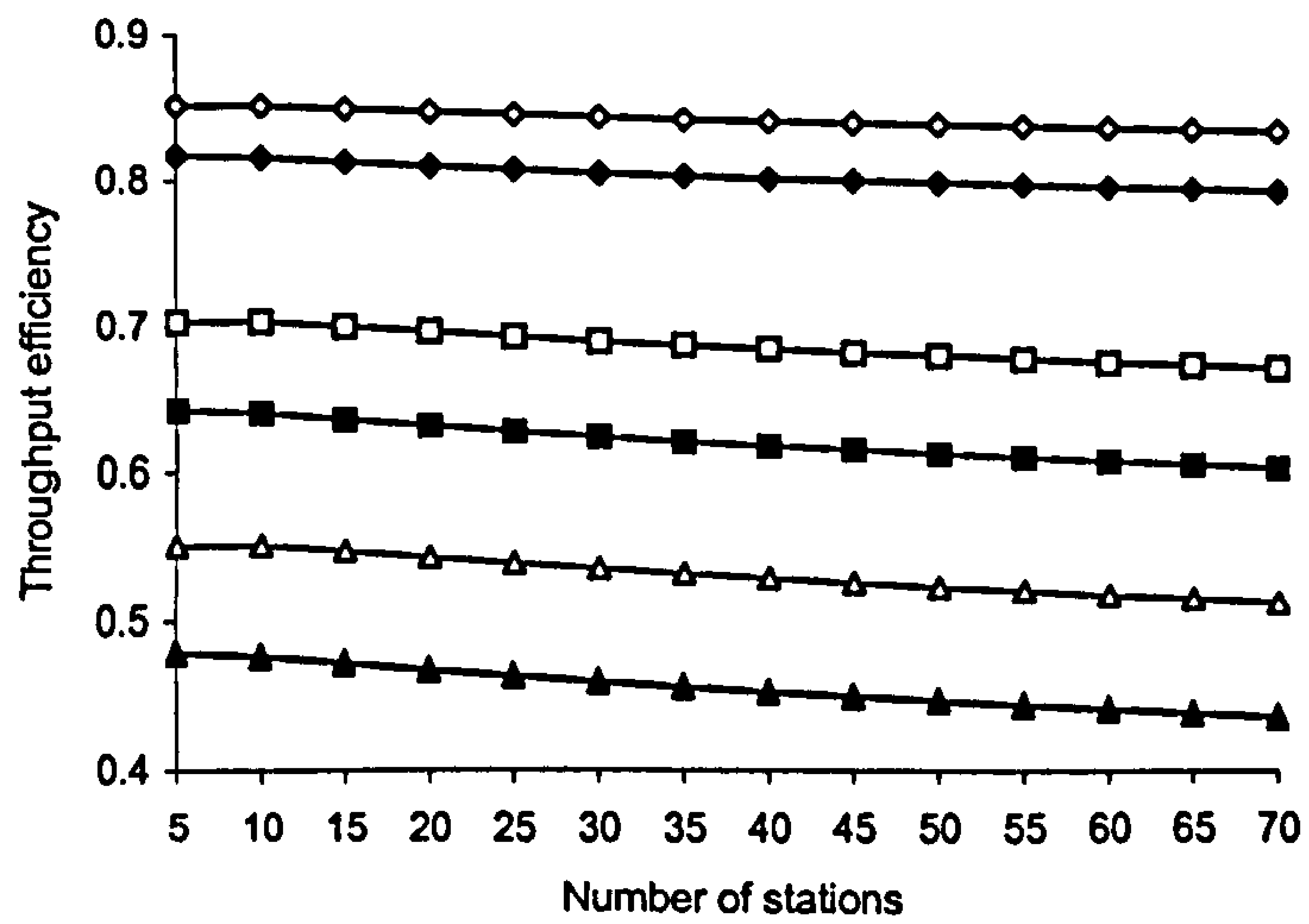
◆ $C = C_{con}$ = 2 Mbit/s, long
◇ $C = C_{con}$ = 2 Mbit/s, short
■ C=5.5 Mbit/s, C_{con} = 2 Mbit/s, long
□ C=5.5 Mbit/s, C_{con} = 2 Mbit/s, short
▲ C=11 Mbit/s, C_{con} = 2 Mbit/s, long
△ C=11 Mbit/s, C_{con} = 2 Mbit/s, short

Figure 3.13 Packet delay versus n, for W=32, m=6, m'=5 and various (C, C_{con}) and headers

An interesting observation in figure 3.14 is that the use of the RTS/CTS appears to be more robust and weakly depends on the number of stations for any data rate due to the shorter collision duration. However, when higher data rates are utilized, (especially C=11 Mbit/s), the basic access scheme seems to appears to achieve a better performance than RTS/CTS even in the case of congested environments (large network sizes). The surprising result is that the RTS/CTS reservation scheme either is beneficial when the number of stations is greater than 50 (C=5.5 Mbit/s) or even degrades performance (C=11 Mbit/s). The reason is that although high data rates reduce the transmission time



(a) *Basic access*



(b) *RTS/CTS*

◆ *C*= C_{con} = *2 Mbit/s, long* ◇ *C*= C_{con} = *2 Mbit/s, short*

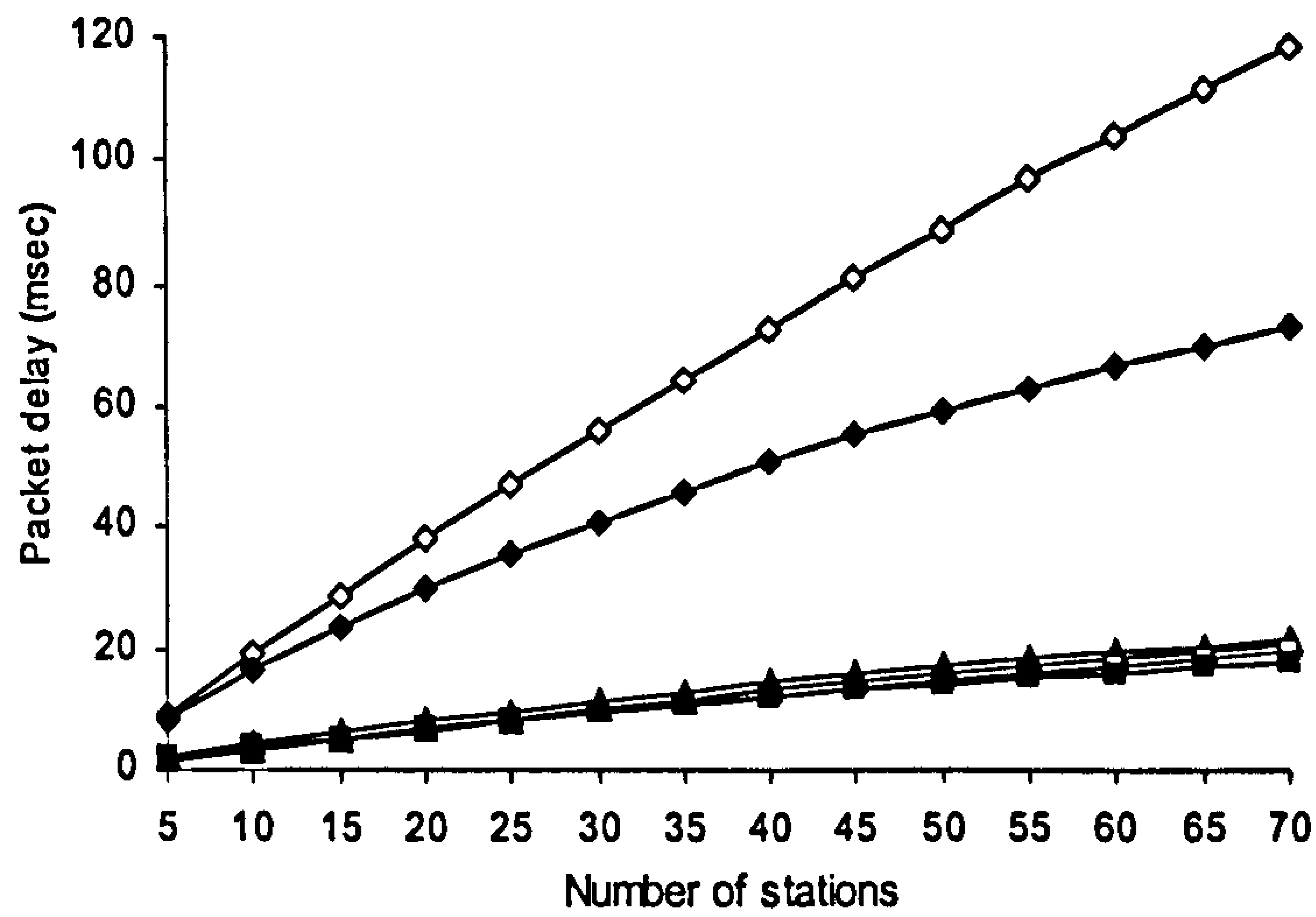
■ *C=5.5 Mbit/s,* C_{con} = *2 Mbit/s, long* □ *C=5.5 Mbit/s,* C_{con} = *2 Mbit/s, short*

▲ *C=11 Mbit/s,* C_{con} = *2 Mbit/s, long* △ *C=11 Mbit/s,* C_{con} = *2 Mbit/s, short*

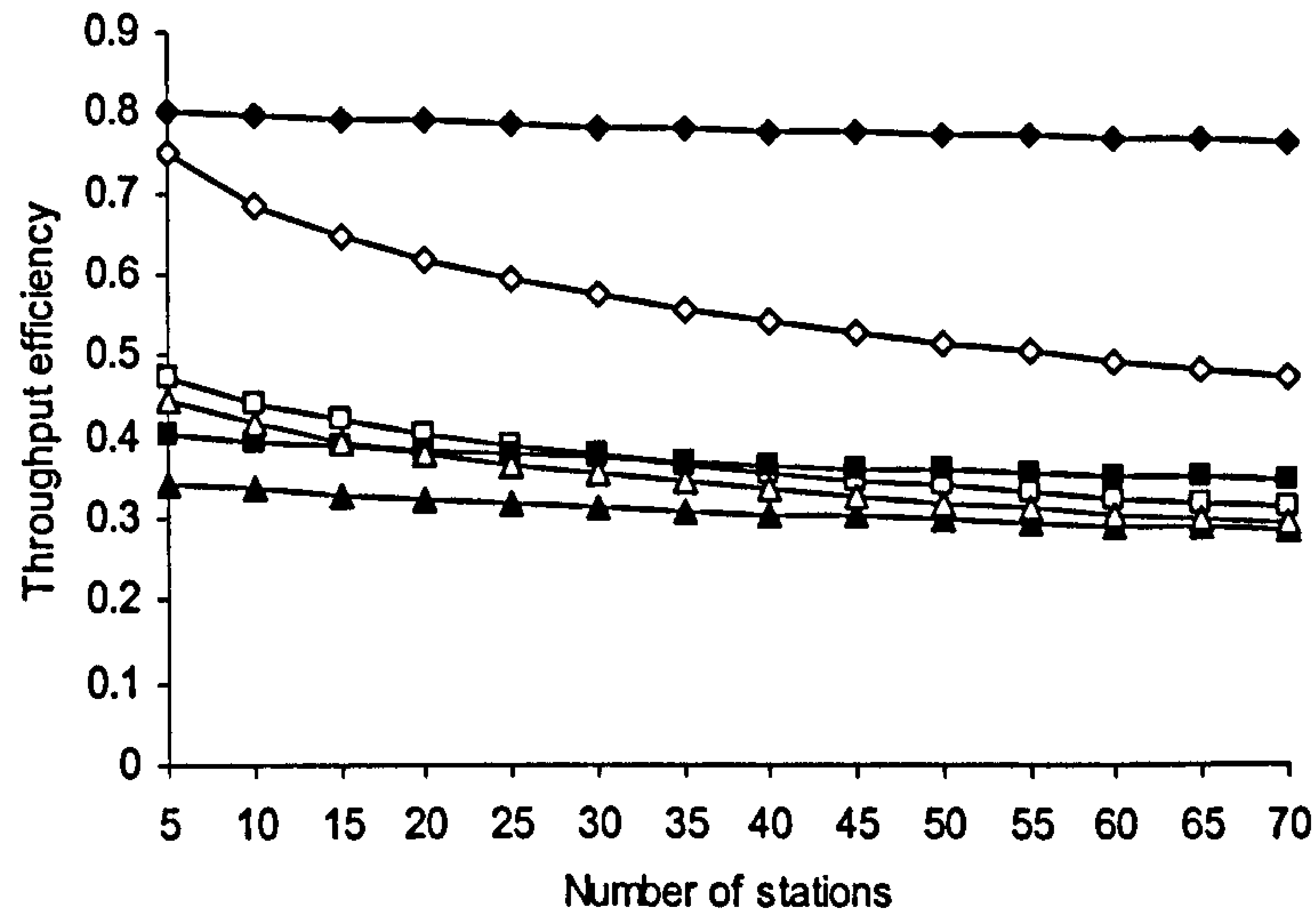
Figure 3.14 Throughput efficiency versus n, for W=32, m=6, m'=5 and various (C, C_{con}*) and headers*

for data packets, the RTS and CTS control packets are still being transmitted at the low control rate (2 Mbit/s), resulting in a considerable communication delay. Furthermore, the Short PLCP PHY_{hdr} (smaller packet overhead) mainly reduces the overhead of RTS and CTS control packets. Thus, the main drawback of the RTS/CTS scheme can be minimized and it could be employed effectively even for smaller network sizes.

Figures 3.15(a) and 3.15(b) plot packet delay and throughput efficiency against the number of contending stations for IEEE 802.11a physical layer and for three different pairs of data and control rates in IEEE 802.11a. When the link data and control rates are



(a) *Packet delay*

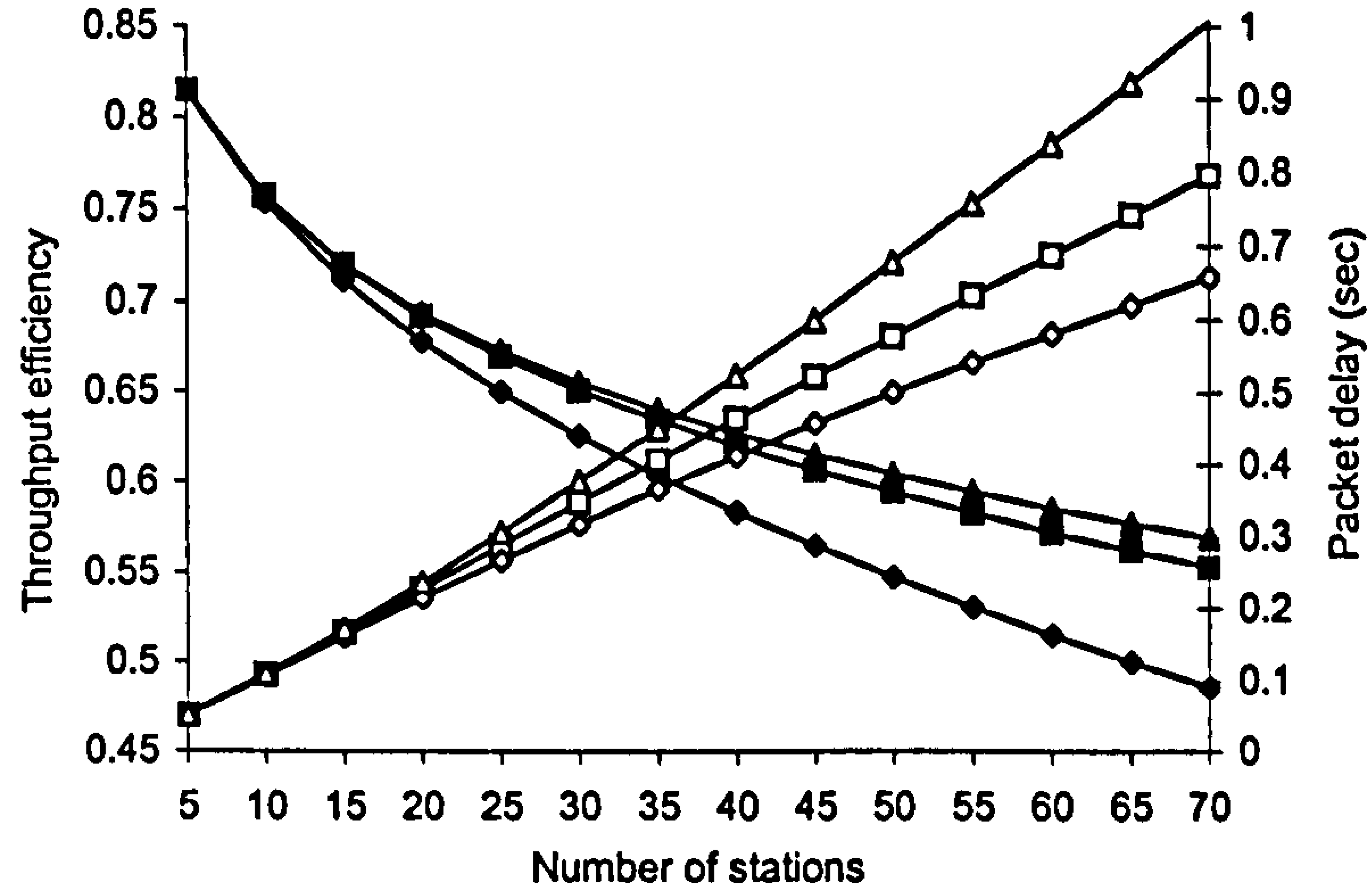


(b) *Throughput efficiency*

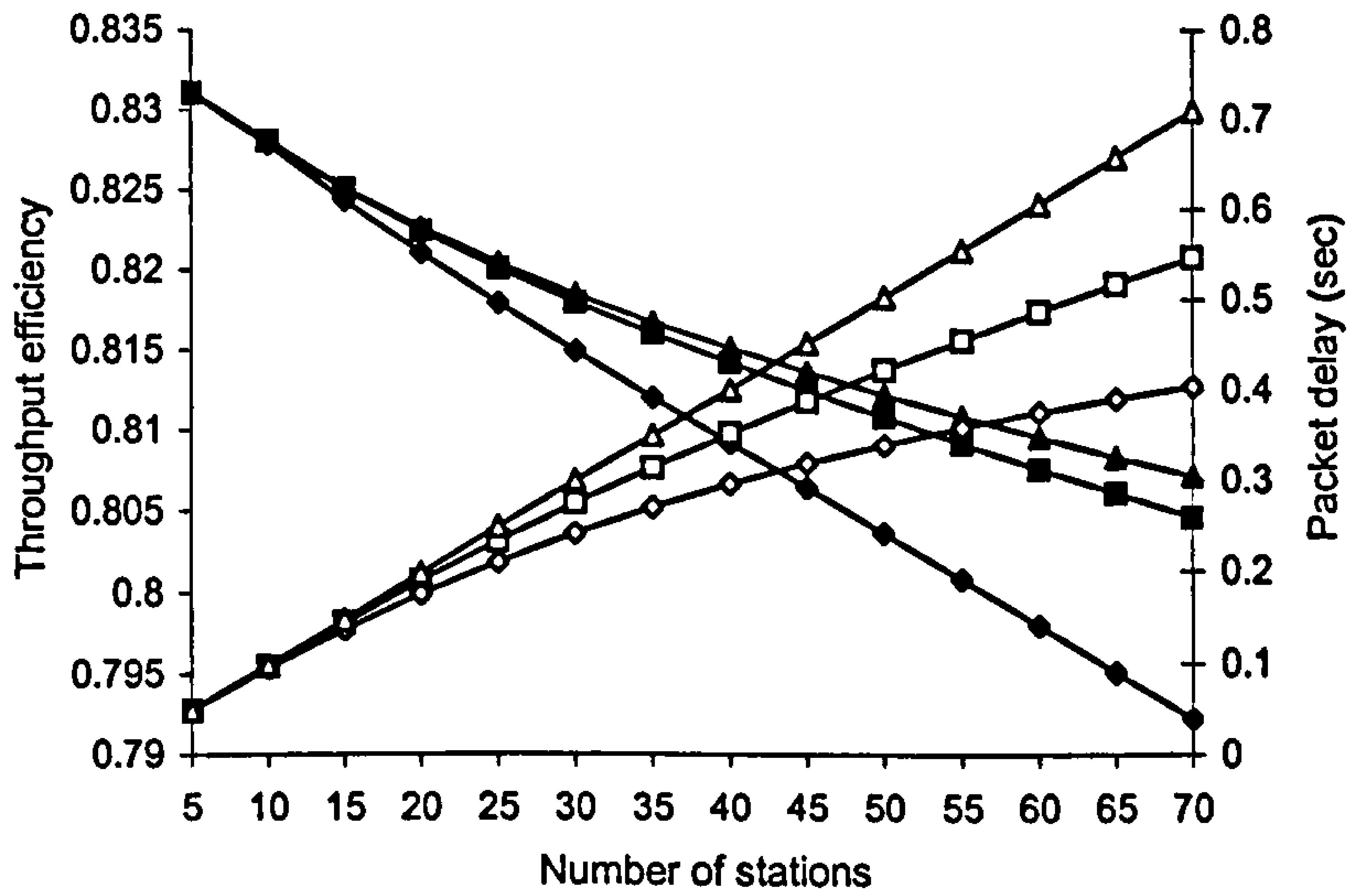
◆ *RTS/CTS, C= C_{con} = 6 Mbit/s*	◇ *Basic access, C= C_{con} = 6 Mbit/s*
■ *RTS/CTS, C=54Mbit/s, C_{con} = 24 Mbit/s*	□ *Basic access, C=54Mbit/s, C_{con} = 24 Mbit/s*
▲ *RTS/CTS, C=54Mbit/s, C_{con} = 6 Mbit/s*	△ *Basic access, C=54Mbit/s, C_{con} = 6 Mbit/s*

Figure 3.15 Packet delay and throughput efficiency versus n, for W=16, m=m'=6 and various (C, C_{con})

the same (6 Mbit/s), the RTS/CTS reservation scheme always achieves better performance than the basic access due to the shorter collision duration, which is consistent with the conclusion derived in [6] for a data rate of 1 Mbit/s. On the contrary, when the highest data rate of 54 Mbit/s is utilized combined with the lowest control rate of 6 Mbit/s, the basic access scheme outperforms RTS/CTS for any network size since the much lower control rate considerably degrades performance. Furthermore, for the 54 Mbit/s data rate and in the best-case scenario for the highest possible control rate of 24 Mbit/s, the RTS/CTS scheme attains higher throughput efficiency than the basic access scheme for network sizes n >35.



(a) *Basic access*



(b) *RTS/CTS scheme*

◆ *Throughput efficiency, m=4* ◇ *Packet delay, m=4*

■ *Throughput efficiency, m=6* □ *Packet delay, m=6*

▲ *Throughput efficiency, no retry limits* △ *Packet delay, no retry limits*

Figure 3.16 Throughput efficiency and packet delay for basic access varying retry limit

3.4.2 The effect of packet retry limit (*m*)

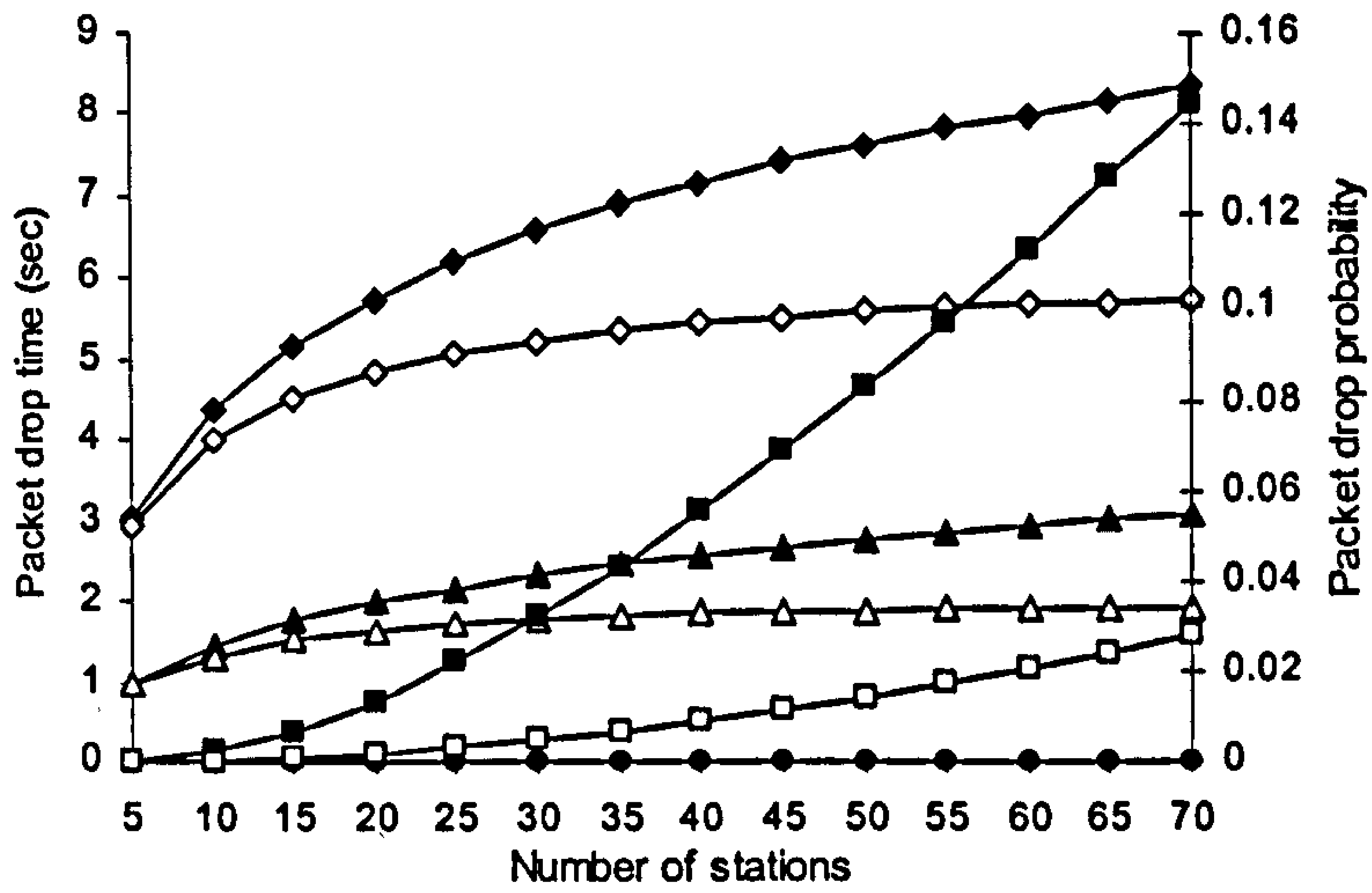
The dependency of the average packet delay the retry limit is examined for both basic access and RTS/CTS mechanisms, respectively, in figures 3.16(a) and 3.16(b). The two figures report throughput efficiency and packet delay values for two different packet retry limits (m=4 and m=6) as well as for the case of no retry limits[2]. Results

[2] The IEEE 802.11 standard proposes the value 6 for the packet retry limit.

show that the retry limit considerably affects the throughput performance of the 802.11 protocol. Both figures illustrate that the average packet delay increases as the retry limit increases and that the packet delay decreases if a smaller retry limit than the proposed value is employed, in both the basic access and RTS/CTS mechanisms. Especially in the case of low retry limit values and for medium or large network size ($n > 15$) the packet delay attains a lower value but at the expense of more packets being dropped. On the other hand, if no retry limits are considered, both throughput and packet delay performance is overestimated due to the fact that all packets are eventually transmitted successfully and are never dropped.

Figure 3.17 depicts the effect of the retry limit on the packet drop probability and the packet drop time. As shown in equation (3.30), packet drop probability depends on the retry limit and the collision probability. Since the packet drop probability does not depend on access mechanism, the results presented in figure 3.17 are applicable to both basic access and RTS/CTS. More specifically, packet drop probability increases as the number of stations increases. For small values of the retry limit and a large network size, the packet drop probability increases rapidly (packet drop probability of 0.14 is obtained for $m = 4$ and n=70). Figure 3.17 also allows us to answer the question on the dependence of the packet drop time on the retry limit. In particular, the RTS/CTS mechanism always achieves a lower value for the average drop time, with respect to the basic access mechanism, mainly observable when large network size values lead to a higher collision probability. A small value of the retry limit ($m = 4$), results in a low average drop time. Obviously, for the no retry limit case, both packet drop probability and packet drop time are equal to zero since there are not any dropped packets.

Figure 3.18 illustrates the equivalent performance results for packet inter arrival time utilizing the previous different retry limit values for basic access and RTS/CTS schemes. When basic access and a low retry limit (m=4) are employed, packet inter arrival time attains the highest value compared to the other two cases. On the other hand, when the RTS/CTS reservation scheme is utilized, packet inter arrival time is not practically affected for any m value. Finally, for the no retry limit case, packet inter arrival time obtains exactly the same values as packet delay and since there are not any discarded packets.



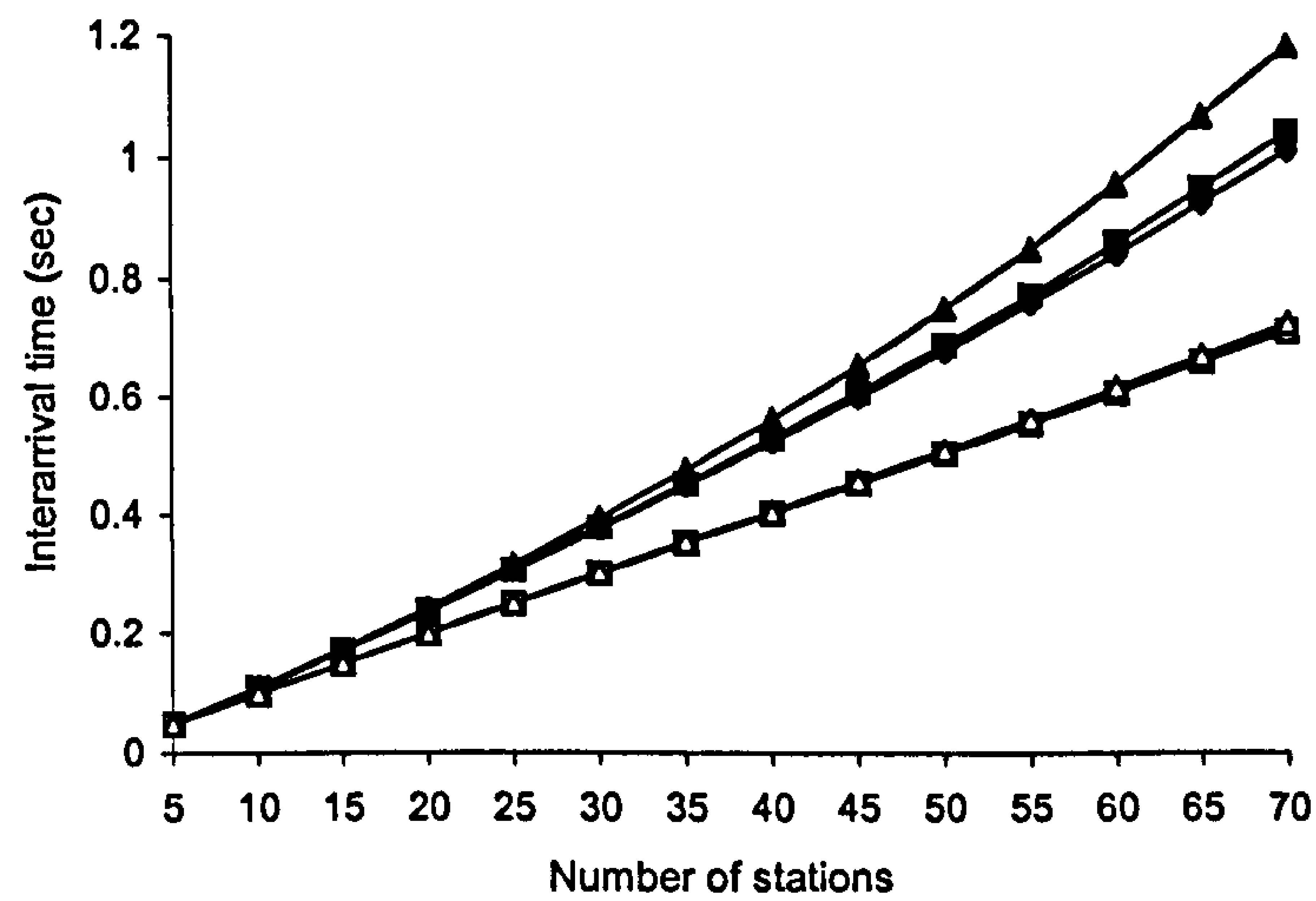
◆ *Packet drop time, basic access, m=6* ◇ *Packet drop time, RTS/CTS, m=6*

▲ *Packet drop time, basic access, m=4* △ *Packet drop time, RTS/CTS, m=4*

■ *Packet drop prob., both access schemes, m=4* □ *Packet drop prob., both access schemes, m=6*

● *Packet drop time and packet drop prob., both access schemes, no retry limits*

Figure 3.17 Packet drop time and packet drop probability varying retry limit



▲ *Inter arrival time, basic access, m=4* △ *Inter arrival time, RTS/CTS, m=4*

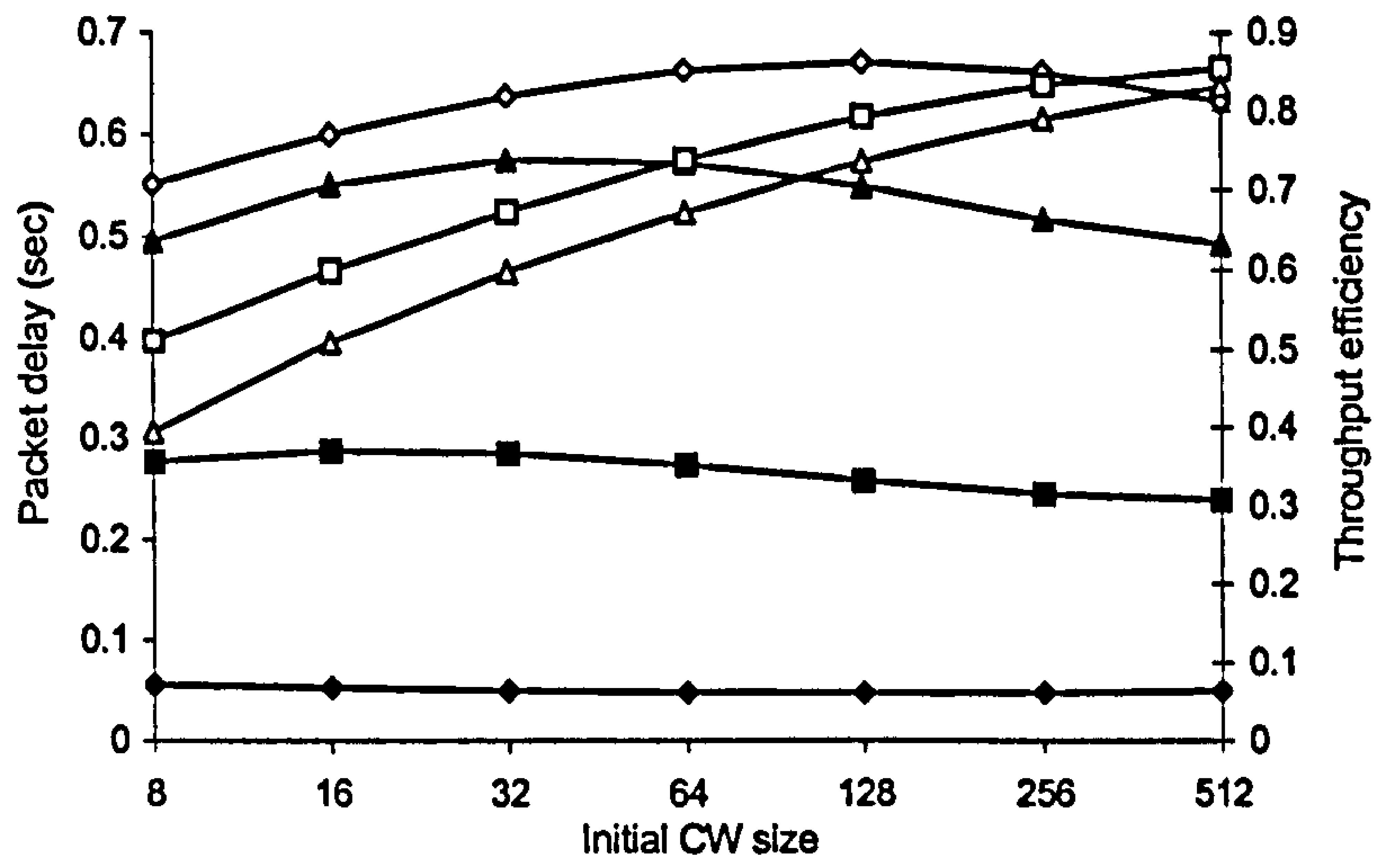
■ *Inter arrival time, basic access, m=6* □ *Inter arrival time, RTS/CTS, m=6*

◆ *Inter arrival time, basic access, no retry limits* ◇ *Inter arrival time, RTS/CTS, no retry limits*

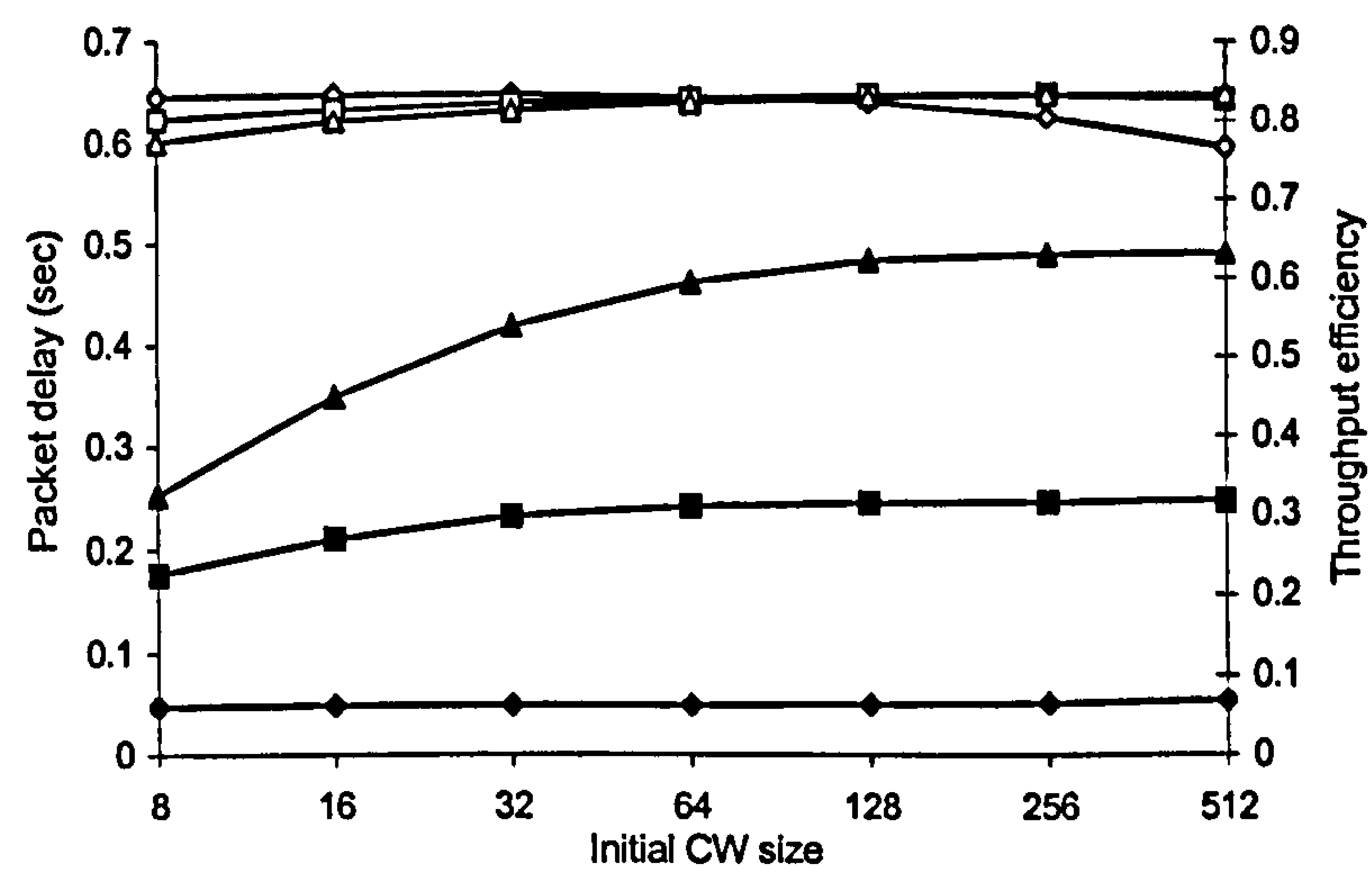
Figure 3.18 Packet inter arrival time for basic access and RTS/CTS, varying retry limit

3.4.3 The effect of Contention Window (CW)

The following figures examine the dependency of packet delay, throughput efficiency, packet drop probability, packet drop time and packet inter arrival time on the initial contention window size *W*. The figures study both the basic access and the RTS/CTS mechanisms and report three different network sizes (n=5, 25 and 50).



(a) *Basic access*



(b) *RTS/CTS scheme*

◆ *Packet delay, n =5*	◇ *Throughput efficiency, n =5*
■ *Packet delay, n =25*	□ *Throughput efficiency, n =25*
▲ *Packet delay, n =50*	△ *Throughput efficiency, n =50*

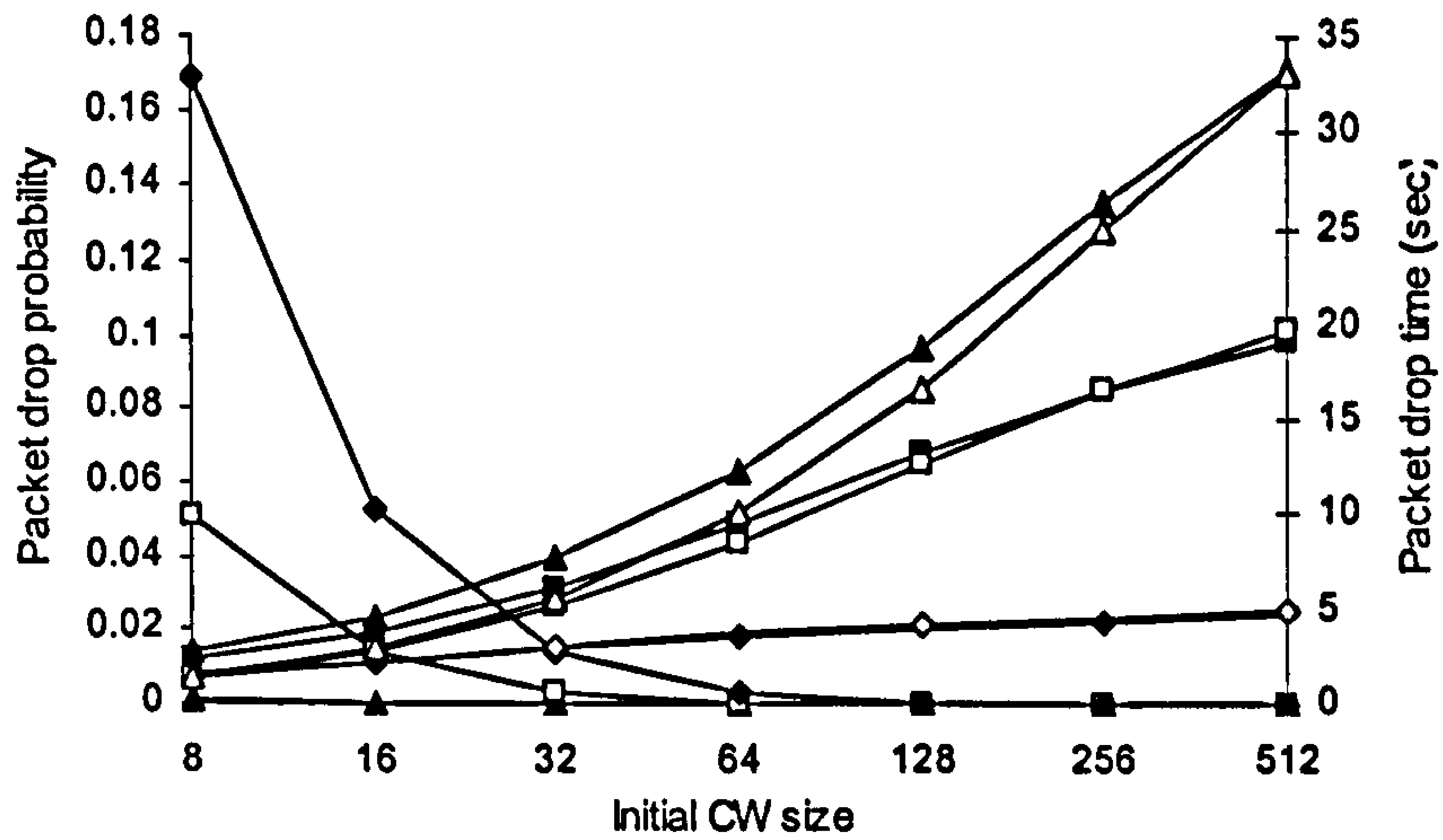
Figure 3.19 Packet delay and throughput efficiency for basic access and RTS/CTS schemes

Figure 3.19(a) plots packet delay and throughput efficiency versus initial contention window (*CW*) size for the basic access scheme and for various network sizes. The figure shows that when the basic access mechanism is employed, throughput improves as initial contention window increases. The situation is explained since when *CW* increases, the number of collisions decreases and the system throughput gets higher. The only exception is when n=5 and $CW \geq 128$, throughput drops off due to the increased number of idle slots. Furthermore, packet delay is not greatly affected from the increase of the initial contention window, in small network sizes. In large network scenarios, when initial contention size grows, more packets are transmitted successfully (figure 3.20(a)). A notable result is that packet delay increases with the increase of initial contention size, especially when $CW \leq 32$, as a result of the fact that the additional packets contain large delays. In the case of $CW \geq 64$, packet delay drops off as a result of fewer collisions that take place. The figure also indicates that a very small initial contention window is not effective for large networks due to the increased number of collisions. In contrast, a large value of W is unsuitable for a small network size ($n \leq 5$) due to many idle slots.

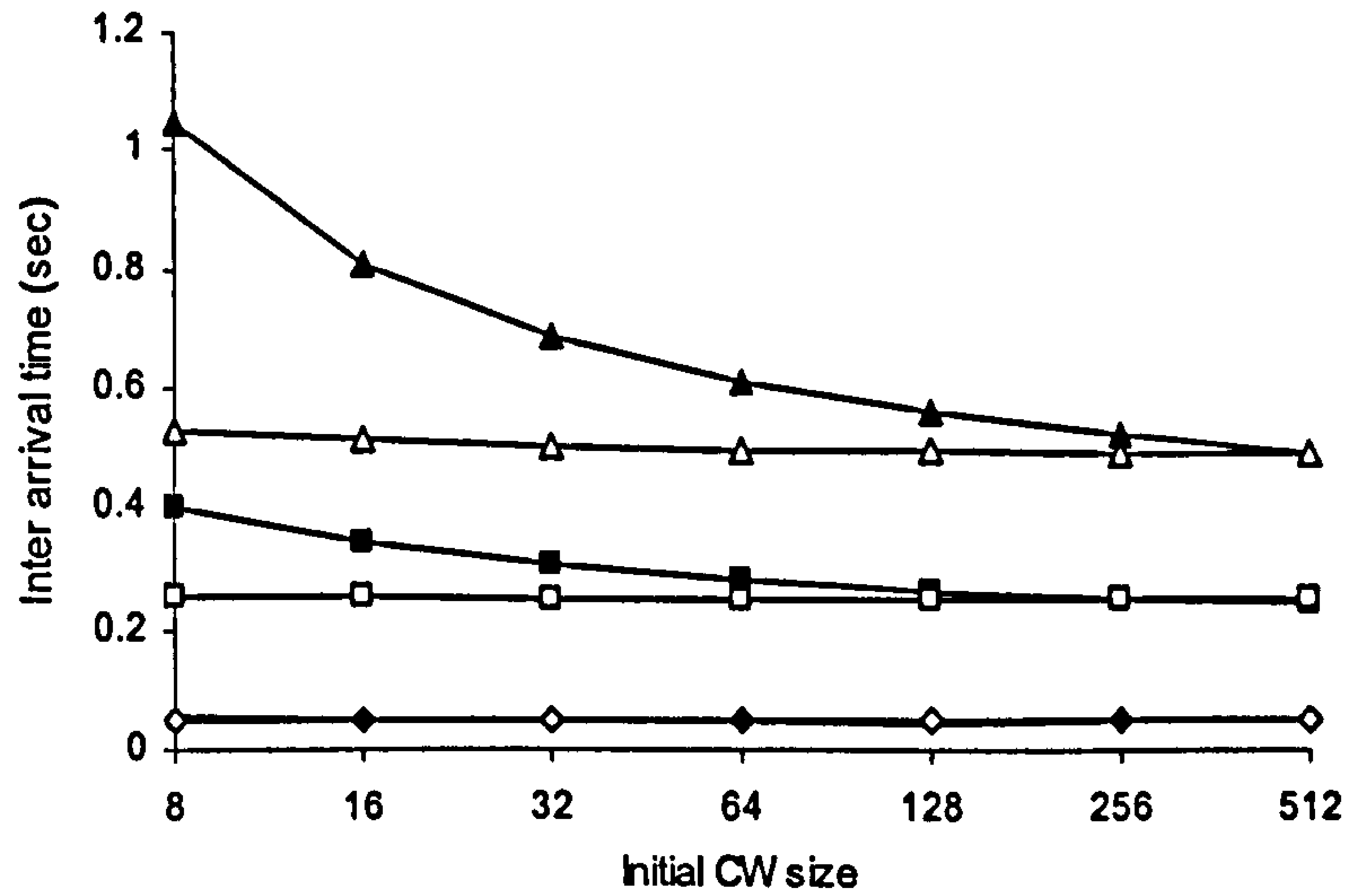
Figure 3.19(b) explores the effect of CW when the RTS/CTS is employed and indicates that the choice of initial contention window does not significantly affect throughput due to the shorter collision duration of the RTS packets. When n=5 and $CW \geq 128$ throughput slightly decreases due to the increase of idle slots. For a large network size, the throughput improves for high values of CW but remains constant as long as CW is greater than 64 due to the fact that high CW values can effectively cope with the increased number of collisions. In contrast, for large network sizes, packet delay increases when CW increases due to the fact that more packets are being successfully transmitted (illustrated in figure 3.20(a) with the aid of packet drop probability). However, the choice of the initial contention window size does not affect packet delay when $CW \geq 128$ with a large network size. For a small number of contending stations (n=5), packet delay is not affected by changing the values of initial contention window size since the number of packets that are transmitted successfully is about the same regardless the value of CW (figure 3.20(a)).

Figure 3.20(a) shows that the adjustment of the initial CW size to higher values in large network scenarios highly benefits packet drop probability; fewer packets are discarded since higher values of CW reduce the number of collisions. On the other hand, for a small number of stations (n=5), the packet drop probability is not considerably affected as a result of the low collision probability. Figure 3.20(b) illustrates that higher values of CW cause an increase on packet drop time for both the basic access and RTS/CTS mechanisms mainly due to the increase of idle slots.

Figure 3.20(b) plots packet inter arrival time against the initial CW size for three different network sizes (n=5, 25 and 50). Small CW values result in a high inter arrival time when basic access scheme is employed; on the contrary, when RTS/CTS is utilized and for medium and large network scenarios (n=25, 50), packet inter arrival time is only marginally affected by varying the CW values.



(a) *Packet drop probability and packet drop time*



(b) *Packet inter-arrival time*

◆ *Basic, n =5* ◇ *RTS/CTS, n =5*

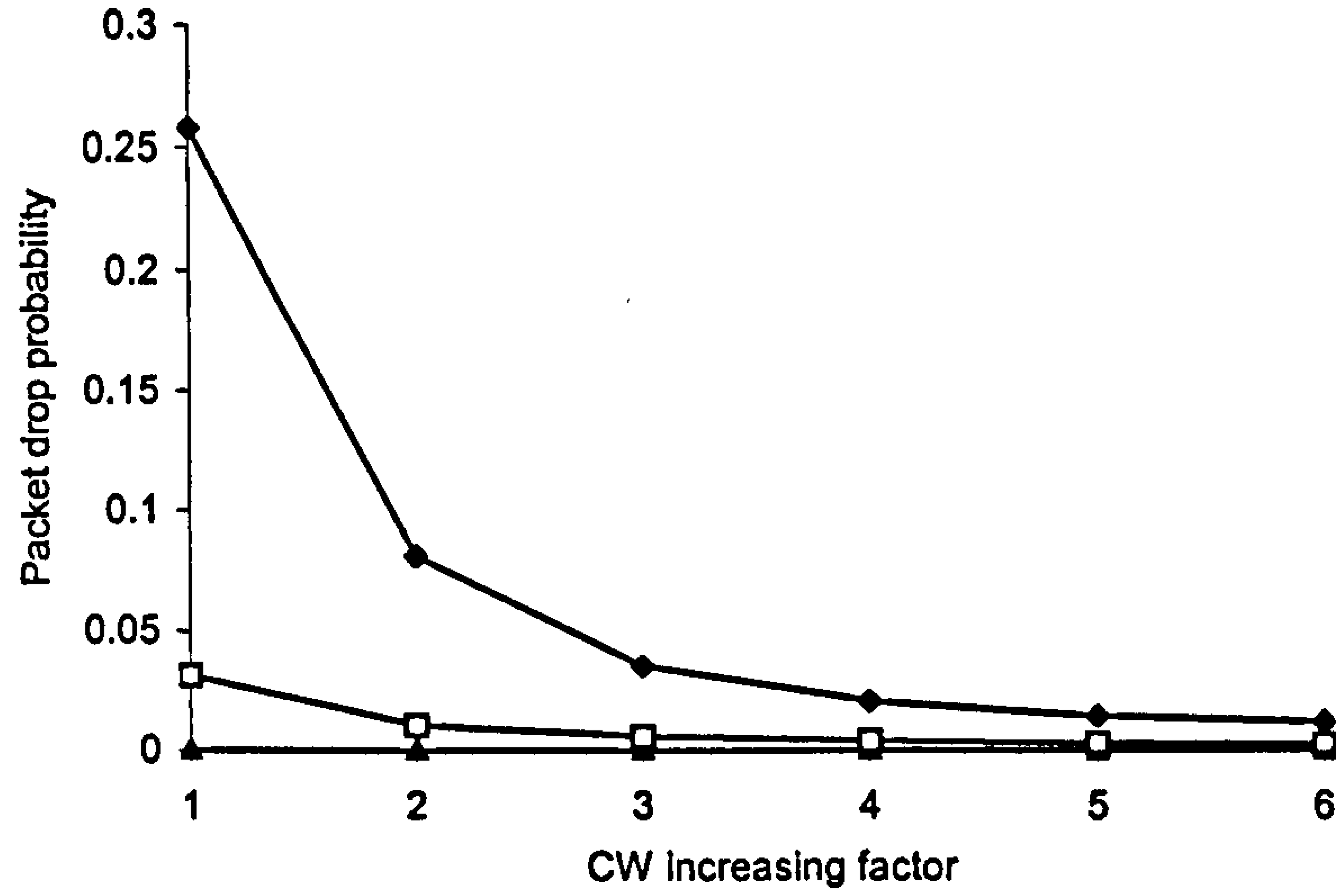
■ *Basic, n =25* □ *RTS/CTS, n =25*

▲ *Basic, n =50* △ *RTS/CTS, n =50*

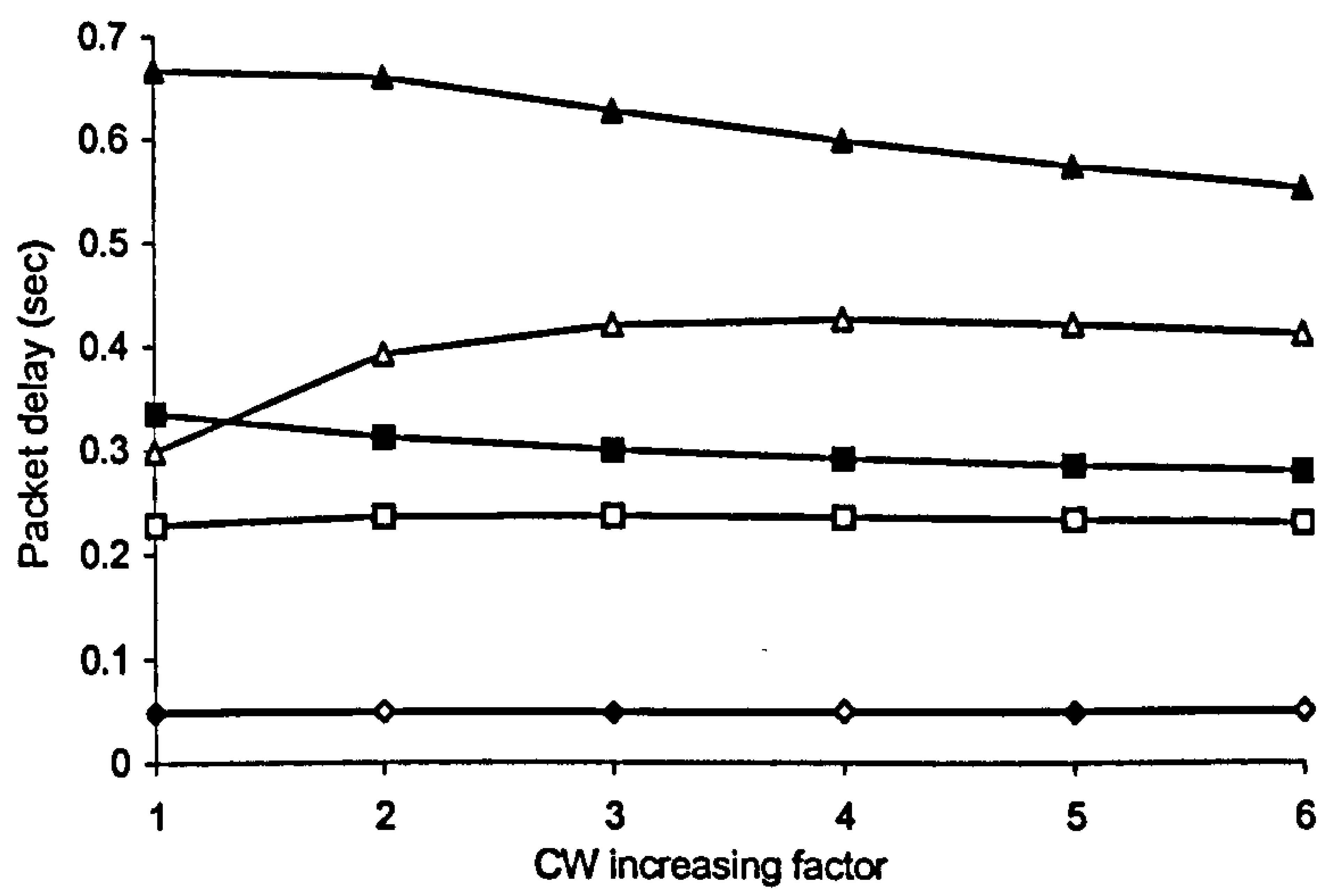
Figure 3.20 Packet drop probability, drop time and interarrival time for basic access and RTS/CTS

3.4.4 The effect of maximum *CW* size (*m'*)

Figures 3.21, 3.22 and 3.23 study the effect of the maximum *CW* size (by varying the *CW* increasing factor *m'*) on packet drop probability, packet drop time, packet delay and throughput efficiency for the basic access and the RTS/CTS mechanisms and for three different network size. Figure 3.21(a) plots packet drop probability against the *CW* increasing factor for various network sizes. The figure illustrates that the increase of *m'* is beneficial for packet drop probability; fewer packets are dropped since higher values of *m'* deal with the increased number of packet collision.



(a) *Packet drop probability*

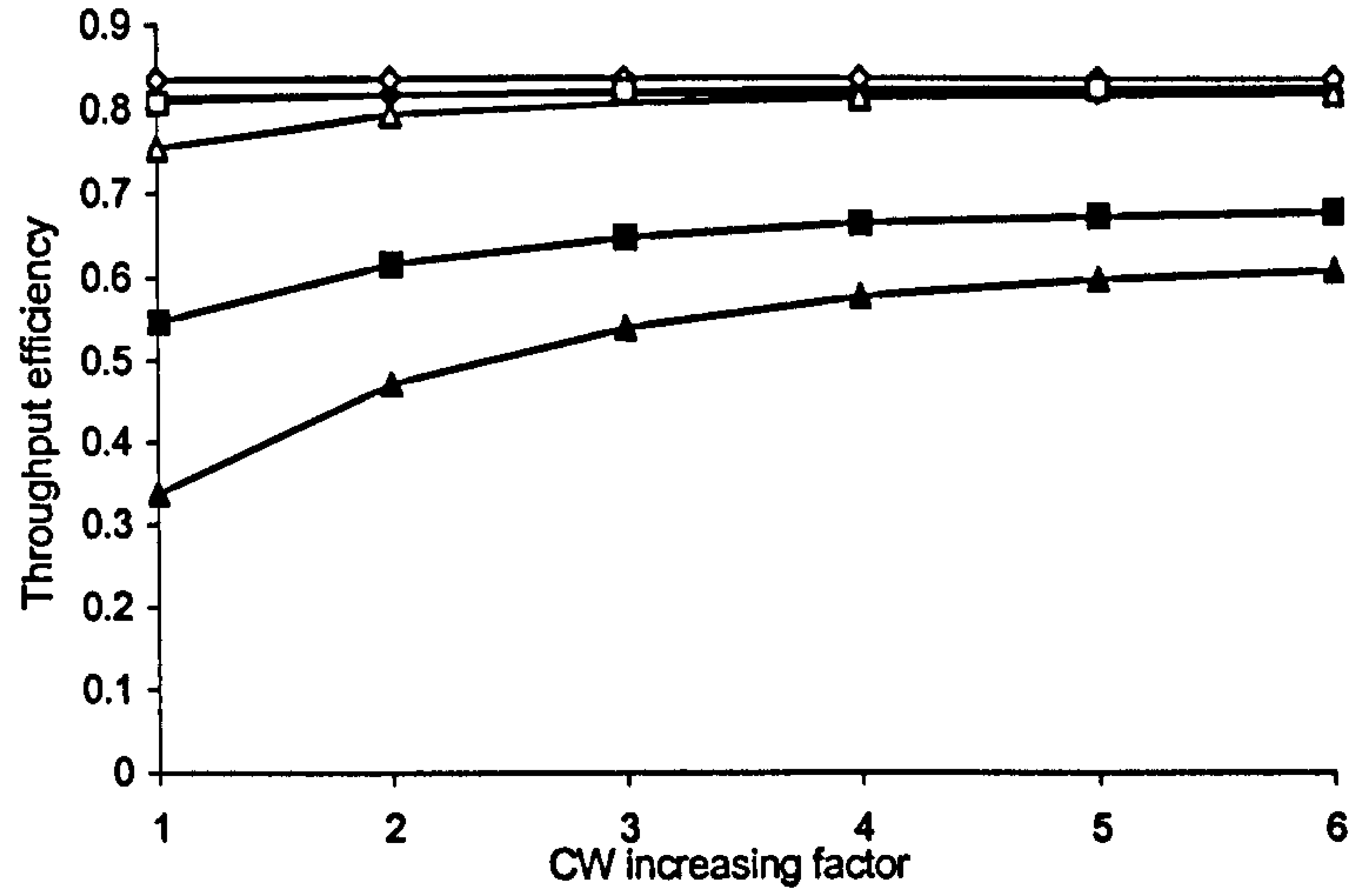


(b) *Packet delay*

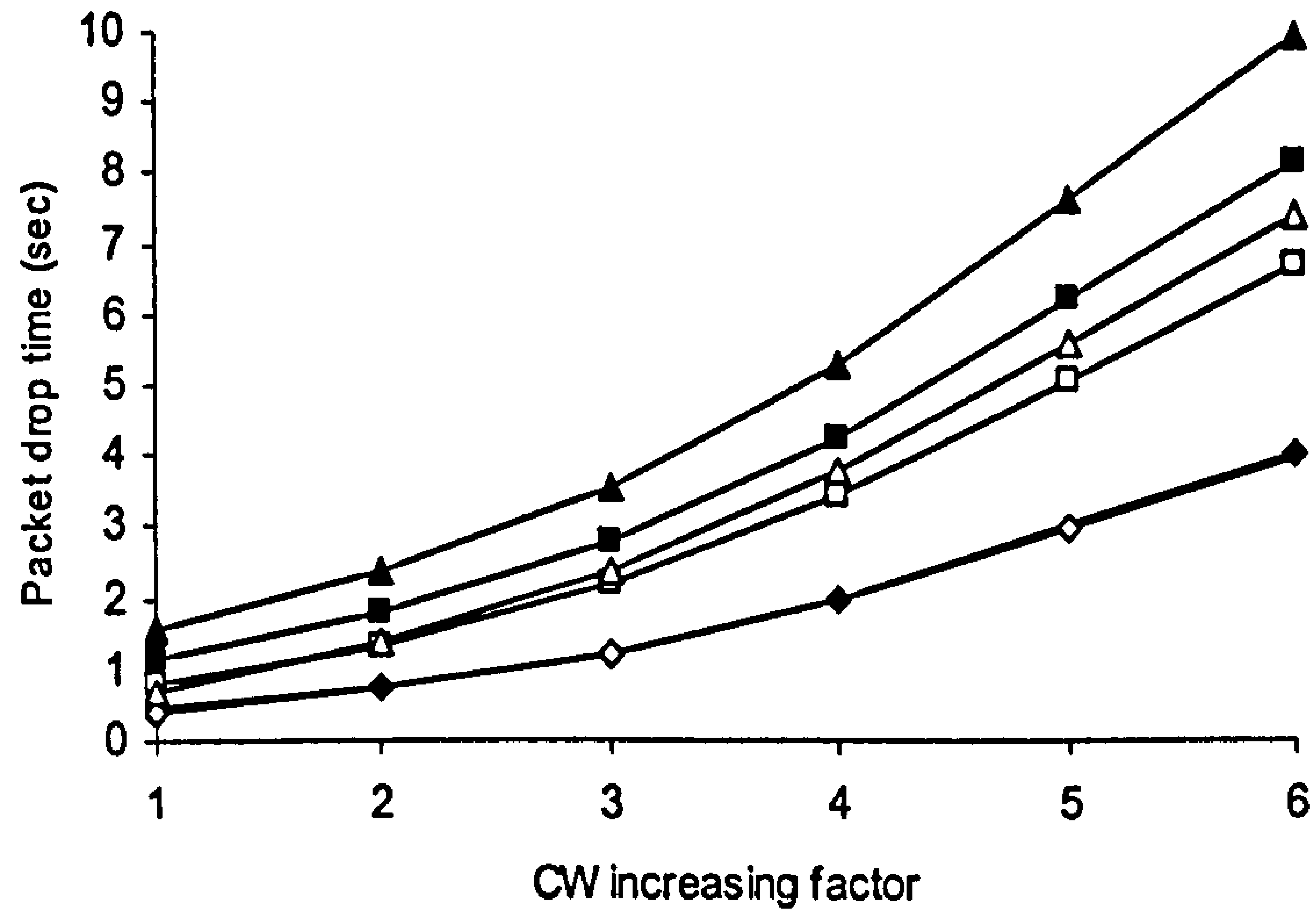
◆ *Basic, n =5* ◇ *RTS/CTS, n =5*
■ *Basic, n =25* □ *RTS/CTS, n =25*
▲ *Basic, n =50* △ *RTS/CTS, n =50*

Figure 3.21 Packet drop probability and packet delay for basic access and RTS/CTS

In figure 3.21(b), packet delay is plotted against the *CW* increasing factor for both basic access and RTS/CTS mechanisms. The figure depicts that packet delay mainly depends on the number of contending stations and increases when n increases. Moreover, the use of RTS/CTS scheme appears to be beneficial as it offers lower packet delay especially for large networks, while basic access experiences higher packet delay. In all cases, packet delay is not significantly affected when $m' \geq 5$ for any network size and access scheme. The reason is that when $m' \geq 5$, less collisions are taking place and many packets are eventually transmitted successfully as it is shown in figure 3.21(a).



(a) *Throughput efficiency*



(b) *Packet drop time*

◆ *Basic, n =5*	◇ *RTS/CTS, n =5*
■ *Basic, n =25*	□ *RTS/CTS, n =25*
▲ *Basic, n =50*	△ *RTS/CTS, n =50*

Figure 3.22 Throughput efficiency and packet drop time for basic access and RTS/CTS

Figure 3.22(a) studies the effect of the *CW* increasing factor on throughput efficiency for three different networks sizes, for both basic access and RTS/CTS mechanisms. The throughput performance of basic access scheme increases as the *CW* increasing factor increases, whereas RTS/CTS mechanism appears more robust and constantly achieves high throughput values. Moreover, the *CW* increasing factor does not affect throughput efficiency when $m' \geq 5$ and $m' \geq 4$ for the basic access and RTS/CTS schemes respectively.

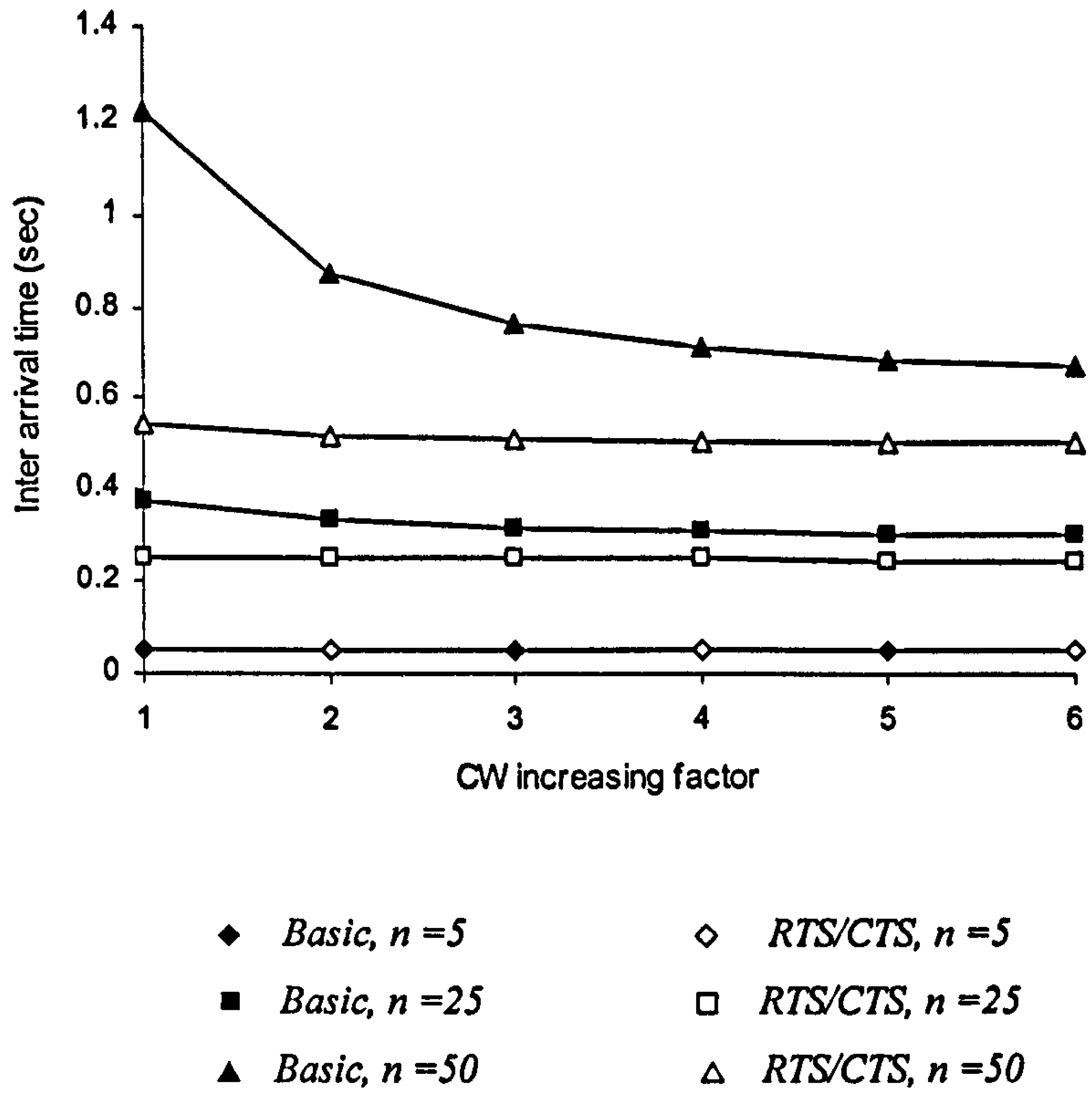


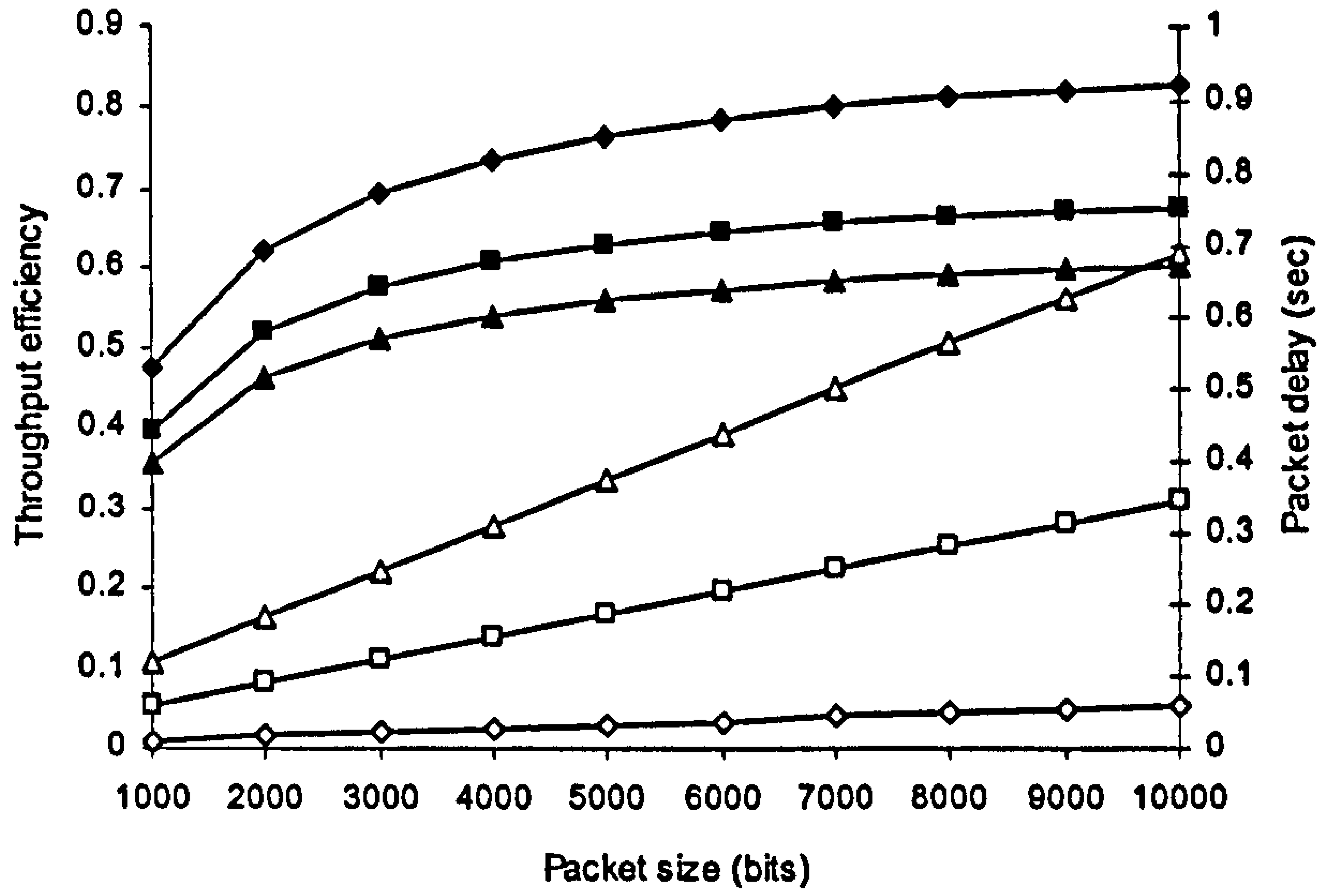
Figure 3.23 Packet inter arrival time for basic access and RTS/CTS

Figure 3.22(b) reveals that packet drop time depends significantly on m' values. Moreover, it is obvious that the medium access mechanism has a significant effect on packet drop time; RTS/CTS scheme achieves a considerably lower packet drop time compared to the basic access scheme.

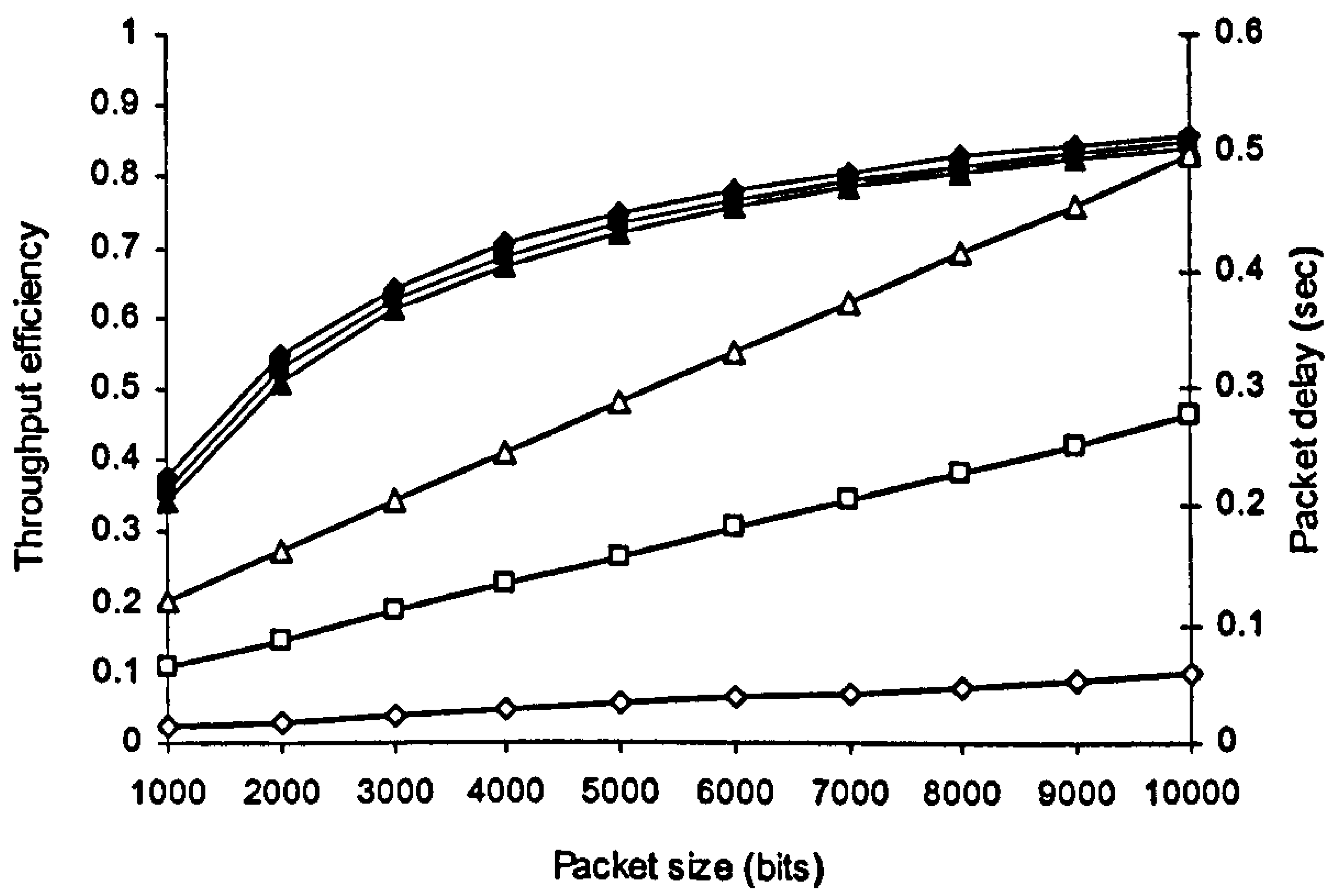
Similar conclusions are derived for the packet inter-arrival time in figure 3.23. For medium (n=25) or large (n=50) network scenarios, basic access always attains higher packet inter arrival time values compared to RTS/CTS scheme. On the contrary, for small network sizes (n=5), packet inter arrival time is almost the same, for both medium access schemes.

3.4.5 The effect of packet payload size (l)

The effect of packet payload size (l) on performance is illustrated in figure 3.24. The figure plots throughput and packet delay against packet size for three representative network sizes (n=5, 25 and 50) and for both access mechanisms. Figure 3.24(a) shows that when the basic access scheme is employed both network size and packet size significantly affect performance. On the other hand, it seems that the RTS/CTS reservation scheme is almost independent of network size since the negative impact of packet collisions is considerably reduced by the shorter collision duration comparing to the basic access scheme. We also observe that the smaller the packet size is, the lower



(a) *Basic access*



(b) *RTS/CTS*

◆ *Throughput efficiency, n=5*	◇ *Packet delay, n=5*
■ *Throughput efficiency, n=25*	□ *Packet delay, n=25*
▲ *Throughput efficiency, n=50*	△ *Packet delay, n=50*

Figure 3.24 Throughput efficiency and packet delay for basic access and RTS/CTS versus packet size

the packet delay is. This is a strong indication that we are dealing with a trade off on delay/throughput performance especially as the network size increases.

3.4.6 Refinement of protocol parameters

There are a variety of performance requirements according to the various communication needs or application desires. For example, time bounded applications that exchange query-like messages, require low packet loss and low delivery delay. Conversely, applications that provide delay insensitive services (i.e. email, ftp) are not concerned much with packet timely deliverance and maximising throughput performance is of prime importance in this case. Additionally, there are many applications that lie somewhere in the middle and may demand low delivery delay but will not be sensitive to some loss of packets or may demand low loss but not small delay. For example, multimedia applications are not able to tolerate high delay or jitter but may tolerate some packet loss whereas HTTP-like applications can tolerate delay but require minimum data loss.

In order to fulfill specific communication needs, we propose the adjustment of certain protocol parameters to different values than those proposed by the IEEE standard. Three parameters are being examined; the initial contention size (W), the packet retry limit (m) and the number of backoff stages (m'). Our performance analysis examines the following metrics as good indicators for the performance of the IEEE 802.11 protocol, namely the throughput efficiency, the average packet delay, the packet drop probability as well as the average time to drop a packet.

Various sets of protocol parameter values have been examined and compared with parameter values that the IEEE 802.11 standard proposes in order to identify potential improvements on protocol performance. After an extensive performance study, we have identified three sets of parameter values. Each set of parameter values achieves better performance on some particular metrics and it can be employed according to the specific communication needs. For example, one set of parameter values can significantly improve the throughput efficiency whereas another combination of parameters can considerably reduce the packet drop probability or the packet drop time.

The following three sets of parameter values that are being employed for the basic access scheme, for the case of "long" packets of l=1500 bytes[3] and compared with the values that the IEEE 802.11 protocol proposes (W=32, m=6, m'=5) are:

a) W=64, m=5, m'=4

b) W=64, m=5, m'=3

[3] Results for the RTS/CTS scheme and other packet sizes such as "short" VoIP packets of l=200 bytes have reached exactly the same conclusions, denoting that the proposed improvement does not depend on the employed access scheme or the packet payload size.

c) W=64, m=7, m'=3

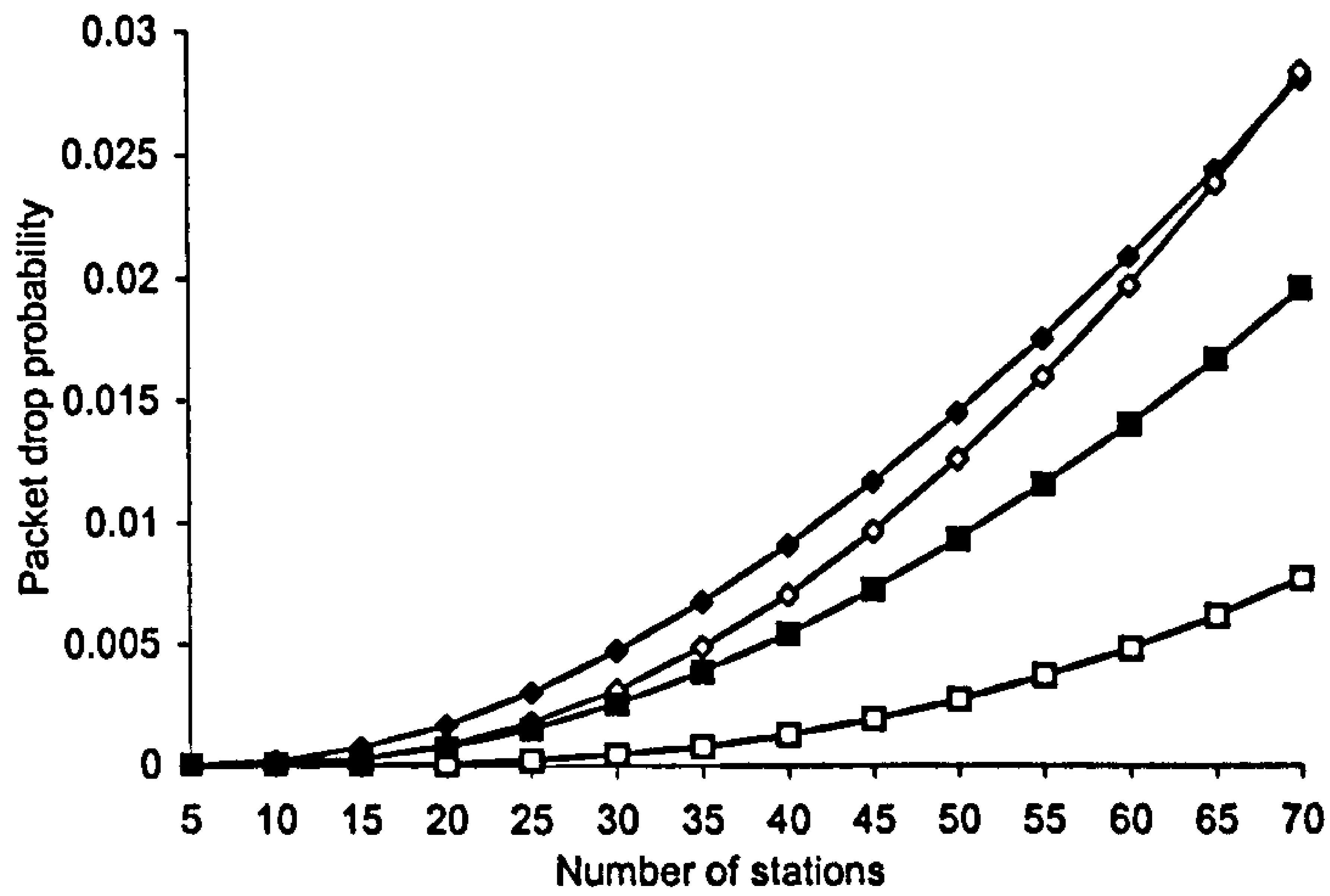
In all considered cases we increase the value of W to reduce the number of collisions. In the first case, the CW_{max} value that the standard proposes (CW_{max}=1024) is utilized by decreasing m' to 4; a lower retry limit (m=5) is considered sufficient since increasing W to 64 reduces the collision probability. In the second set, we study the effect of reducing CW_{max} to 512 by decreasing m' to 3; this set is expected to improve the average packet delay. Finally, in the last set the retry limit is increased to the value of 7. As a result, a contending station utilizes two more times the (relatively) small last backoff stage (CW_{max}=512) aiming to reduce the packet drop probability while keeping a fairly low packet delay.

At a first glance, it might seem that the choice of a higher value for the initial contention window size (W=64) comparing to the value of the standard (W=32) will cause a performance decrease in a small network scenario. A closer study to the case of a small network size ($2 \leq n \leq 6$) was performed and table 3.4 presents the packet delay and throughput efficiency for the two different values of the initial contention window W. The table illustrates that the adjustment of W to a higher value does not cause a considerable effect on both the packet delay and throughput efficiency for very small networks; on the contrary performance is improved in networks with five or more contending stations.

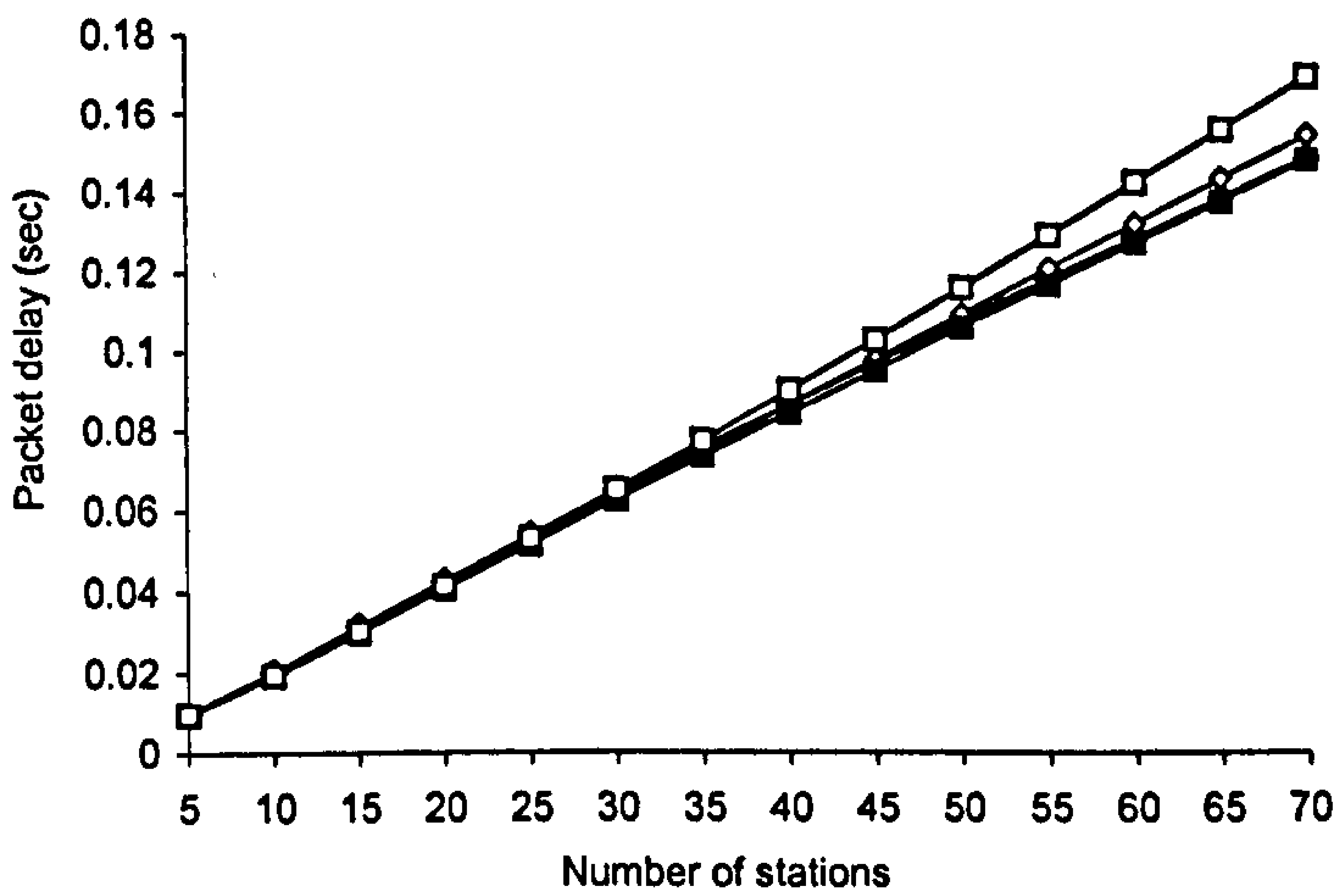
Number of stations	IEEE 802.11 standard W=32, m=6, m'=5		W=64, m=6, m'=5	
	Packet delay (sec)	Throughput efficiency	Packet delay (sec)	Throughput efficiency
n=2	0.003779	0.577334	0.004049	0.538847
n=3	0.005664	0.577849	0.005843	0.560091
n=4	0.007624	0.572318	0.007683	0.567978
n=5	0.009647	0.565203	0.009564	0.570292
n=6	0.011722	0.557878	0.011485	0.569902

Table 3.4 Packet delay and throughput efficiency for a small network size (l=1500 bytes)

The efficiency of each set of parameter values on the packet drop probability is explored in figure 3.25(a) against the number of contending stations. When the standard proposed values are employed, a packet suffers the highest drop probability compared to the other three cases. The choice of a higher W value improves the drop probability



(a) *Packet drop probability*



(b) *Packet delay*

◆ W=32, m=6, m'=5 ■ W=64, m=5, m'=4

◇ W=64, m=5, m'=3 □ W=64, m=7, m'=3

Figure 3.25 Packet drop probability and packet delay against number of stations (l=1500 bytes)

since fewer collisions are taking place. When W=64, m=5, m' =3 are employed, the packet drop probability increases rapidly and gradually attains the same value with the standard proposed values in a large network scenario (n=70). This is justified by noting that employing W=64 and m'=3, the maximum value of the contention window size will be lower (CW_{max} = 512) compared to the one that the IEEE standard proposes (CW_{max} = 1024) resulting in an increased number of collisions when the number of contending stations is high. The lowest packet drop probability is achieved when W=64, m=7 and m' =3 since the packet drop probability is reduced up to 75% compared to the IEEE standard proposed values despite of the decrease of CW_{max}.

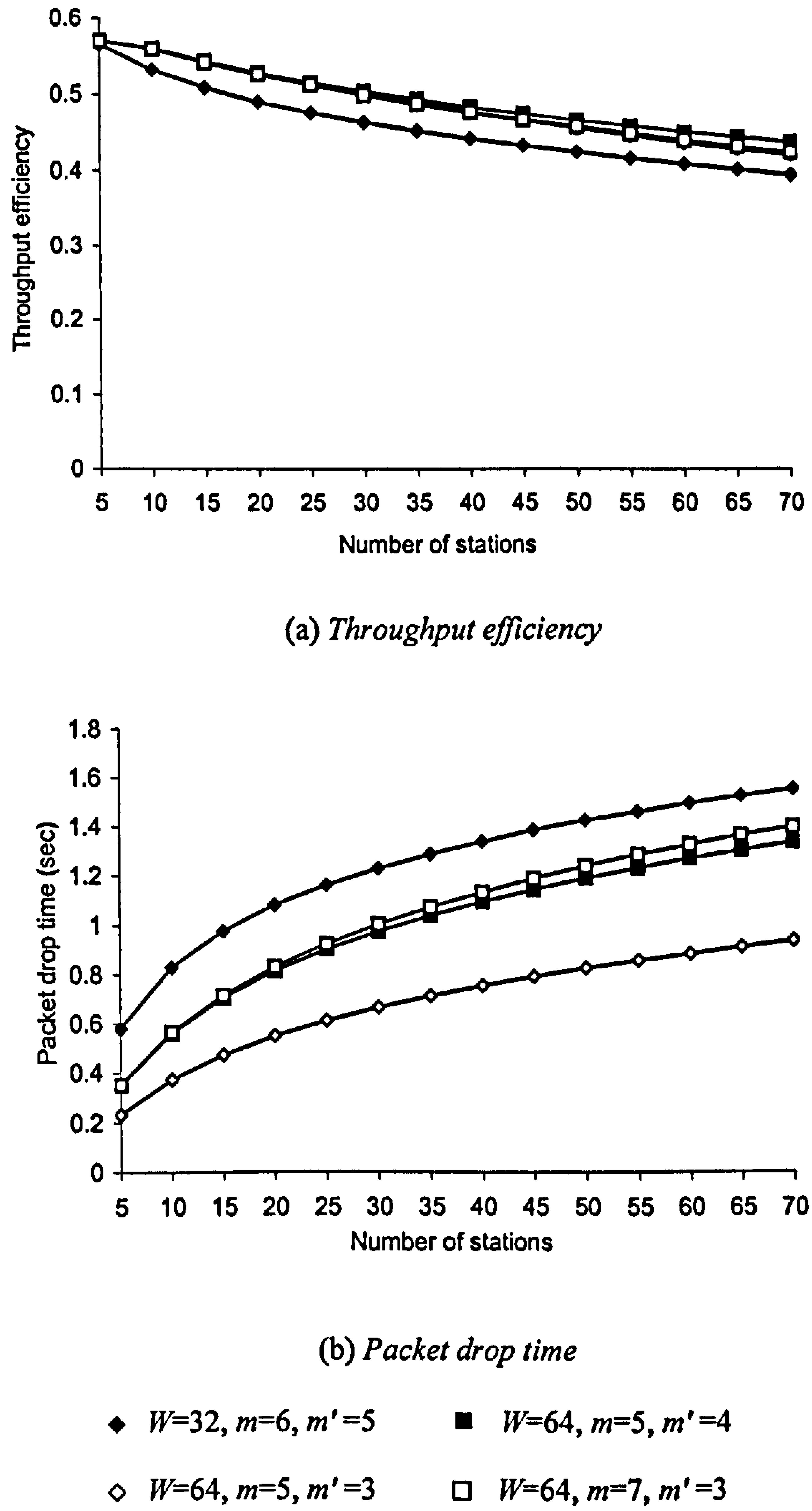


Figure 3.26 Throughput efficiency and packet drop time against number of stations (l=1500 bytes)

Figure 3.25(b) depicts that the packet delay increases when the network size grows in all cases due to the higher number of collisions. The figure also shows that the packet delay is not significantly affected by the employment of different parameter values. The only exception is when W=64, m=7, m' =3, the packet delay increases slightly faster than in the other cases when $n \geq 35$ and a packet experiences a small increase on delay of up to 10% in a large network (n=70). However, by means of figure 3.25(a) the situation is easily explained since a larger number of packets is transmitted successfully and not discarded. The small increase of the packet delay is a good price to pay for significantly decreasing the packet drop probability.

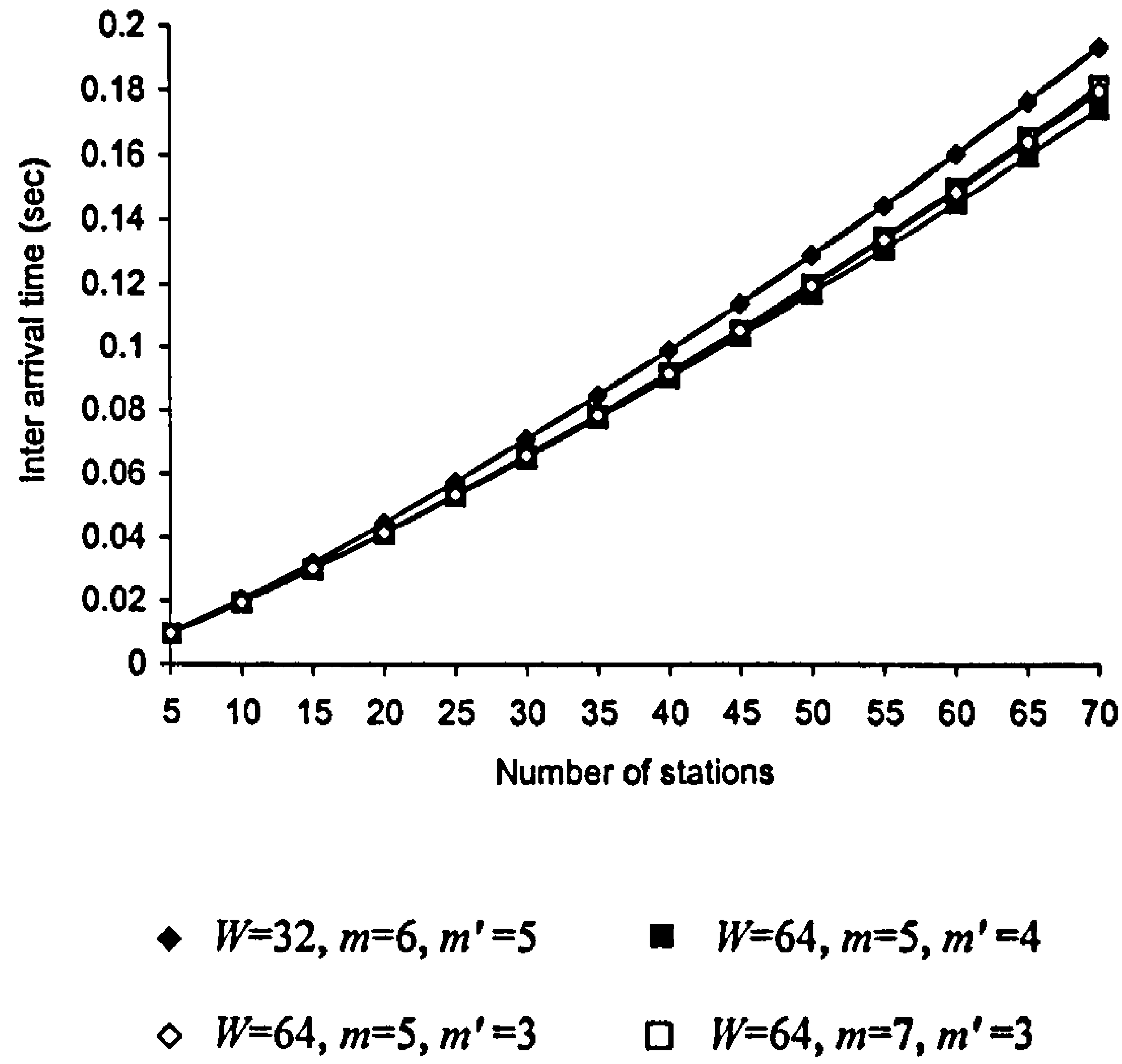


Figure 3.27 Packet inter arrival time against number of stations (l=1500 bytes)

Figure 3.26(a) examines the throughput efficiency that each considered set of parameter values achieves with varying the number of contending stations. Any of the proposed value sets achieves throughput efficiency higher compared to the IEEE 802.11 standard parameter values mainly because the larger W value decreases the number of collisions. Especially when W=64, m=5, m'=4 the increase on throughput can be up to 10% compared to the standard parameters.

Figure 3.26(b) plots the average time to drop a packet when it reaches the maximum retransmission limit against the number of contending stations. For all sets of parameter values, the packet drop time increases when the network size increases. The figure shows that the employment of any of the considered sets of parameter values, as compared to the IEEE standard parameters, results in a significant improvement on the packet drop time. The highest packet drop time is attained using the parameter values suggested in the standard, whereas the case of W=64, m=5, m'=3 achieves the lowest packet drop time with a reduction of about 40% for a large network size (n=70).

Finally, figure 3.27 studies packet inter arrival time, which is defined as the time interval between two successful packet receptions at the receiver. As expected, packet inter arrival time for the IEEE 802.11 standard parameters is considerably higher than any other case. This can be easily justified by noting that packet inter arrival time also includes the time for packets that have being discarded; this time is much greater for the case of W=32, m=6, m'=5 due to the high drop probability values (figure 3.25(a)).

Performance results reported in the previous figures show that when (W=64, m=5,

m' =4), lower packet drop probability, packet drop time, packet inter arrival time and better throughput performance are achieved compared to the values proposed by the standard. When the CW_{max} is decreased to a lower value (CW_{max} = 512) for the same retry limit (m=5), we attain the lowest packet drop time comparing to any other case but the drop probability increases considerably. On the contrary, the adjustment of the retry limit to a higher value (W=64, m=7, m'=3), results in the lowest packet drop probability and a small increase of packet drop time and delay due to larger number of packets being transmitted successfully. Each combination of parameters achieves an improved performance on some specific metrics compared to the standard proposed values and the choice of which set of protocol parameters should be employed depends on the specific communication requirements.

This chapter has presented an elegant and intuitive analysis that takes into account packet retry limits and leads to simple equations for additional performance metrics to throughput efficiency such as the average packet delay, the packet drop probability, the average time to drop a packet and the packet inter-arrival time for both basic access and RTS/CTS medium access schemes. Based on the derived mathematical model, an extensive and detailed study was carried out on the influence on performance of physical layer, data rate, initial CW size, maximum CW size and packet payload size on protocol performance. The next chapter develops three different approaches in improving performance for the IEEE 802.11 protocol; packet bursting, optimization of the RTS/CTS mechanism and an alternative backoff algorithm named DIDD (Double Increment Double Decrement).

CHAPTER 4

IEEE 802.11 Proposed enhancements

This chapter develops three different approaches in improving performance for the IEEE 802.11 protocol. Firstly, a different approach in enhancing performance through reducing overhead costs like backoff time and RTS/CTS exchanges is proposed. The concept of transmitting more than one data packets after winning DCF contention, named packet bursting, can be easily implemented through the fragmentation mechanism of the IEEE 802.11 protocol. The mathematical model for the legacy IEEE 802.11 (which was derived in the previous chapter) is extended in order to consider packet bursting. Results obtained for different scenarios showed that the application of packet bursting significantly enhances both throughput and packet delay performance. Furthermore, fairness is explored for both the legacy DCF and packet bursting cases in short-time and long-time scales.

Secondly, the mathematical model developed in the chapter 3 is utilized to study the effectiveness of the RTS/CTS scheme in reducing collision duration at high data rates for both IEEE 802.11b and 802.11a protocols. An all-purpose expression for the RTS threshold value is derived that actually maximizes performance by employing the RTS/CTS reservation scheme whenever it is beneficial for both the packet delay and throughput performance.

Finally, a new easy-to-implement backoff algorithm named DIDD (Double Increment Double Decrement) is introduced. An alternative and simpler mathematical analysis is developed based on elementary conditional probability arguments rather than bi-dimensional Markov chains. Detailed results are presented to identify the improvement of DIDD in throughput and packet drop performance comparing to the binary exponential backoff algorithm utilized in the legacy IEEE 802.11.

4.1 Performance improvement via packet bursting

The concept of transmitting more than one data packets after winning DCF contention is called packet bursting. It is included in the latest 802.11e draft specification and has been discussed in [56]. The number of pending data packets that a station will transmit with packet bursting depends on the data and control rate it is employing. The advantage of packet bursting is the increased throughput due to the reduction of contention periods and RTS/CTS exchanges at the cost of short-time unfairness.

A. Implementation issues of packet bursting

Figures 4.1 and 4.2 illustrate how packet bursting is applied to both basic access and RTS/CTS schemes. The presented implementation of packet bursting is based on the fragmentation mechanism of the IEEE 802.11 protocol. This mechanism provides a simple and practical way for stations to hold the medium for multiple packet transmissions when high data rates are utilized. A station that implements packet bursting transmits a burst of *ppb* (packets-per-burst) packets before releasing the medium. The receiving station individually acknowledges every DATA packet by sending an ACK packet after a SIFS interval and the transmitting station sends the next DATA packet upon reception of this ACK (again after SIFS). If any DATA packet transmission fails (an ACK is not received) the burst is terminated and the station shall attempt to contend for the medium and retransmit the failed DATA packet and the packets following it. Since the SIFS interval is shorter than the DIFS, it is ensured that the sender retains control over the medium and that no other station can go into contention and start transmitting until all the packets that belong to the burst are transmitted.

A description for the NAV usage in packet bursting when the RTS/CTS mechanism is employed (figure 4.2) is given next. The duration information included in the RTS and CTS packets is used to update the NAV of the stations to indicate that the channel is busy until the end of ACK 1. Both DATA 1 and ACK 1 packets contain duration information to update the NAV of all receiving stations to indicate a busy channel until the end of ACK 2. This carries on until the last DATA packet, which carries the duration of one ACK time plus one SIFS time in its duration field. The ACK for the last DATA packet has the duration field set to zero. Thus, each DATA/ACK pair acts as virtual RTS/CTS for the next DATA/ACK exchange and no further RTS/CTS packet exchange is necessary. Also every DATA packet (except the last one) has the *more fragments* flag in the MAC header set to 1 in order to indicate the use of the fragmentation mechanism. The MAC header of the DATA packets also carries the packet number that is used by the destination to arrange the order of the DATA packets (in the case of a single packet transmission, this field is set to 0).

An alternate mechanism to transmitting a specific number of packets after winning the DCF contention is to allow stations to transmit consecutive packets provided that the total access time does not exceed a certain limit (TXOP limit). This mechanism is introduced in IEEE 802.11e and the implemented number of packets per burst depends on the transmission rate and on the signal quality at the receiver. As stations implementing the packet bursting mechanism utilise the standard backoff procedure and thus experience the same delays but transmit more information after winning the contention for the medium, it is expected that packet bursting should improve performance. When a station that implements packet bursting has only one packet available in the station's queue, normal DCF procedures are used and the system has the same performance as without packet bursting.

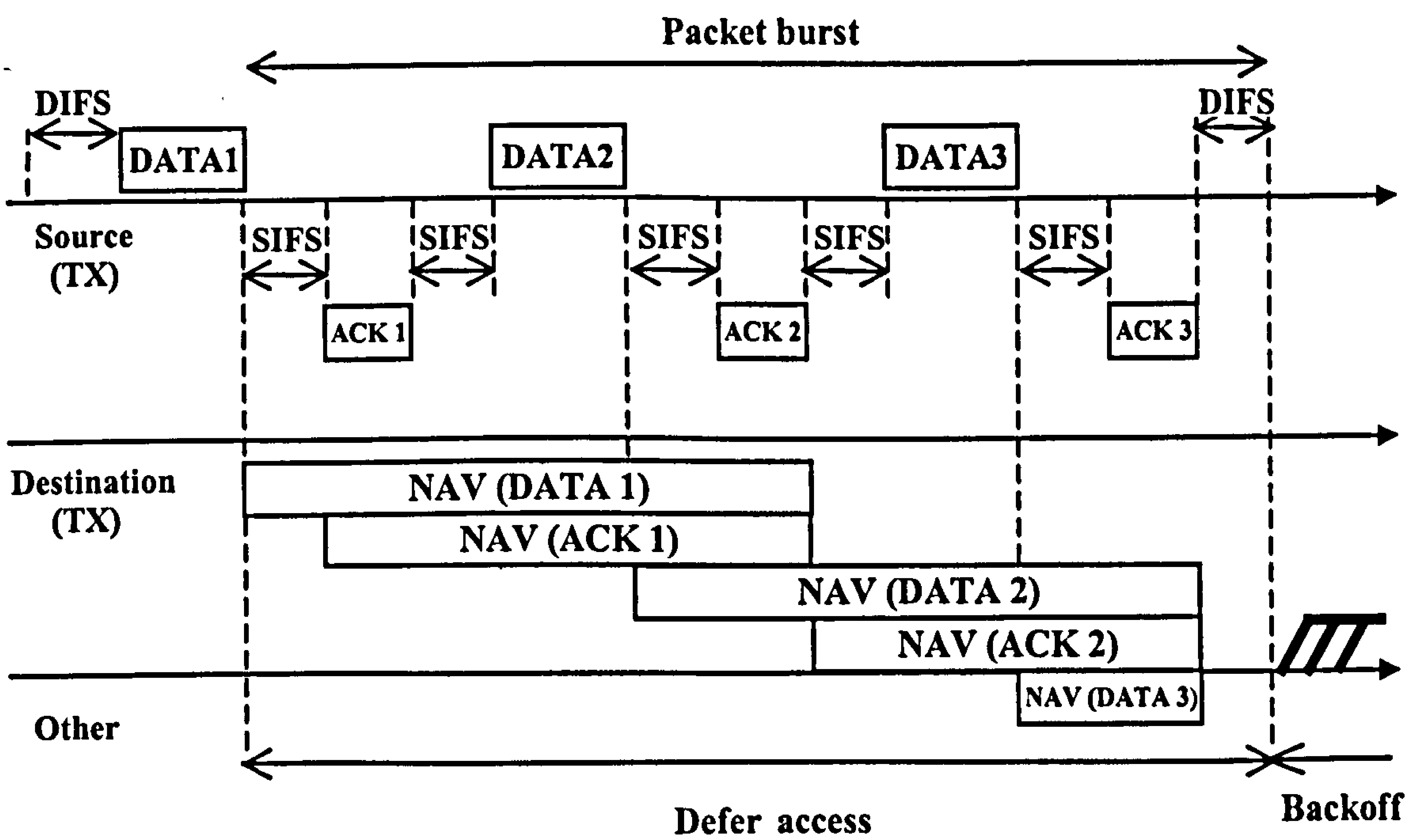


Figure 4.1 Implementation of packet bursting to basic access scheme (ppb=3)

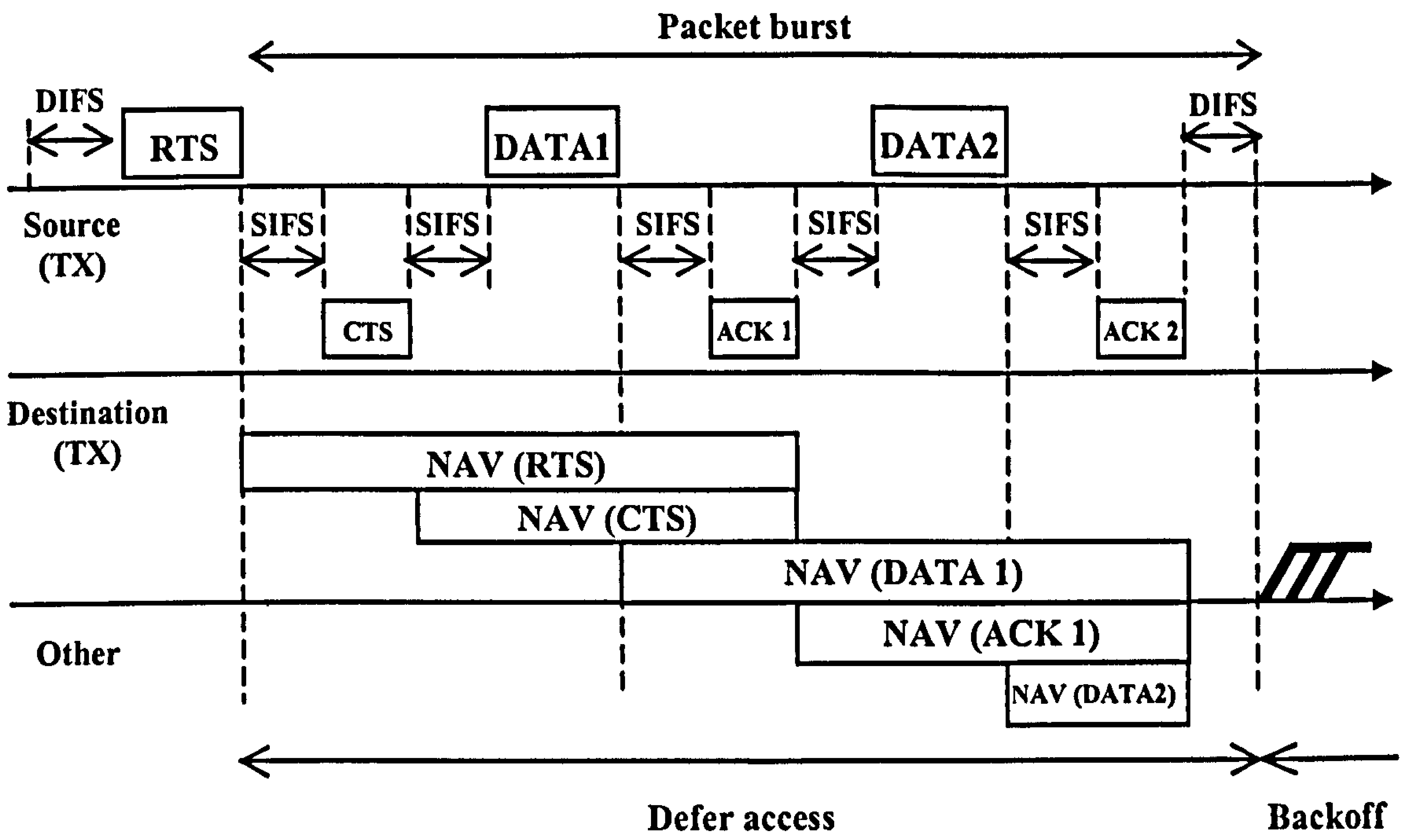


Figure 4.2 Implementation of packet bursting to RTS/CTS scheme (ppb=2)

B. Analytical modelling of packet bursting

The saturation throughput S in the case of packet bursting is computed as:

$$S_{burst} = \frac{P_{tr}P_S\, ppb\, l}{E'[slot]} = \frac{P_{tr}P_S\, ppb\, l}{(1-P_{tr})\sigma + P_{tr}P_S T_S' + P_{tr}(1-P_S)T_C'} \quad (4.1)$$

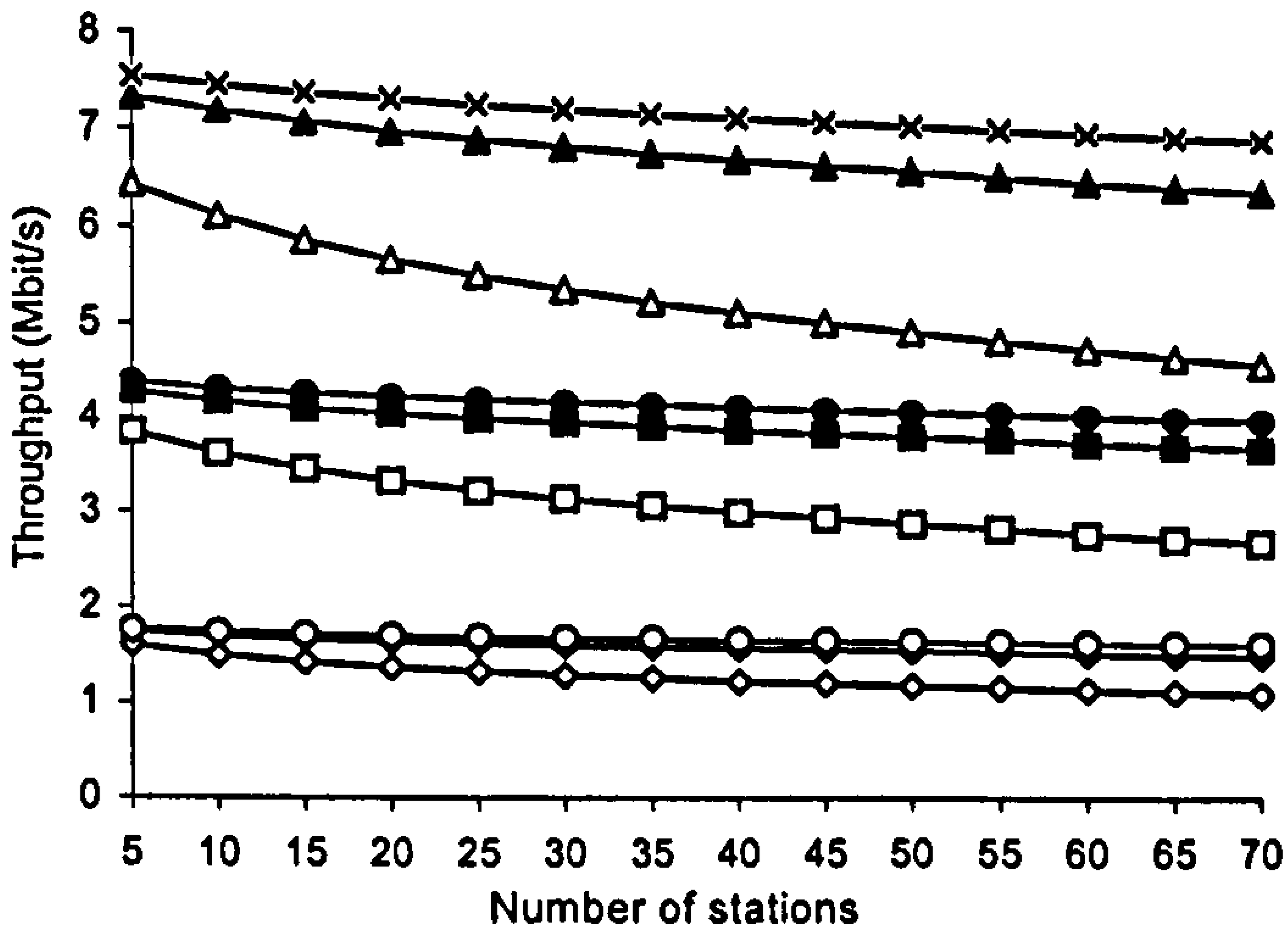
where *ppb* is the number of packets per burst employed by all stations, $E'[slot]$ is the average slot time when packet bursting is used, T_S' and T_C' are the average durations the medium is sensed busy due to a collision and a successful transmission respectively for packet bursting transmissions. The values of T_S' and T_C' for the basic and the RTS/CTS access mechanisms are given by (4.2)-(4.3), respectively.

$$\begin{cases} T_S'^{\,bas} = DIFS + ppb\; T_{DATA} + (2ppb-1)SIFS + ppb\,T_{ACK} + 2\,ppb\;\delta \\ T_C'^{\,bas} = DIFS + T_{DATA} + SIFS + T_{ACK} + 2\,\delta \end{cases} \quad (4.2)$$

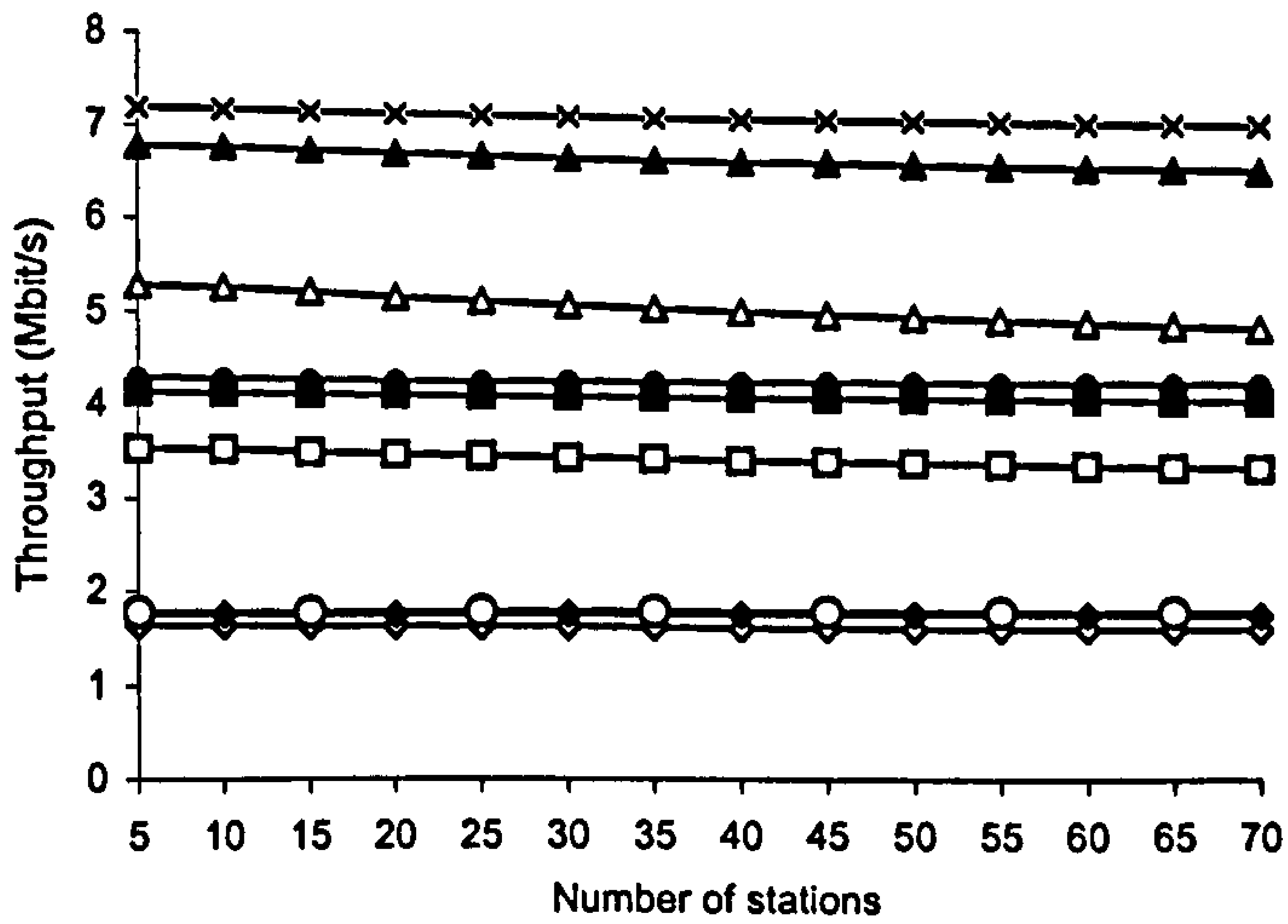
$$\begin{cases} T_S'^{\,RTS} = DIFS + T_{RTS} + T_{CTS} + ppb\; T_{DATA} + (2ppb+1)SIFS + ppb\,T_{ACK} + (2ppb+2)\delta \\ T_C'^{\,RTS} = DIFS + T_{RTS} + SIFS + T_{CTS} + 2\delta \end{cases} \quad (4.3)$$

C. Performance evaluation of packet bursting

Figures 4.3, 4.4, 4.5 and 4.6 illustrate the substantial improvement of packet bursting on performance; they plot throughput and packet delay versus network size for different burst size values and data rates for both basic access and RTS/CTS schemes. All figures clearly show that packet bursting substantially enhances performance by increasing throughput and reducing packet delay. This is explained by considering that packet bursting reduces the overhead by amortizing the cost of the contention period and RTS/CTS packet exchange over several packets.



(a) *Basic access*



(b) *RTS/CTS scheme*

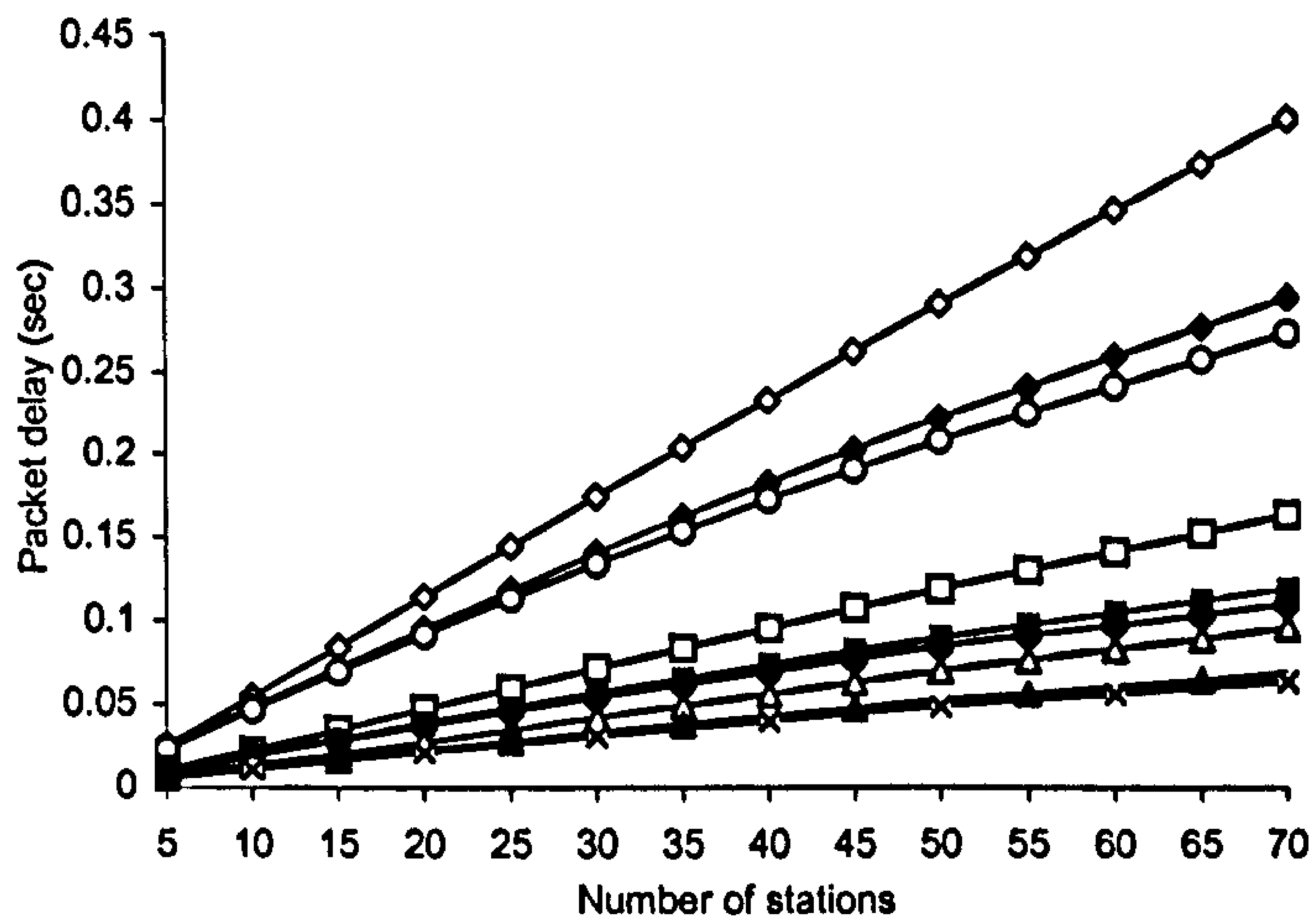
△ *C=11 Mbit/s, ppb=1* ▲ *C=11 Mbit/s, ppb=3* x *C=11 Mbit/s, ppb=5*

□ *C=5.5 Mbit/s, ppb=1* ■ *C=5.5 Mbit/s, ppb=3* • *C=5.5 Mbit/s, ppb=5*

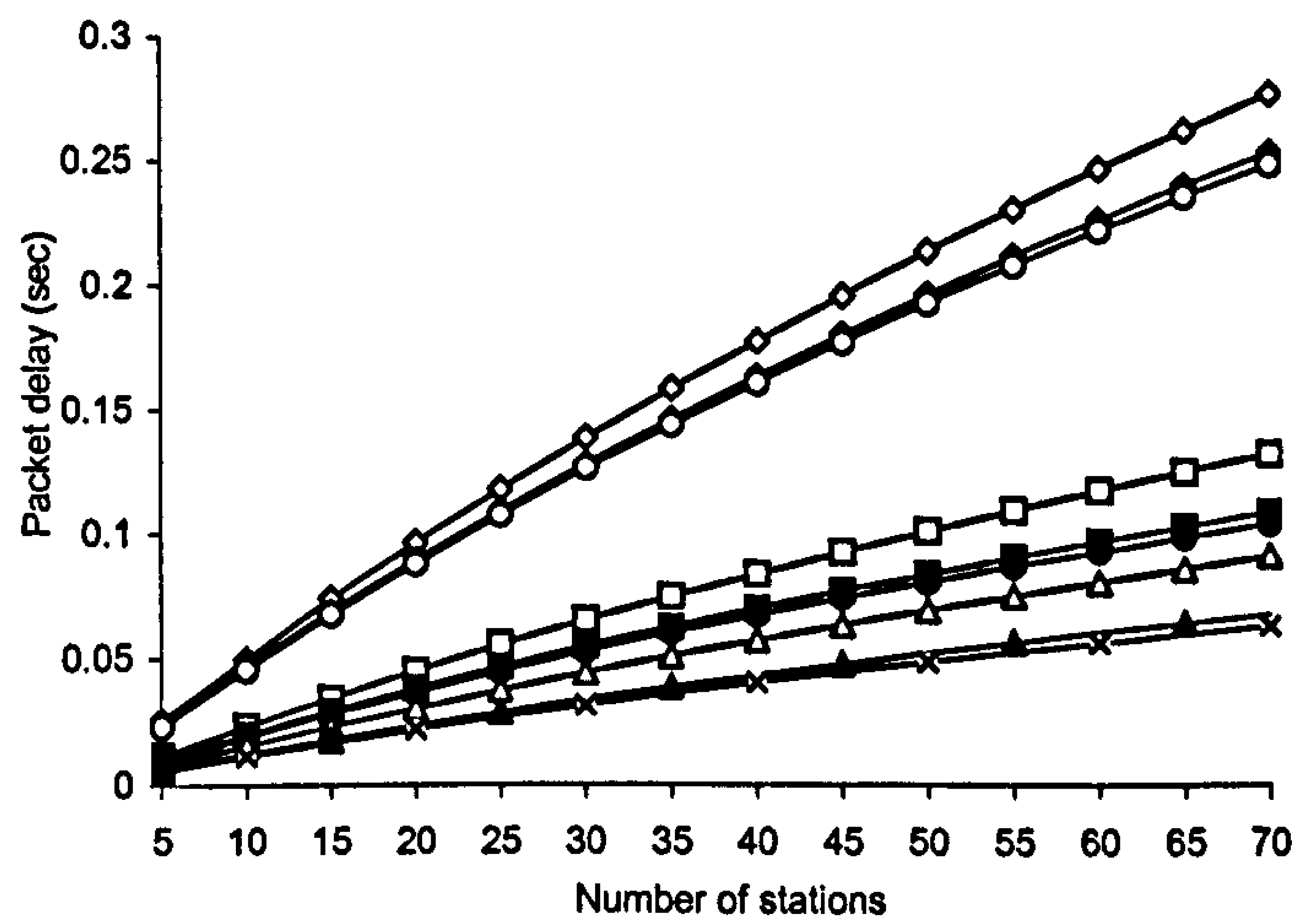
◇ *C=2 Mbit/s, ppb=1* ◆ *C=2 Mbit/s, ppb=3* o *C=2 Mbit/s, ppb=5*

Figure 4.3 Throughput enhancement of packet bursting for basic access and RTS/CTS mechanisms

It is quite interesting to study why and how packet bursting increases performance in different scenarios. When no packet bursting is implemented (*ppb*=1) and the basic access is used (figure 4.3(a)), throughput considerably decreases for all data rates when the network size increases due to the increased packet collision probability. When *ppb*=1 and the RTS/CTS scheme is used, throughput is not significantly affected from network size increase for *C*=2 Mbit/s because the increased packet collision probability does not degrade performance due to the short collision duration. However, for higher data rates (*C*=5.5 Mbit/s and *C*=11 Mbit/s), throughput degrades with network size



(a) *Basic access*



(b) *RTS/CTS scheme*

△ *C=11 Mbit/s, ppb=1* ▲ *C=11 Mbit/s, ppb=3* x *C=11 Mbit/s, ppb=5*

□ *C=5.5 Mbit/s, ppb=1* ■ *C=5.5 Mbit/s, ppb=3* • *C=5.5 Mbit/s, ppb=5*

◇ *C=2 Mbit/s, ppb=1* ◆ *C=2 Mbit/s, ppb=3* o *C=2 Mbit/s, ppb=5*

Figure 4.4 Throughput enhancement of packet bursting for basic access and RTS/CTS mechanisms

increase because the collision duration is high compared to the data rate as the RTS and CTS control packets are always transmitted at the lower control rate of 2 Mbit/s.

When packet bursting is utilized (*ppb*=3 and *ppb*=5) for the basic access scheme, throughput considerably increases, especially for large networks with increased collision probability, mainly because packet bursting shortens the duration of collisions as compared to the duration of successful transmissions! Collisions involve only the first DATA packet of the packet burst because the lack of the first ACK packet forces the transmitting stations to contend again for medium access; successful medium accesses last much longer as they involve the transmission of a burst of packets.

When packet bursting is utilized (*ppb*=3 and *ppb*=5) for the RTS/CTS scheme, throughput is not significantly increased for *C*=2 Mbit/s due to the relatively short RTS/CTS collision duration. However, at higher data rates, (*C*=5.5 Mbit/s and *C*=11 Mbit/s), throughput is considerably increased because packet bursting reduces the number of medium reservations that involve the transmission of RTS and CTS packets at the low data rate of *C*=2 Mbit/s.

D. Fairness issues

The main purpose of a successful packet bursting implementation is the selection of a reasonable packet burst size value that improves performance and, at the same time, prevents stations from capturing the medium for long periods. Medium capture is undesirable and creates fairness problems. The fairness of a protocol is measured in terms of how resources are assigned to different stations over a period of time. Based on the length of this time period, the fairness can be measured on short-term or on long-term basis.

Intuitively, short-term fairness of a protocol refers to its ability to allocate the channel bandwidth equally to competing stations over short time periods; long-term fairness, in contrast, measures the same ability over longer time periods. The short-term fairness automatically implies long-term fairness, but not the vice versa.

To measure fairness, this work utilizes the average fairness index proposed by Jain [68]:

$$F_J = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} x_i^2} \tag{4.4}$$

where n is the number of stations and x_i is the throughput of station i during the considered window size of w successful packet transmissions. Absolute fairness is achieved when $F_J = 1$ (all stations equally share the medium) and absolute unfairness (a station monopolizes the channel) is achieved when $F_J = 1/n$.

In figure 4.5, we examine the fairness of packet bursting (utilizing the average Jain's fairness index) by considering two window size values that represent a short-term scale (*w*=1000 packets) and long-term scale (*w*=10000 packets). The figure reveals the weak fairness of both the packet bursting and the legacy IEEE 802.11 on a short-term scale (a small window size exhibits high unfairness). In fact, the fairness index is considerably lower than one when packet bursting is not utilized, especially for large network size

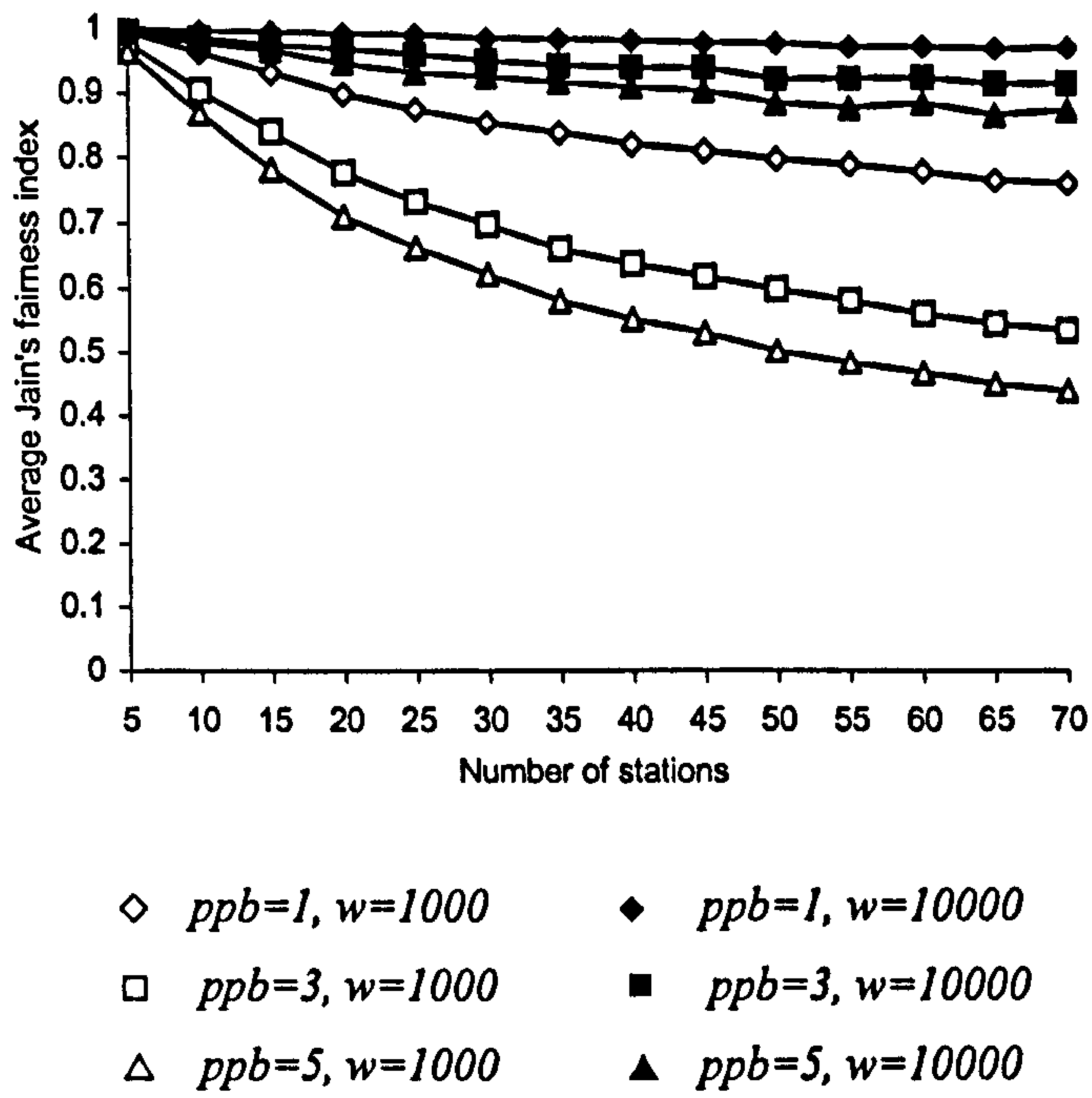


Figure 4.5 Fairness of packet bursting over short and long time scale, Basic access, C=2 Mbit/s

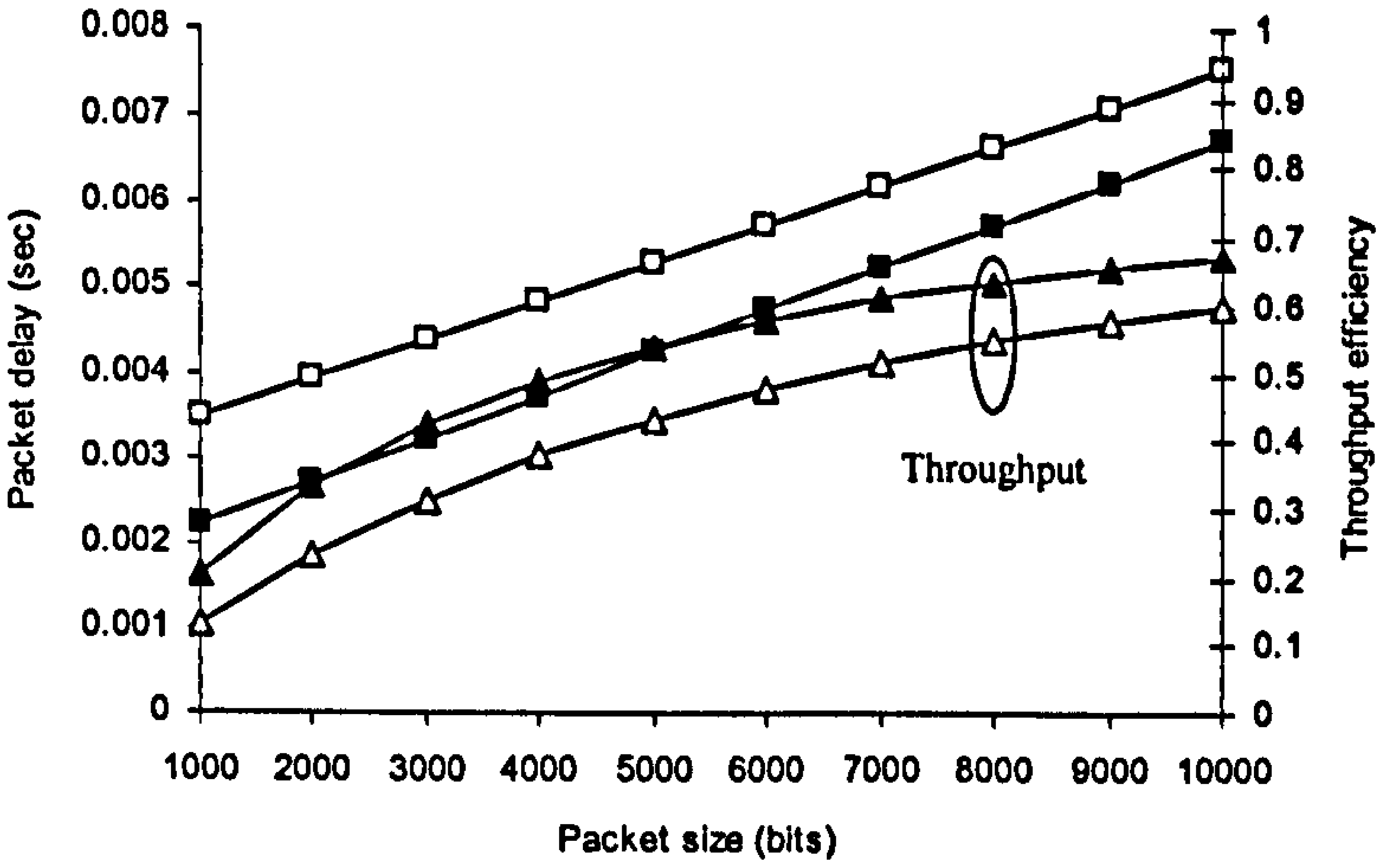
values. However, fairness improves in both cases when the window size used for measurement is increased, ensuring long-term fairness (in long-term all contending stations experience on average the same number of collisions).

4.2 Optimisation of the RTS/CTS reservation mechanism

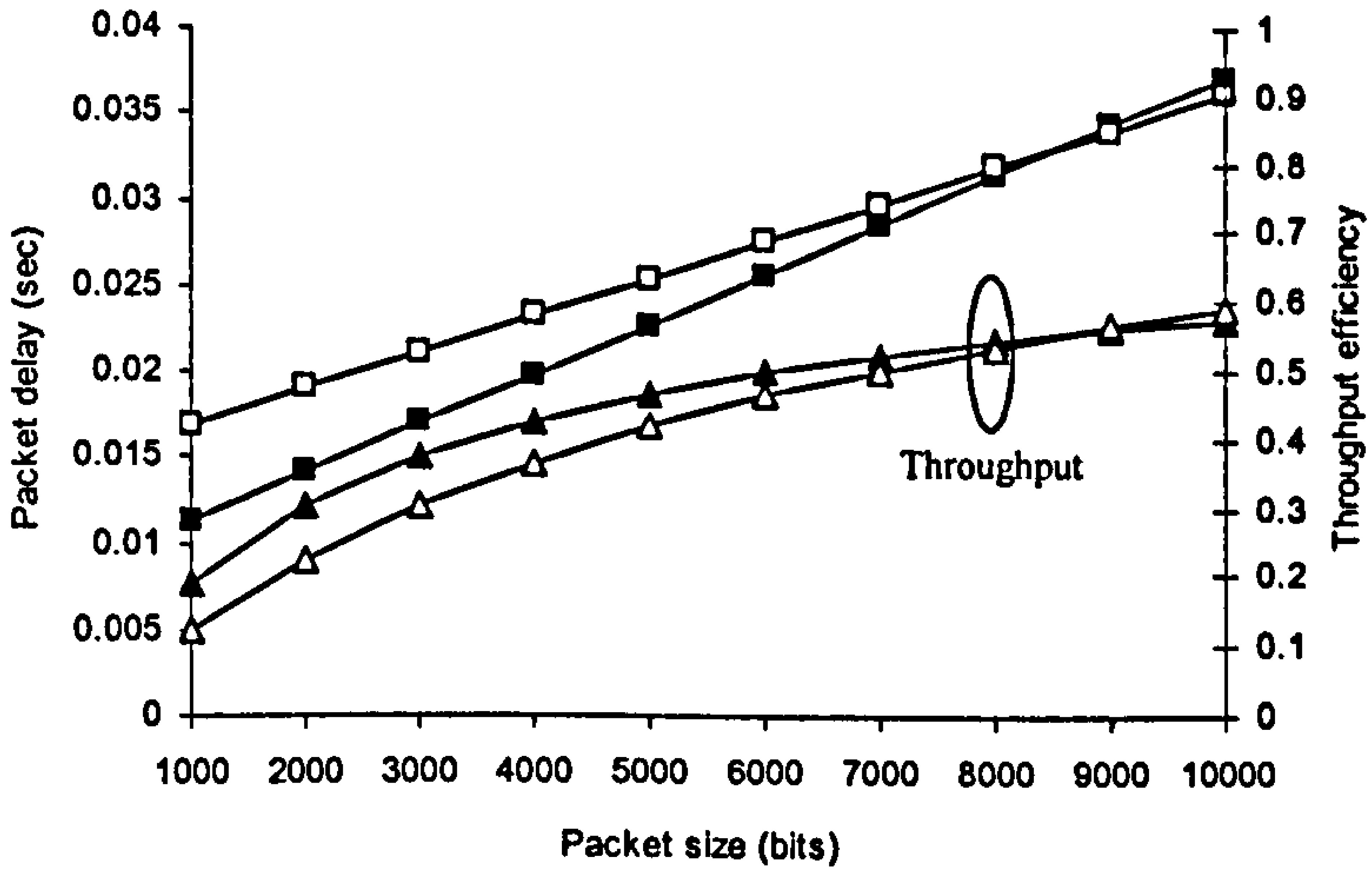
4.2.1 Inefficiency of RTS/CTS scheme at high-data rates

By utilizing the mathematical model for throughput and delay as 'performance metrics', we explore the effectiveness of RTS/CTS reservation mechanism for collision duration decrease at high-data rate IEEE 802.11b and IEEE 802.11a Wireless LANs and for different data and control transmission rates without. Actually, this section proves that the overall WLAN performance suffers significantly when the lower rate RTS/CTS exchange is combined with higher transmission data rates.

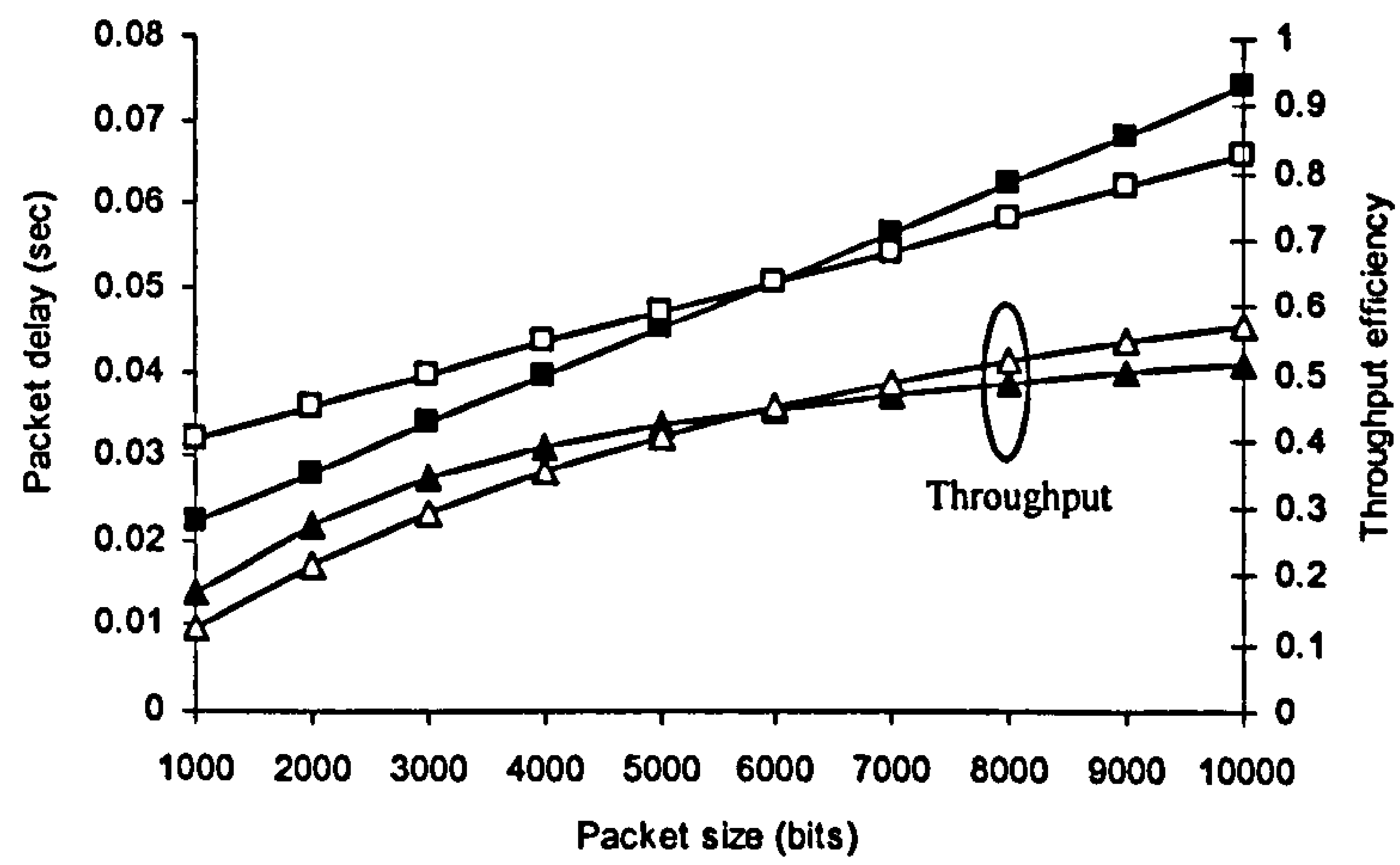
Figures 4.6(a), 4.6(b) and 4.6(c) study the effectiveness of the RTS/CTS scheme for IEEE 802.11b WLANs in high data rates (C=11 Mbit/s) by plotting throughput and average packet delay versus packet size for small (n=5), medium (n=25) and large (n=50) network sizes, respectively. The best-case scenario is considered where control packets (RTS, CTS and ACK) are transmitted at the highest possible control rate (2 Mbit/s) and the short PHY header is utilized. The figures demonstrate that both packet delay and throughput increase, as the data packet size increases. Note that the curves for packet delay and throughput cross in exactly the same point in both the basic access and RTS/CTS schemes.



(a) *n=5*



(b) *n=25*



(c) *n=50*

■ *Packet delay, Basic* ▲ *Throughput, Basic*

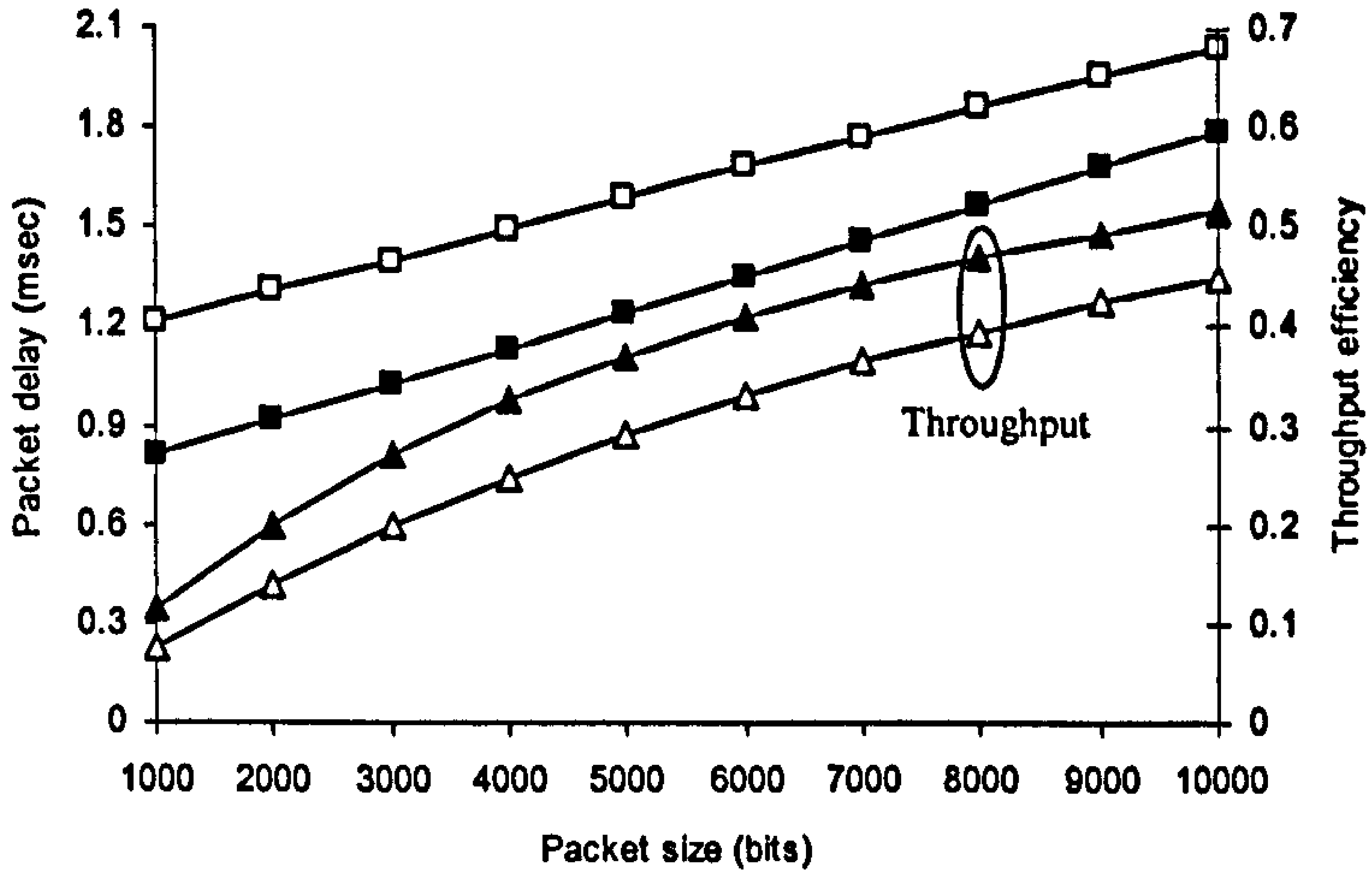
□ *Packet delay, RTS/CTS* △ *Throughput, RTS/CTS*

Figure 4.6 Packet delay and throughput versus packet size, C= 11 Mbit/s, C_{con}= 2 Mbit/s

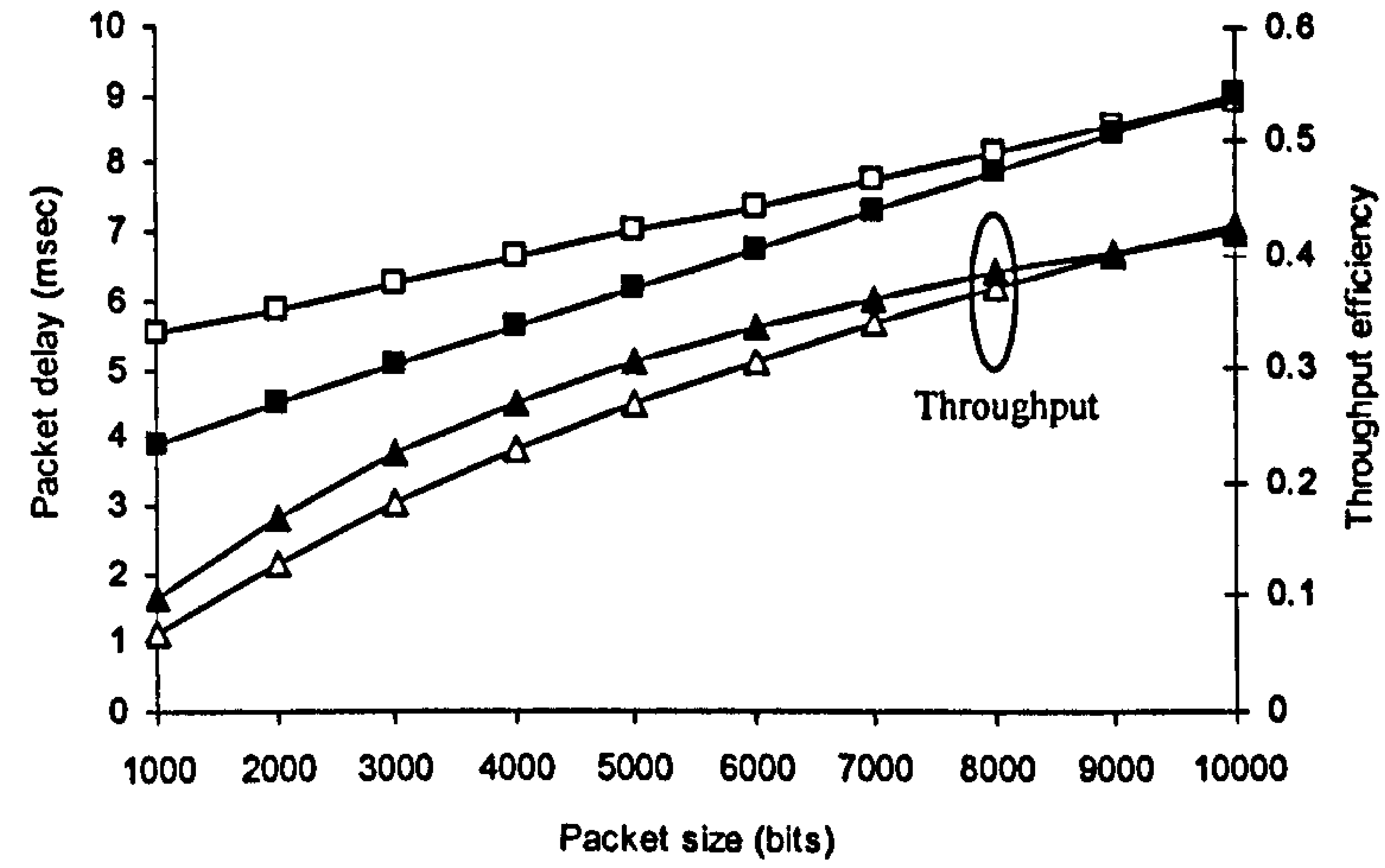
Figure 4.6(a) illustrates that the basic access outperforms RTS/CTS when the number of contending stations is relatively small (n=5) for all packet size values. This expected outcome confirms that the RTS/CTS reservation scheme is not beneficial for small size networks due to the low collision probability and is consistent with the conclusion derived in [6] for the data rate of 1 Mbit/s. Figure 4.6(b) illustrates the case of a medium network size (n=25) with a much higher collision probability; the RTS/CTS scheme attains lower packet delay and higher throughput than the basic access scheme for packet sizes l >8500 bits. This RTS threshold value is large due to the much lower control rate considerably degrades performance. Furthermore, figure 4.6(c) shows that even when the collision probability increases significantly as a result of the large number of contending stations (n=50), the RTS/CTS scheme is advantageous to basic access for relatively large packets (l> 6000 bits). Similar figures (not shown due to correspondence) for intermediate network size values of n=20, 30 and 40, show that the RTS/CTS scheme enhances performance only when the length of data packets exceeds 9500, 8000 and 6500 bits, respectively.

Figures 4.7(a), 4.7 (b) and 4.7 (c) investigate the performance of the RTS/CTS scheme for the IEEE 802.11a in high data rates by plotting throughput efficiency and average packet delay versus packet size for three representative network sizes (n=5, 25 and 50, respectively). Once more, the best-case scenario is considered where control packets are transmitted at the highest possible control rate (24 Mbit/s). The figures demonstrate that both throughput efficiency and packet delay increase, as the data packet size increases. The figures also show that similar results are acquired for the case of IEEE 802.11a as in IEEE 802.11b; the RTS/CTS reservation scheme is superior to basic mechanism only for medium or large network sizes due to the overhead ratio problem, as explained before. This benefit of RTS/CTS scheme in congested environments is justified since the more contending stations are, the heavier traffic load, and the more advantage the RTS/CTS mechanism can gain. On the other hand, the higher the data rate and the smaller packet size, the larger the overhead ratio of the RTS/CTS mechanism. This degrades the overall performance in terms of packet delay and throughput efficiency. Thus, the RTS/CTS scheme enhances performance compared to basic access only when the length of data packets exceeds 9000 in medium network (n=25) and 6500 bits in large network sizes (n=50).

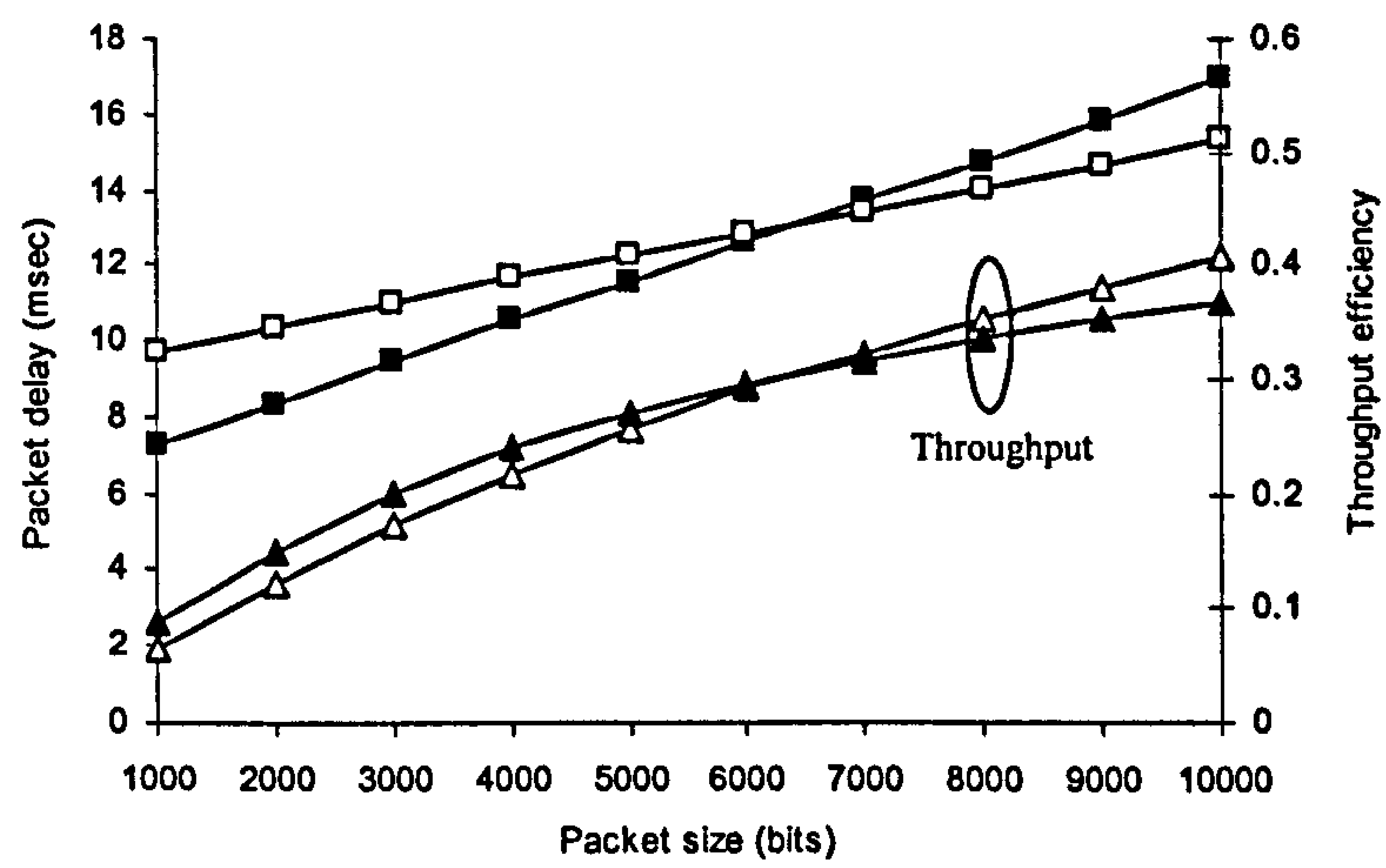
The presented performance results demonstrate the deficiency of the RTS/CTS scheme for high data rates (11 Mbit/s in IEEE 802.11b and 54 Mbit/s in IEEE 802.11a), unlike common expectation. We find that only very large packet size values render the



(a) *n=5*



(b) *n=25*



(c) *n=50*

■ *Packet delay, Basic* ▲ *Throughput, Basic*

□ *Packet delay, RTS/CTS* △ *Throughput, RTS/CTS*

Figure 4.7 Packet delay and throughput versus packet size, C= 54 Mbit/s, C_{con}= 24 Mbit/s

RTS/CTS beneficial compared to the basic access scheme. This result holds true even when the highest possible control rate (2 Mbit/s and 24 Mbit/s, respectively) is utilized and is explained by considering that the exchange of the RTS and CTS reservation packets at a much lower control rate results in a significant delay in communication.

4.2.2 Derivation of the RTS threshold

Performance results presented in the previous section indicate that the use of the RTS/CTS reservation scheme must strike a balance between the reduced collision duration and the increased overhead for the transmission of the RTS and CTS control packets. Therefore, the desire for optimal use of the RTS/CTS reservation scheme makes essential the derivation of an all-purpose expression for the threshold value, which determines when the RTS/CTS reservation scheme should be employed.

We indicate with D^{BAS} and D^{RTS} the average delay of a packet transmitted by the basic access and RTS/CTS mechanism, respectively. The threshold value should satisfy the following condition[1]:

$$\begin{aligned} D^{RTS} &= D^{BAS} \\ E[X]\ E[slot]^{RTS} &= E[X]\ E[slot]^{BAS} \\ P_S\, T_S^{RTS} + (1-P_S)\, T_C^{RTS} &= P_S\, T_S^{BAS} + (1-P_S)\, T_C^{BAS} \\ P_S\, (T_S^{RTS} - T_S^{BAS}) &= (1-P_S)\, (T_C^{BAS} - T_C^{RTS}) \end{aligned} \tag{4.5}$$

Let O_{RTS} be the overhead introduced by the RTS/CTS scheme where $O_{RTS} = T_S^{RTS} - T_S^{BAS} = \frac{l_{RTS}}{C_{con}} + 2\,SIFS + \frac{l_{CTS}}{C_{con}}$ for the case of the IEEE 802.11b physical layer and $O_{RTS} = T_S^{RTS} - T_S^{BAS} = T_{RTS} + 2\,SIFS + T_{CTS}$ for the IEEE 802.11a physical layer[2]. Moreover, $T_C^{BAS} - T_C^{RTS} = \frac{l}{C} + O_h$ where $O_h = T_{header} - T_{RTS} = \left(\frac{MAC_{hdr} + FCS}{C} + \frac{PHY_{hdr}}{C_{con}}\right) - \frac{l_{RTS}}{C_{con}}$ for the IEEE 802.11a physical layer and $O_h = \left(\frac{22 + MAC_{hdr} + FCS}{C} - \frac{22 + l_{RTS}}{C_{con}}\right)$ for the IEEE 802.11a physical layer is the extra length of the data packet header with respect to the RTS packet size. Thus, equation (4.5) becomes:

$$P_S\, O_{RTS} = (1-P_S)\left(\frac{l}{C} + O_h\right)$$

---

[1] Although, the derived expression is derived in order to minimize packet delay, the same approach can be followed for maximising throughput performance.

[2] Note that the values for T_{RTS} and T_{CTS} can be found by equations (3.23)-(3.24) for IEEE 802.11a.

$$\frac{P_S}{1-P_S} O_{RTS} = \frac{l}{C} + O_h$$

$$l_{threshold} = \left(\frac{P_S}{1-P_S} O_{RTS} - O_h \right) C \tag{4.6}$$

Equation (4.6) provides the threshold value $l_{threshold}$ over which it is beneficial to switch to the RTS/CTS mechanism. The value of this threshold size depends on the probability of a successful transmission P_S (it will be calculated next), the control C_{con} and the data rate C as well as the employed physical layer (802.11b or 802.11a) of the IEEE protocol. As already explained in Chapter 3, the conditional probability P_s is equal to:

$$P_S = n \cdot \tau \cdot (1-\tau)^{n-1} / 1-(1-\tau)^n \tag{4.7}$$

Recall that the collision probability p as a function of the transmission probability is:

$$p = 1 - (1 - \tau)^{n-1} \tag{4.8}$$

from which the number of stations n can be found as a function of p and τ:

$$n = \frac{ln(1 - p)}{ln(1 - \tau)} + 1 \tag{4.9}$$

Thus, after some algebra, equation (4.7) finally becomes:

$$P_S = \frac{\left(\frac{ln(1-p)}{ln(1-\tau)} + 1 \right) \tau (1-p)}{1-(1-p)(1-\tau)} \tag{4.10}$$

If we substitute the value of P_S into equation (4.6), then we can calculate the optimal RTS threshold for each station. More specifically, each station can measure the collision probability by counting the number of collisions and dividing them by the total number of transmission attempts for each packet. Furthermore, an alternative way to estimate p is proposed and discussed in [7].

4.2.3 Performance evaluation using the optimal RTS threshold

The following figures provide the RTS threshold values above which the employment of the RTS/CTS mechanism considerably enhances performance for both IEEE 802.11b and 802.11a physical layers.

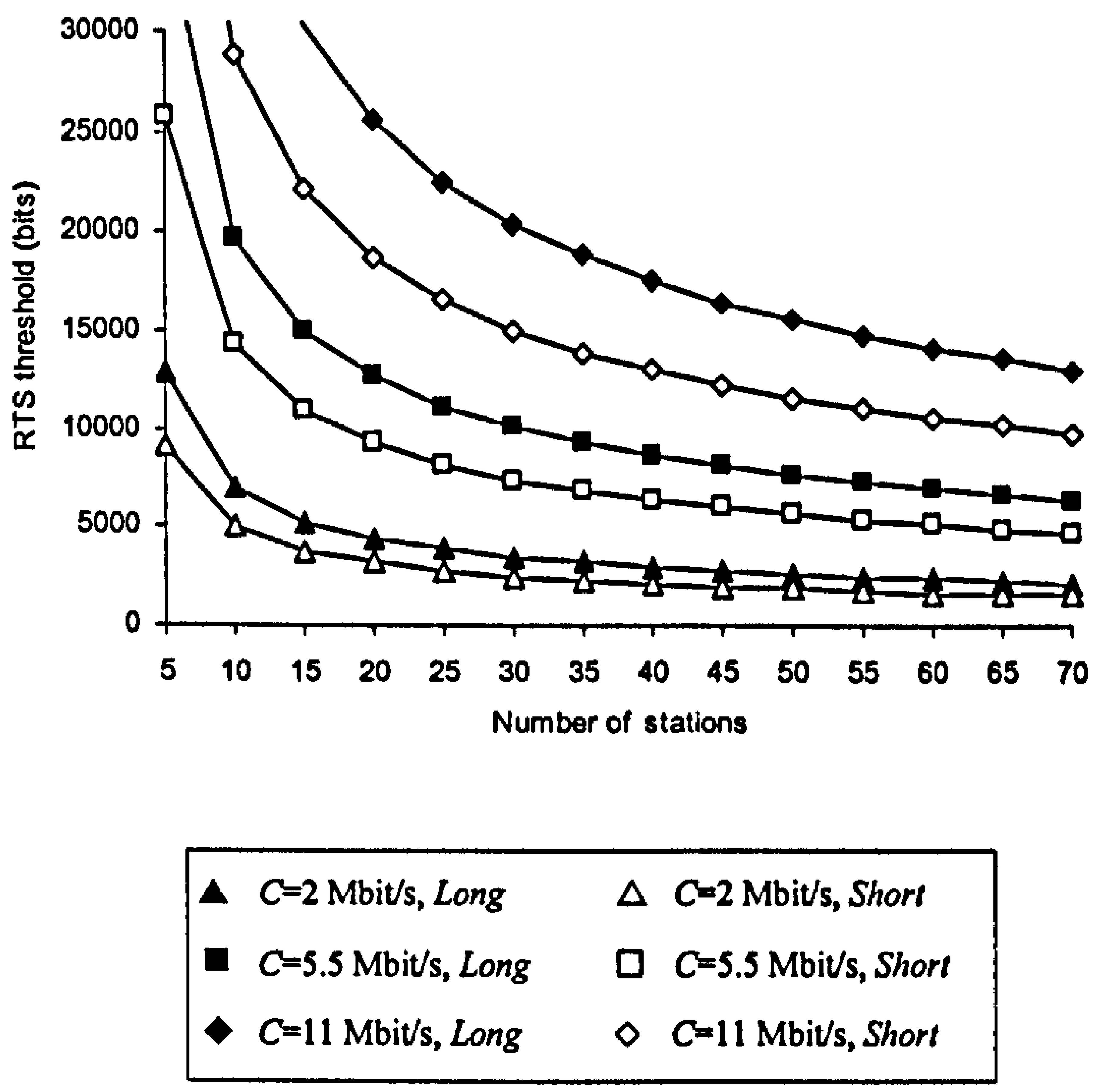


Figure 4.8 Effect of data rate and PLCP header on RTS threshold
(C= 11 Mbit/s, C_{con}= 2 Mbit/s)

Figure 4.8 plots RTS threshold versus network size for three data rates (C = 2, 5.5, and 11 Mbit/s) as well as for a short and long PHY packet overhead. According to figure 4.8, the packet size threshold is highly dependent on the data rate. When the data rate increases, the threshold values increase significantly. The reason is that although high data rates reduce the transmission time for data packets, the RTS and CTS control packets are still being transmitted by the low control rate, resulting in delay in communication. Moreover, the use of a short PHY header, which results in a shorter transmission time comparing to the long PHY header's transmission time, considerably decreases the packet size threshold value. This can easily be explained by considering that smaller packet overhead mainly reduces the overhead that the RTS and CTS control packets introduce. Thus, the main drawback (increased overhead) of the RTS/CTS scheme is minimized denoting that it can be employed for even smaller data packets.

We next study the effect of packet retry limit and initial contention window. Figure 4.9 plots the RTS threshold versus m and W, respectively, for four representative network sizes (n = 5, 25, 50 and 70) and data rate of C=11 Mbit/s. Both figures show that when the number of the contending stations is relatively small (n = 5), the RTS threshold attains high values that exceed the maximum packet size (without employing the fragmentation mechanism as specified by IEEE 802.11b) so the RTS/CTS scheme should not me employed due to the low packet collision probability.

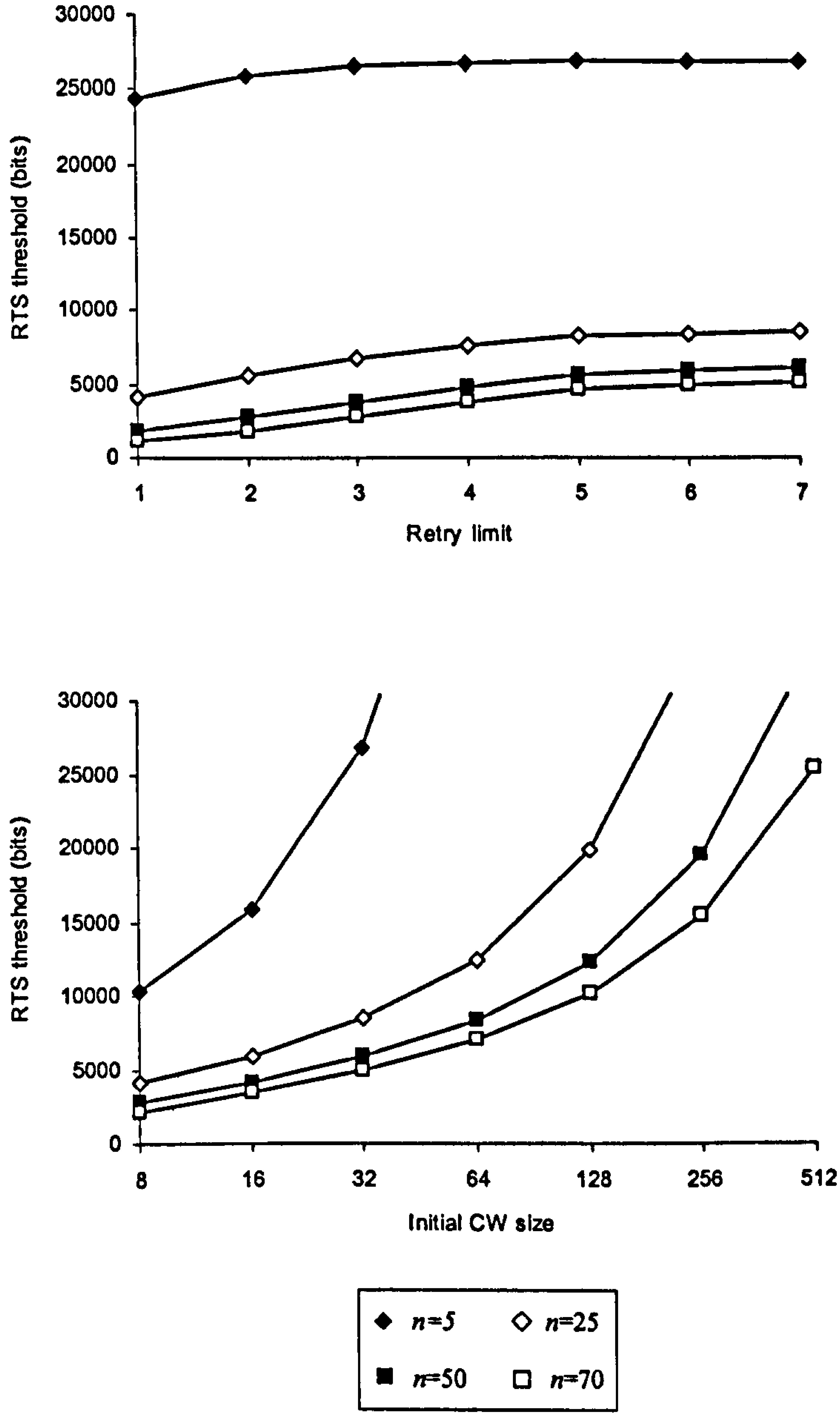


Figure 4.9 RTS threshold versus packet retry limit and initial contention window (CW) size (IEEE 802.11b, C= 11 Mbit/s, C_{con}= 2 Mbit/s)

When the network size increases, the RTS threshold decreases to lower values. This can be justified since large network sizes cause more packet collisions and a much lower successful transmission probability is achieved. We can see that the packet retry limit has a significant effect on RTS threshold; when retry limit increases, the RTS threshold values also increase due to the improved successful transmission probability. An interesting outcome is that for m>6, the RTS threshold is only marginally affected, indicating the proper choice of the retry limit value in the IEEE 802.11 standard. Furthermore, figure 4.9 shows that the RTS threshold values are also highly dependent on the initial contention window. In fact, small network sizes appear to be more sensitive on the initial contention window. A small increase of W results in a greater increase in the RTS threshold for small networks than for large networks.

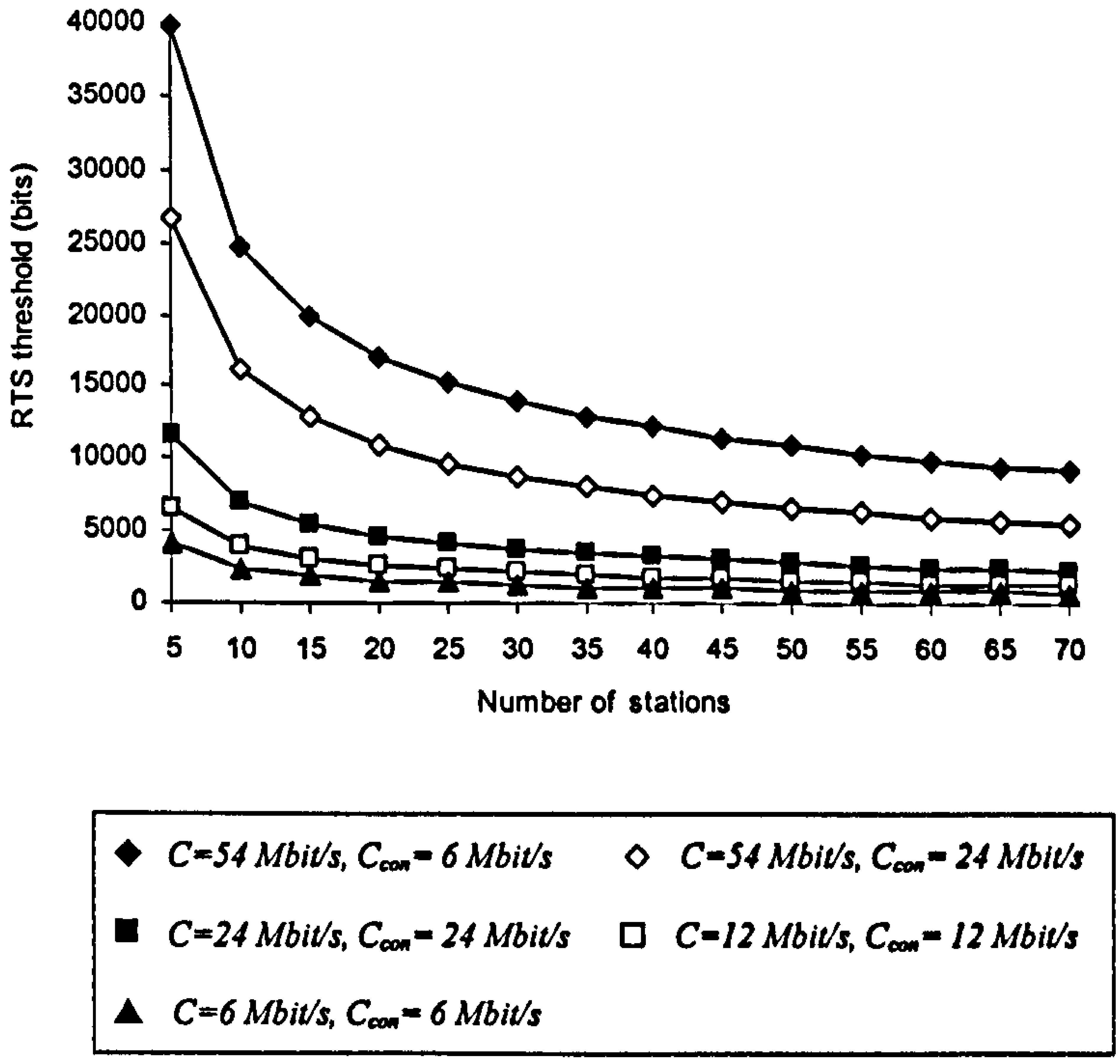


Figure 4.10 Effect of control and data rates on RTS threshold for IEEE 802.11a

Figure 4.10 studies the effect of high-speed rate on RTS threshold values in IEEE 802.11a by utilizing five different pairs of (data, control) rates; (54, 6) (54, 24) (24, 24) (12, 12) and (6, 6). When the link data and control rates are the same (6,6) , (12, 12) or (24, 24) the RTS threshold values are rather low since the RTS/CTS reservation scheme achieves better performance than the basic access for relatively small packet sizes. On the contrary, when the highest data rate of 54 Mbit/s is utilized combined with the lowest control rate of 6 Mbit/s, the RTS/CTS scheme is beneficial only for very large packet sizes (>12000 bits). Even in the best-case scenario for the highest possible control rate of 24 Mbit/s, the RTS/CTS scheme improves performance for large packet sizes (*n =20, l>1100 bits*) or (*n =70, l>5500 bits*).

Figure 4.11 plots RTS threshold versus retry limit and initial *CW* size for four different network sizes (*n*=5, 25, 50 and 70) and the highest data rate of 54 Mbit/s. The best-case scenario is considered where control packets are transmitted at the highest possible control rate (24 Mbit/s). As the figure clearly illustrates, in the case of IEEE 802.11a the same conclusions arise with the ones for 802.11b; 1) only large network sizes render the RTS/CTS beneficial for the overall performance compared to the basic access scheme; and 2) the RTS threshold increases when either for higher retry limit or initial *CW* size values due to the enhanced probability of a successful packet transmission.

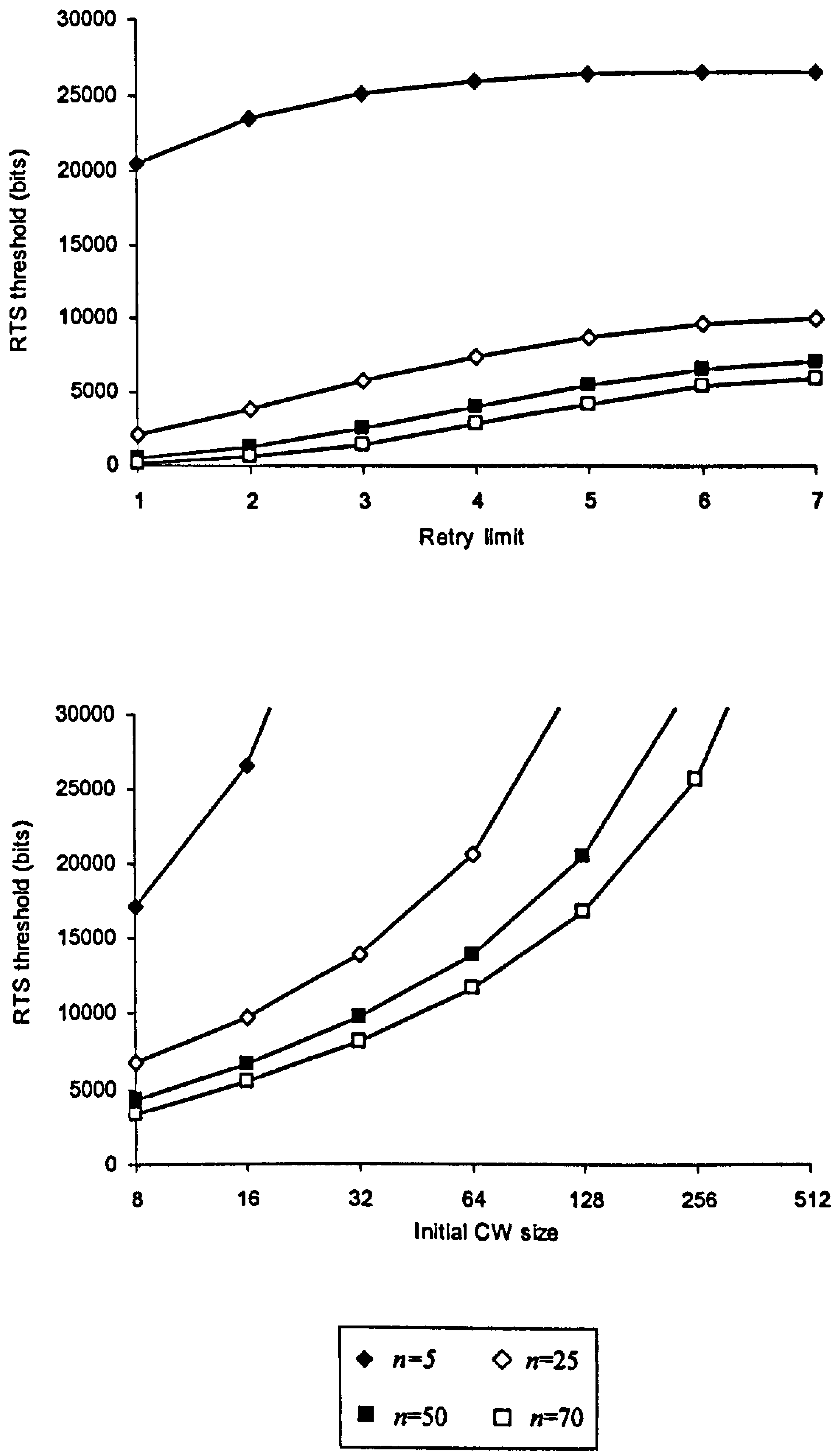


Figure 4.11 RTS threshold versus packet retry limit and initial contention window (CW) size (IEEE 802.11a, C= 54 Mbit/s, C_{con}= 24 Mbit/s)

4.3 Enhancing performance in congested environments by means of the Double Increment Double Decrement backoff scheme

This section proposes an effective and easy-to-implement backoff algorithm, DIDD (Double Increment Double Decrement). The main concept of DIDD is that after a successful packet transmission, the contention window for a new packet will not be reset to CW_{min} in order to avoid new collisions due to congested conditions. The presented performance results show that the proposed DIDD brings several benefits: 1) it obtains higher throughput than traditional DCF especially with large number of competing stations; 2) DIDD does not drop any packets at all; and 3) finally, DIDD is very easy to be implemented, since it does not need to estimate the number of competing stations or modify the packet structure or access procedures in 802.11 DCF.

As explained earlier in chapter 3, the backoff counter for every station depends on the collisions and on the successful packet transmissions experienced by a station in the past. The collision avoidance protocol procedures specify that before transmitting each station selects a random value for its backoff counter in the range [0, W-1]. If a collision encounters, then the protocol employs the exponential backoff i.e. the next backoff value will be selected in the range [0, (2W)-1] and so forth. We define for convenience $W = CW_{min}$. Let m be the 'maximum backoff stage' defined as $CW_{max} = 2^m W$. Since a station may be in stage $i \in [0,m]$, we adopt the following notation:

$$W_i = 2^i W, \quad i \in [0,m] \tag{4.11}$$

where i is defined as the backoff stage that identifies the number of retransmissions a packet has suffered in the past.

Let us denote with TX the event that a station is transmitting during a slot time and with $P(s=i \mid TX)$ the steady state probability that a transmitting station is found in stage i > 0. Since this probability is given by the probability that the station, in the previous transmission slot, was found in stage i − 1 and that the transmission failed (with probability p), it follows that $P(s=i \mid TX)$ can be calculated as:

$$P(s=i \mid TX) = c\left(\frac{p}{1-p}\right)^i \tag{4.12}$$

where c is a constant parameter that we will derive next and p is the probability that a transmission fails due to a collision, when at least one of the $n-1$ remaining stations transmit a packet in the same time slot. If we assume that all stations see the system at steady state and transmit with probability τ, the collision probability p is given by:

$$p = 1-(1-\tau)^{n-1} \tag{4.13}$$

For convenience in further calculations, we set:

$$a = \frac{p}{1-p} \tag{4.14}$$

Since a station is always found in the i stage, we have:

$$\sum_{i=0}^{m} P(s=i\,|\,TX) = 1 \tag{4.15}$$

Substituting (4.12) into (4.15), the value of the parameter c is calculated as:

$$c\sum_{i=0}^{m}\left(\frac{p}{1-p}\right)^{i} = 1$$

$$c\frac{1-\left(\frac{p}{1-p}\right)^{m+1}}{1-\left(\frac{p}{1-p}\right)} = 1$$

$$c = \frac{1-\frac{p}{1-p}}{1-\left(\frac{p}{1-p}\right)^{m+1}} = \frac{1-a}{1-a^{m+1}} \tag{4.16}$$

Using (4.14), (4.12) becomes:

$$P(s=i\,|\,TX) = \frac{1-a}{1-a^{m+1}}a^{i} \tag{4.17}$$

We are ultimately interested in the unconditional probability $\tau = P(TX)$ that the station transmits a packet in a randomly chosen slot. By utilizing Bayes' theorem:

$$P(s=i\,|\,TX) = \frac{P(TX\,|\,s=i)\,P(s=i)}{P(TX)} \tag{4.18}$$

which in turn yields, for all i values in $(0,..m)$:

$$P(TX)\frac{P(s=i\,|\,TX)}{P(TX\,|\,s=i)} = P(s=i) \tag{4.19}$$

This equality yields also for the summation:

$$P(TX)\sum_{i=0}^{m}\frac{P(s=i\,|\,TX)}{P(TX\,|\,s=i)}=\sum_{i=0}^{m}P(s=i)=1 \tag{4.20}$$

A reservation attempt occurs when the backoff counter of the transmitting station becomes equal to zero, regardless of the backoff stage. Thus, the transmission probability τ is equal to:

$$\tau=P(TX)=\frac{1}{\sum_{i=0}^{m}\frac{P(s=i\,|\,TX)}{P(TX\,|\,s=i)}} \tag{4.21}$$

It remains to find an expression for the conditional probability $P(TX\,|\,s=i)$. This probability can be calculated by dividing the average number of slots a station spends in the transmission state (i,0) while in stage i (exactly 1 slot according to the adopted time scale), and the average number of slots that a station spends in the backoff stage i which is equal to $W_i+1/2$. Since the average number of slot times spent for each backoff counter transition is exactly 1 slot, therefore:

$$P(TX\,|\,s=i)=\frac{1}{1+\frac{W_i-1}{2}}=\frac{2}{W_i+1} \tag{4.22}$$

Therefore, equation (4.21) becomes equal to:

$$\tau=\frac{2}{\frac{1-a}{1-a^{m+1}}\left(\sum_{i=0}^{m}(W_i+1)a^i\right)}$$

$$\tau=\frac{2}{\frac{1-a}{1-a^{m+1}}\left(\sum_{i=0}^{m}\left(2^i\cdot W a^i\right)+\sum_{i=0}^{m}a^i\right)}$$

$$\tau=\frac{2}{\frac{1-a}{1-a^{m+1}}\left(\sum_{i=0}^{m}\left((2a)^i\cdot W\right)+\sum_{i=0}^{m}a^i\right)}$$

$$\tau=\frac{2}{\frac{1-a}{1-a^{m+1}}\left(\frac{1-(2a)^{m+1}}{1-2a}\cdot W+\frac{1-a^{m+1}}{1-a}\right)}$$

$$\tau=\frac{2}{\frac{1-a}{1-a^{m+1}}\left(\frac{\left(1-(2a)^{m+1}\right)(1-a)W+(1-2a)\left(1-a^{m+1}\right)}{(1-2a)(1-a)}\right)}$$

Finally, the probability τ that a station transmits a packet in a randomly chosen slot time is given by[3]:

$$\tau = \frac{2(1-2a)(1-a^{m+1})}{(1-(2a)^{m+1})(1-a)W + (1-2a)(1-a^{m+1})} \tag{4.23}$$

Equations (3) and (14) represent a non-linear system with two unknowns p and τ. This system can be solved by utilizing numerical methods (it has a unique solution) and evaluating p and τ for a certain W and m combination. Note that $p \in [0,1]$ and $\tau \in [0,1]$.

A. *Packet delay analysis*

This section develops an elegant method to calculate the average packet delay for a successfully transmitted packet, based on the analysis derived in the previous section. The average delay $E[D]$ is defined to be the time interval from the time a packet is at the head of its MAC queue ready for transmission, until its successful reception. The average packet delay $E[D]$ can be found by:

$$E[D] = E[X]\,E[slot] \tag{4.24}$$

where $E[X]$ is the average number of time slots needed for a successful packet transmission and $E[slot]$ is the average length of a slot time. $E[X]$ is calculated as:

$$E[X] = \frac{1}{\tau(1-p)} = \frac{1}{\tau(1-\tau)^{n-1}} \tag{4.25}$$

Finally, if we substitute $E[slot]$ into equation (4.24), the average packet delay $E[D]$ can be simply calculated.

B. *Performance evaluation of DIDD backoff scheme*

The OPNET 802.11 simulator developed in Chapter 3 was appropriately modified in order to model the proposed DIDD backoff scheme. Once more, we consider a LAN of n stations operating at saturation conditions under ideal an error free medium and no hidden stations. Figure 4.12 shows the resulting throughput and packet delay obtained through the analytical model developed in the previous section and OPNET simulation outcome. Results are given for both the cases of basic access and RTS/CTS schemes for the DSSS physical layer. We can observe that analytical results are very consistent

[3] Note that the above expression for the probability τ is different to the one for the IEEE 802.11 exponential backoff algorithm. From equation (4.14), we observe that the transmission probability τ depends on the collision probability p.

with simulation outcome[4] and both analysis and simulation always show the same results with accuracy. Moreover, the figure illustrates that the RTS/CTS scheme achieves higher throughput and lower packet delay comparing to basic access, for the specific large packet size, due to the shorter collision duration.

Figure 4.13 illustrates the conditional collision probability p and the transmission probability τ as function of the number of stations. The probabilities are shown for both the cases of legacy DCF and DIDD. As expected, the larger the number of stations, the higher the collision probability for legacy DCF comparing to DIDD. In fact, DIDD can decrease the chance of a packet collision by utilizing a higher contention window after a successful transmission instead of resetting it to CW_{min}. Furthermore, more contending stations bring about the decrease of the transmission probability; for large network size scenarios, τ attains roughly the same values having a slight decreasing trend.

Figure 4.14 illustrates the DIDD throughput gain obtained with and without the use of the RTS/CTS mechanism for two different CW values (CW=16, 32). The gain without RTS/CTS is much higher than when RTS/CTS is used. This means that the DIDD scheme is more beneficial when the RTS/CTS is not utilized. The reason is that RTS/CTS reduces the collision time to a small value, which makes the use of DIDD less effective since the collision time is already small. Moreover, we can observe that the initial CW size and the number of stations strongly affect the throughput gain of DIDD. In particular, for small initial CW sizes (CW=16) as well as when the number of stations increases, DIDD gives significant improvements over the legacy DCF. For instance, under the basic access scheme, the percentage of improvement for CW=32 are 2% (n=10), 8% (n=25), 15% (n=25), and 20% (n=70). In the case of CW=16, performance is enhanced even more and the improvements are 6% (n=10), 15% (n=25), 27% (n=25), and 36% (n=70).

Figure 4.15 depicts packet delay and packet drop probability values for DIDD and legacy DCF schemes. As it is illustrated in figure 4.14, the main advantage of the proposed DIDD backoff scheme (apart from the throughput improvement) is that we don't have any packet drops. Under DIDD, every packet is being retransmitted until its successful transmission but with a decreased collision probability compared to the legacy DCF (as it has been shown in figure 4.13). DCF causes many packet drops, especially when there are many competing nodes. On the other hand, DIDD attains

[4] Note that simulation results are acquired with a 95% confidence interval lower than 0.002.

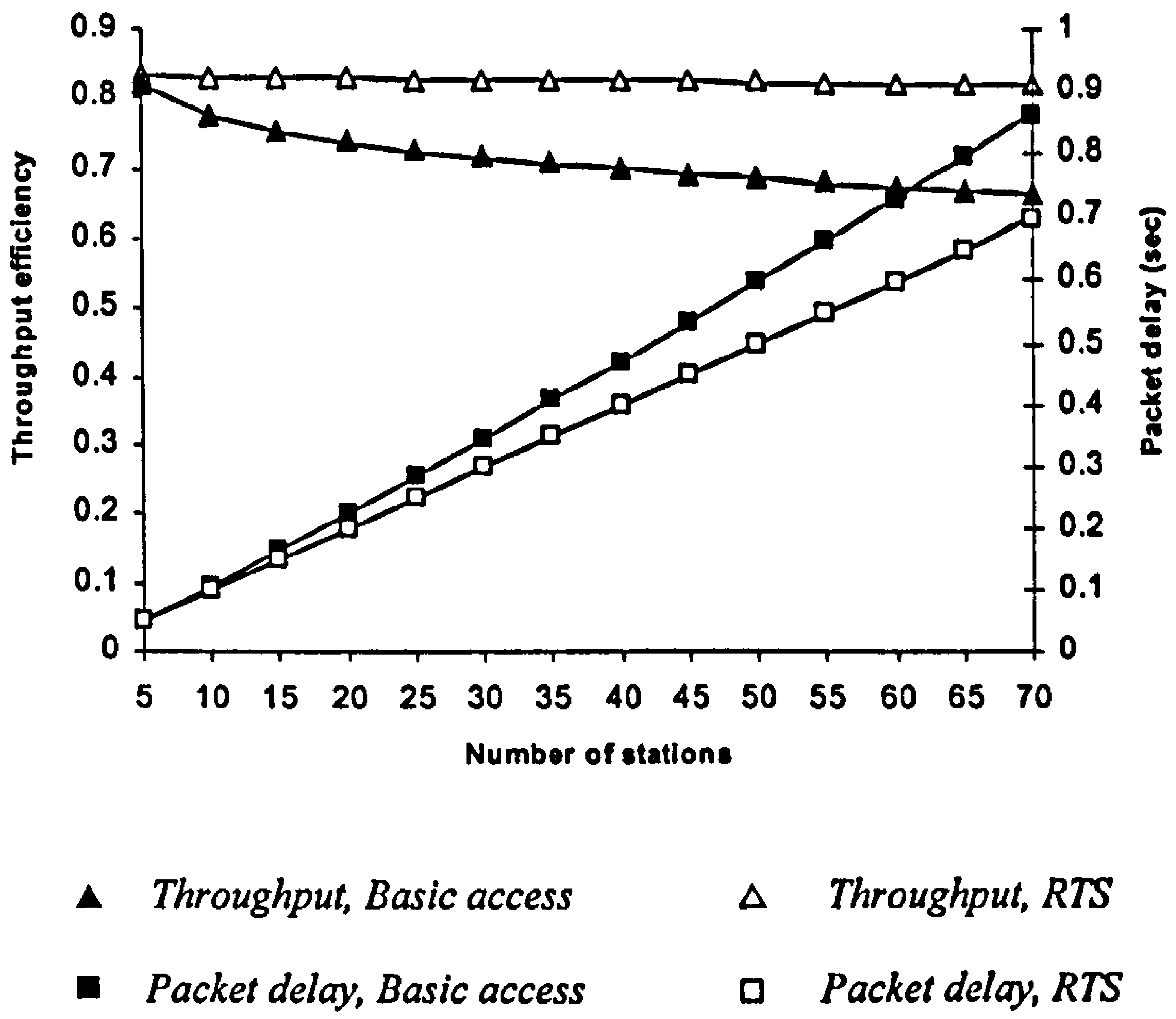


Figure 4.12 Throughput efficiency and packet delay for basic access and RTS/CTS schemes Analysis (lines) versus OPNET simulation(symbols)

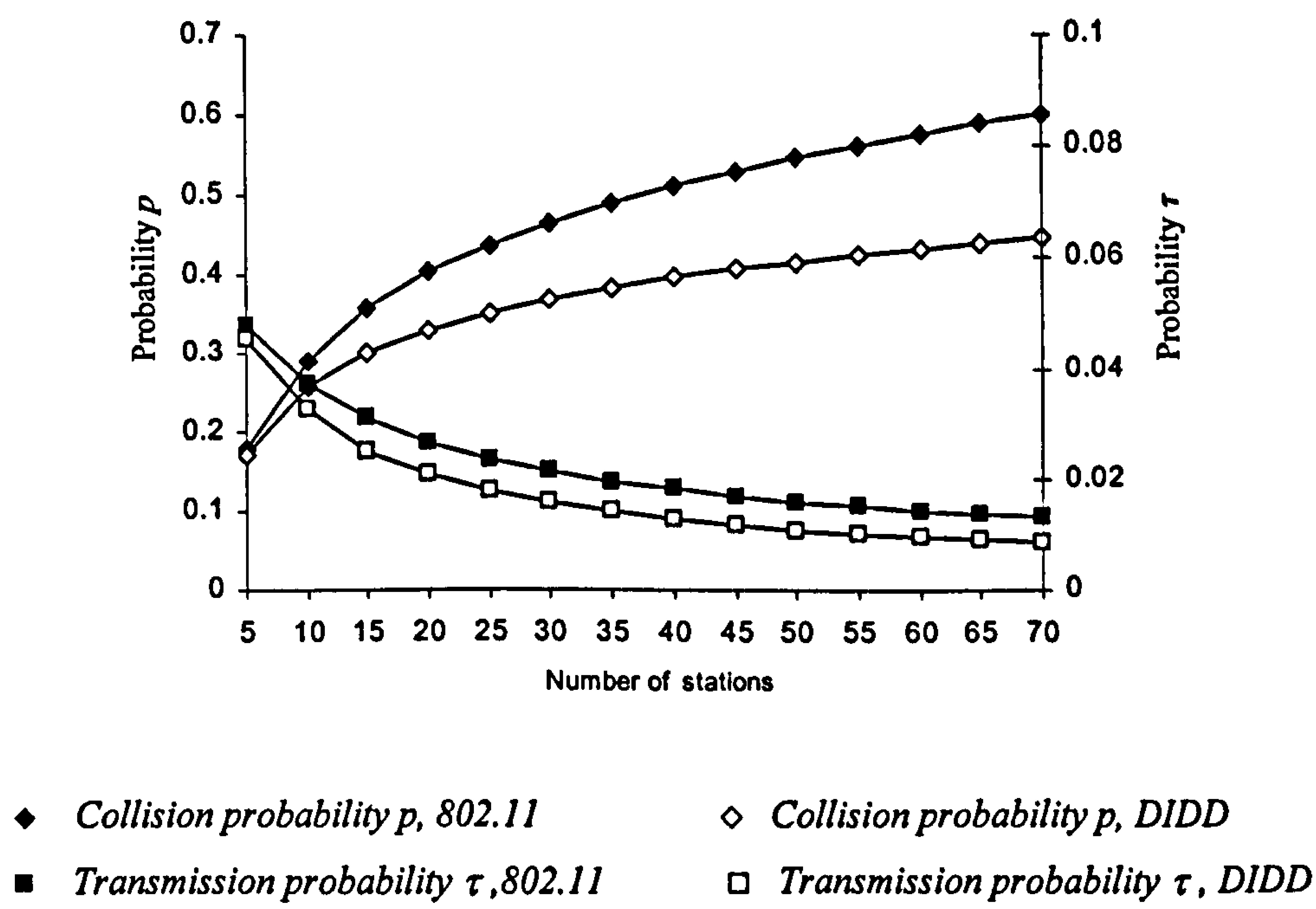
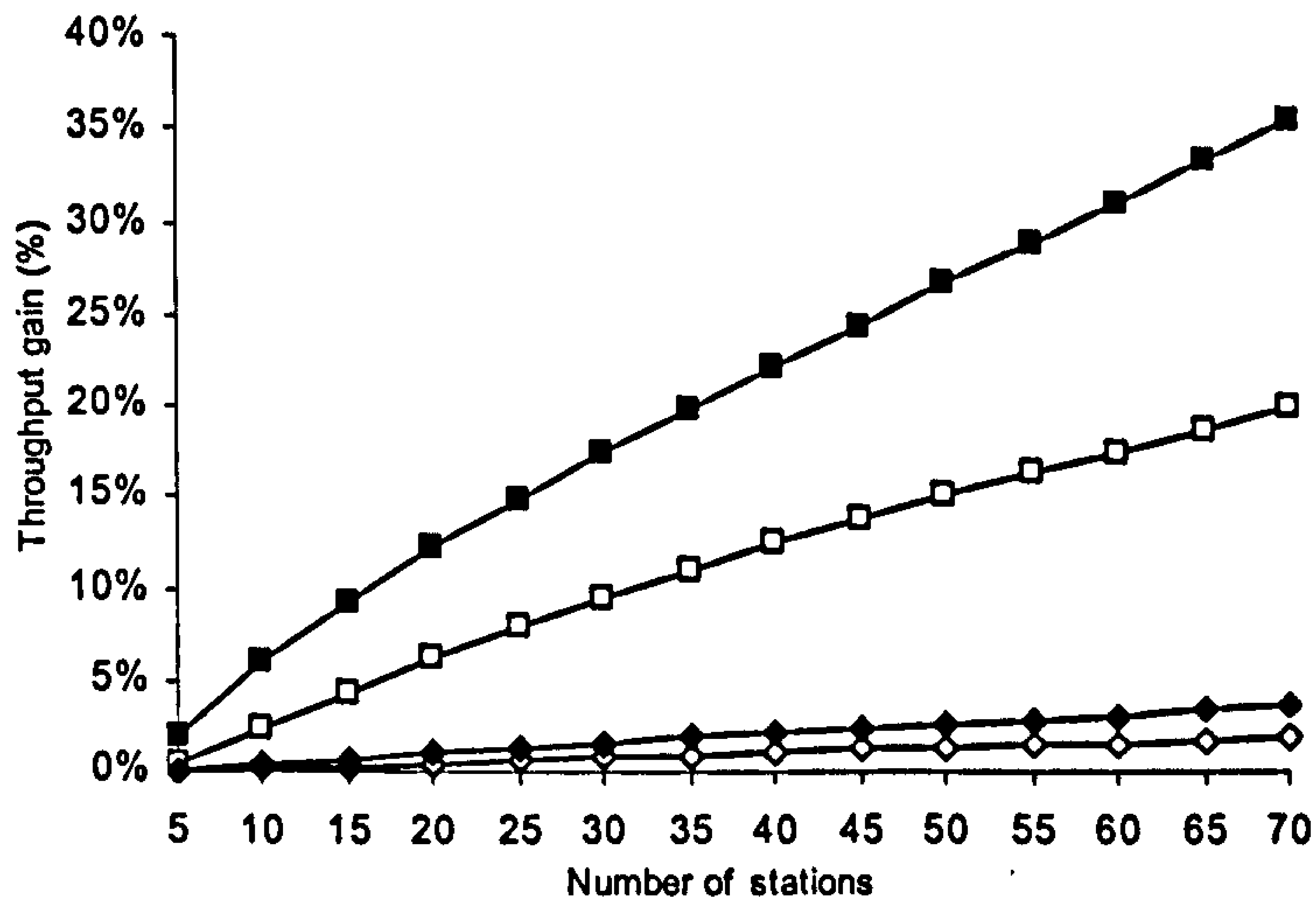


Figure 4.13 Collision and transmission probabilities versus n

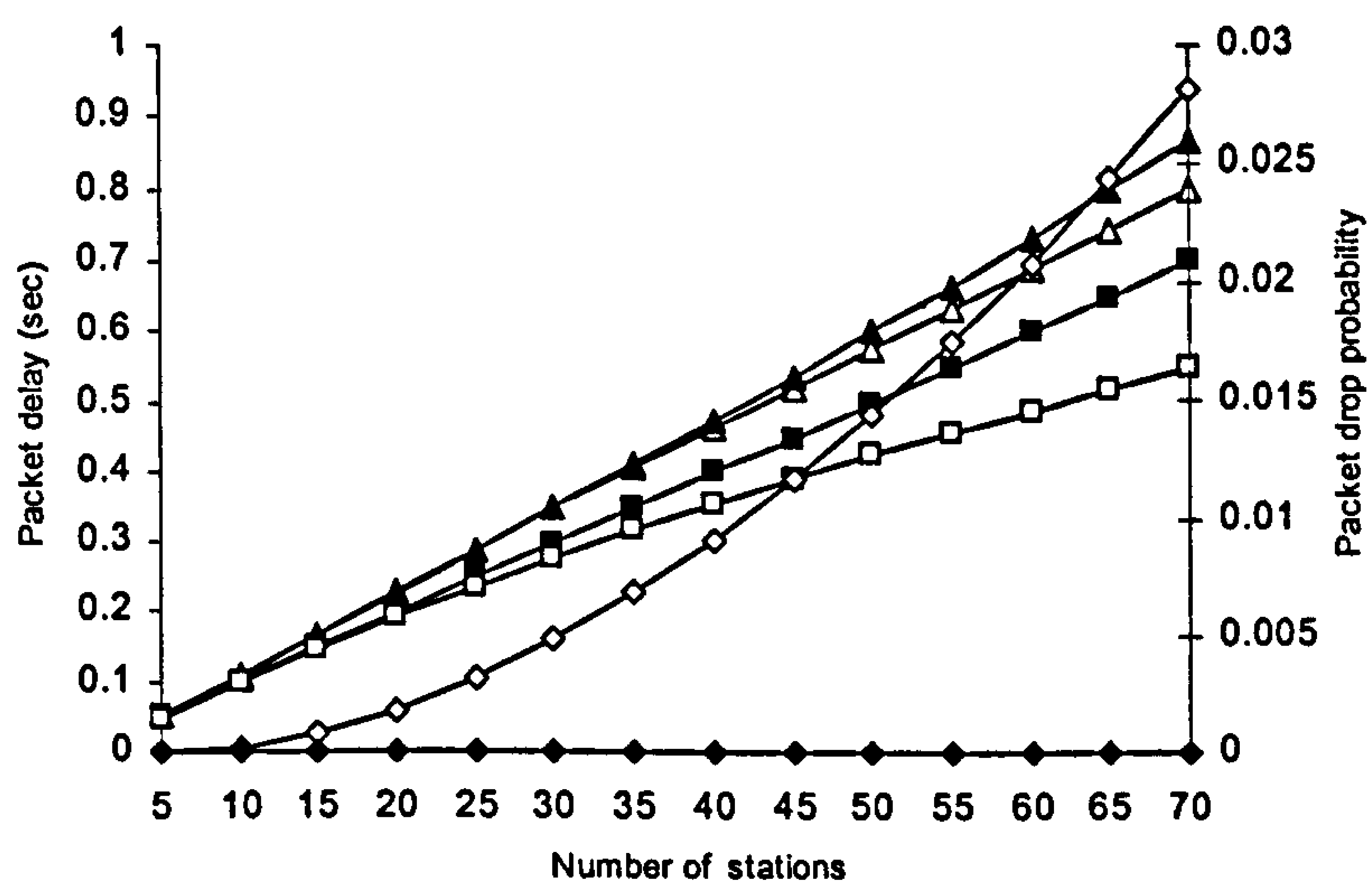
higher packet delay values comparing to the legacy DCF since it includes the time delay of packets that would have been discarded using the legacy DCF. This is the small price we pay in order to have higher throughput performance and not dropped packets at all.



◆ *RTS/CTS, W=16* ◇ *RTS/CTS, W=32*

■ *Basic, W=16* □ *Basic, W=32*

Figure 4.14 Throughput gain (in %) versus n



▲ *Packet delay, DIDD, Basic* △ *Packet delay, 802.11, Basic*

■ *Packet delay, DIDD, RTS/CTS* □ *Packet delay, 802.11, RTS/CTS*

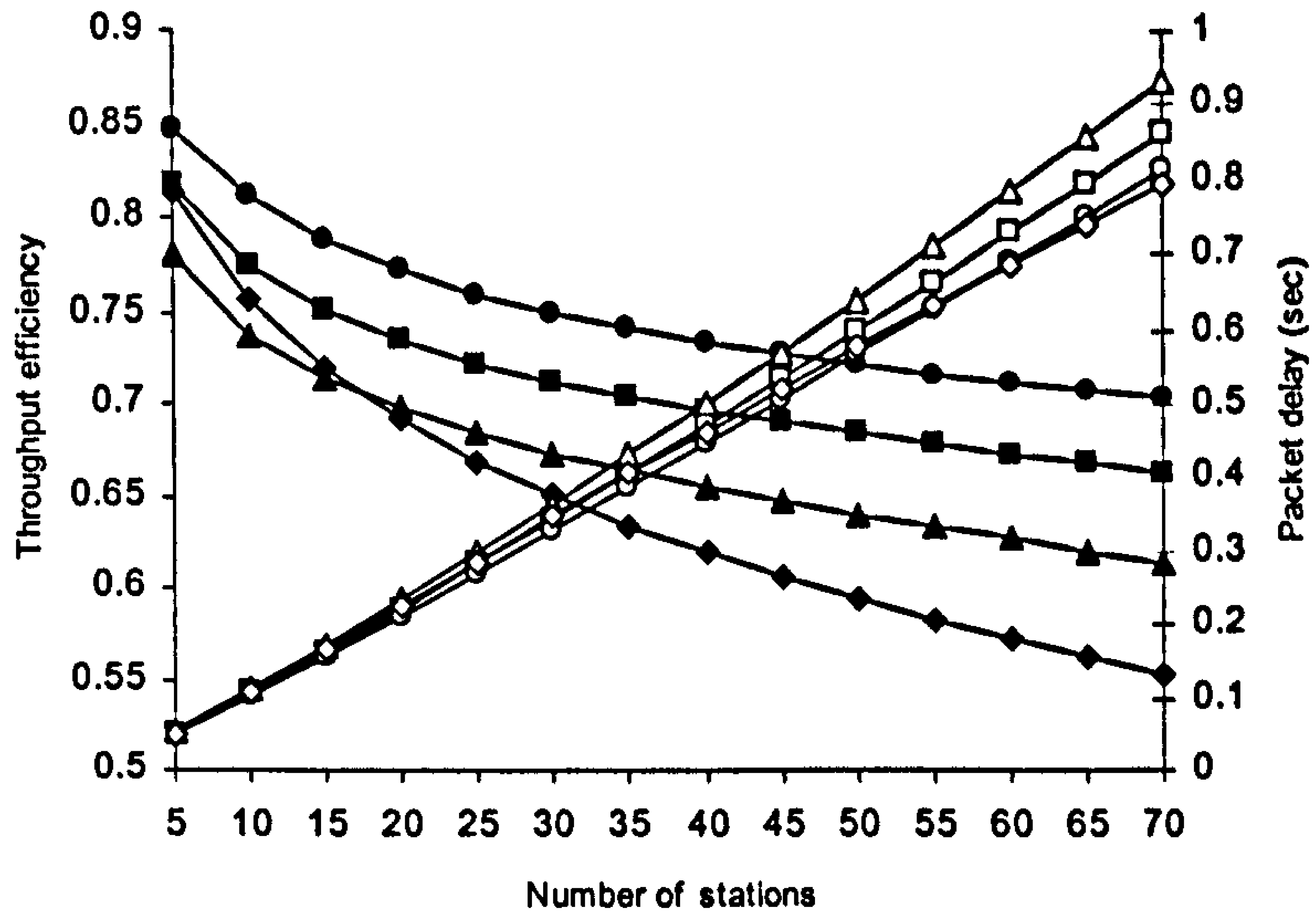
◆ *Packet drop probability, DIDD* ◇ *Packet drop probability, 802.11*

Figure 4.15 Packet delay and packet drop probability versus n

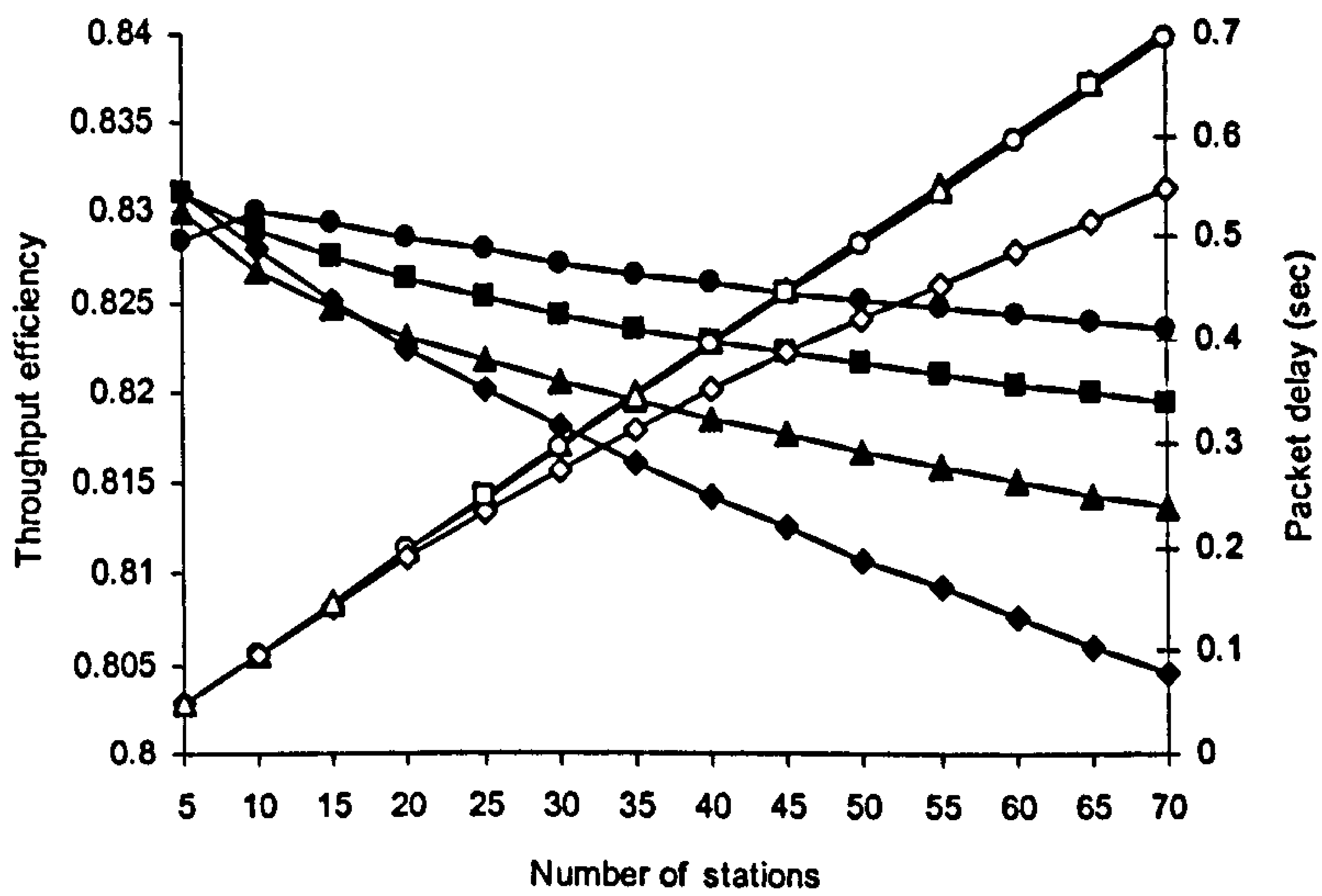
Since the DIDD scheme introduces a different backoff scheme for contention, it is interesting to study how performance is affected by various initial *CW* sizes. Figure 4.16 compares the performance of DIDD with the standard backoff algorithm utilized in legacy DCF with different *CW* values (*CW*=16, 32 and 64) and for both medium access schemes. It is not difficult to conclude that DIDD always achieves a higher throughput since it can decrease the chance of a packet collision by utilizing a higher contention window. Moreover, the basic access mode attains a higher throughput when we choose a larger initial *CW* size. However, for the RTS-CTS access mode, the throughput is less improved even with a larger initial *CW* size. This can be explained by the fact that a larger *CW* window size can decrease the probability of collision and the number of retransmission for the basic access mode. In contrast, the RTS/CTS access mode by itself can avoid long collision and the associated waste of the bandwidth when packet collisions occur.

In Figure 4.17, we examine the throughput and packet delay performance of different backoff parameters (*CW* and *m'*) on both basic access and RTS/CTS schemes. Five different combinations are studied; (*CW*, *m'*) = (32, 3) (32, 5) (32, 7) (64, 3) for DIDD and (32, 5) for legacy DCF. From the figure it can be seen that: 1) DIDD performs better in throughput than legacy DCF for any pair of (*CW*, *m'*); 2) the throughput performance gain obtained by DIDD is more apparent when the number stations is large and under basic access; 3) legacy DCF achieves the lowest packet delay values comparing to any combination of backoff parameters used in DIDD; 4) DIDD packet delay performance under RTS/CTS will be kept at a certain level (for example, the four curves are nearly overlapped); 5) the worst packet delay performance, especially for large network sizes, is for the case of (32, 3) due to the resulting low *CW* size and high collision probability; and 6) by utilizing *CW*=32 and *m'*=7, further throughput improvement is obtained when the number of stations is large. Considering the trade-off between performance decrease under very small network sizes and performance improvement under large network sizes, (*CW*, *m'*) = (32, 7) appears to be the best choice to choose in practical deployment if the number of competing stations cannot be known.

Figure 4.18 plots throughput efficiency and packet delay versus network size for three data rates (C = 2, 5.5 and 11 Mbit/s) using the short PHY packet overhead (preamble and header) defined in the IEEE 802.11b standard. When data rate increases, throughput efficiency decreases as explained in chapter 4; packet delay also decreases since the transmission time of data packets is reduced. When the basic access scheme is employed, we clearly see that throughput performance considerably decreases when the



(a) Basic access



(b) RTS/CTS

● *Throughput, DIDD, W=64* ■ *Throughput, DIDD, W=32*

▲ *Throughput, DIDD, W=16* ◆ *Throughput, 802.11, W=32*

○ *Packet delay, DIDD, W=64* □ *Packet delay, DIDD, W=32*

△ *Packet delay, DIDD, W=16* ◇ *Packet delay, 802.11, W=32*

Figure 4.16 Throughput efficiency and packet delay for various CW sizes

number of stations increases (more packet collisions) and that DIDD achieves a much higher throughput than that of the legacy DCF. For the RTS/CTS mechanism, throughput performance of both the DIDD and legacy DCF schemes is not significantly sensitive to the number of the competing stations for any data rate. At the same time, DIDD achieves slightly higher throughput but considerably higher packet delay than

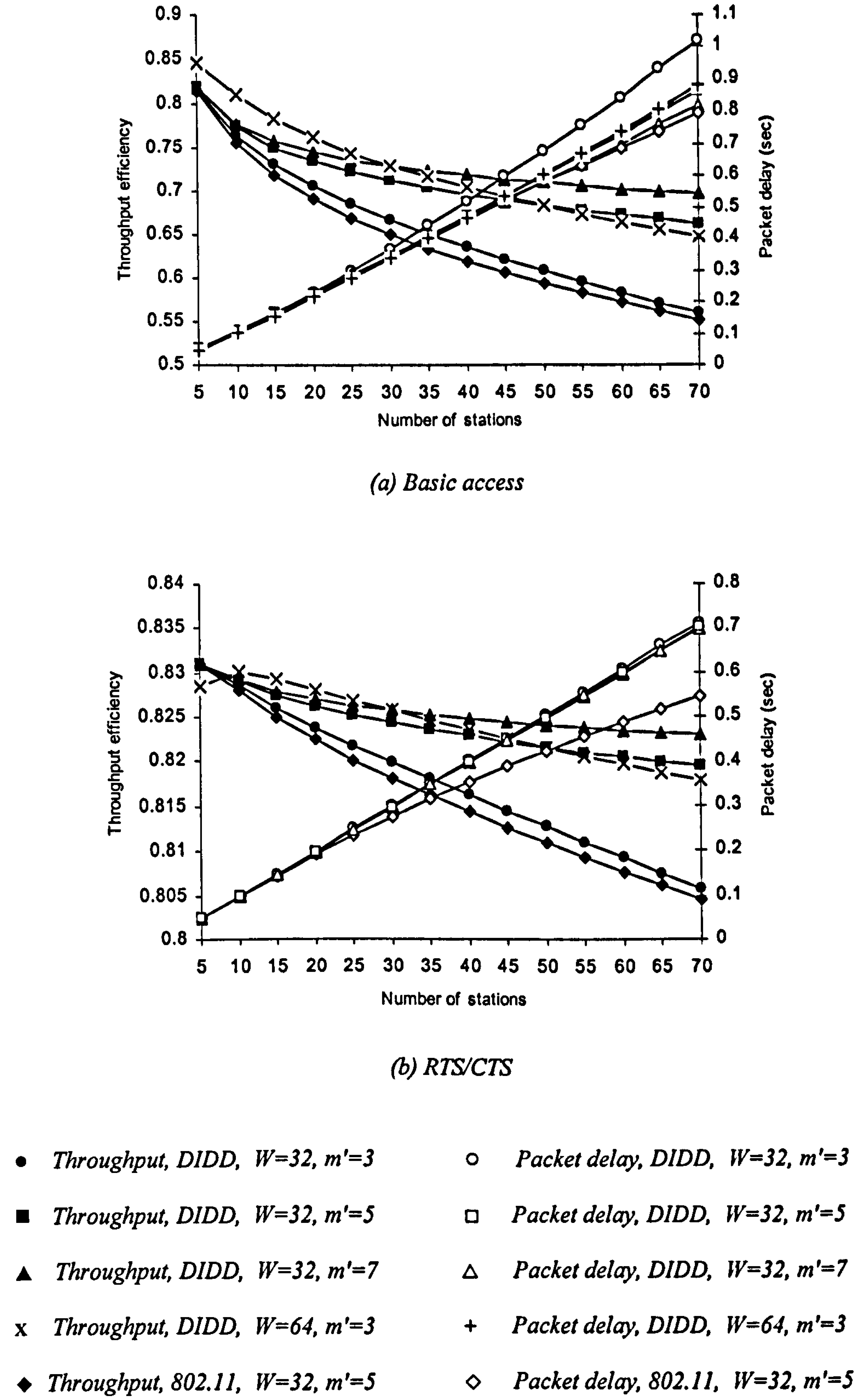
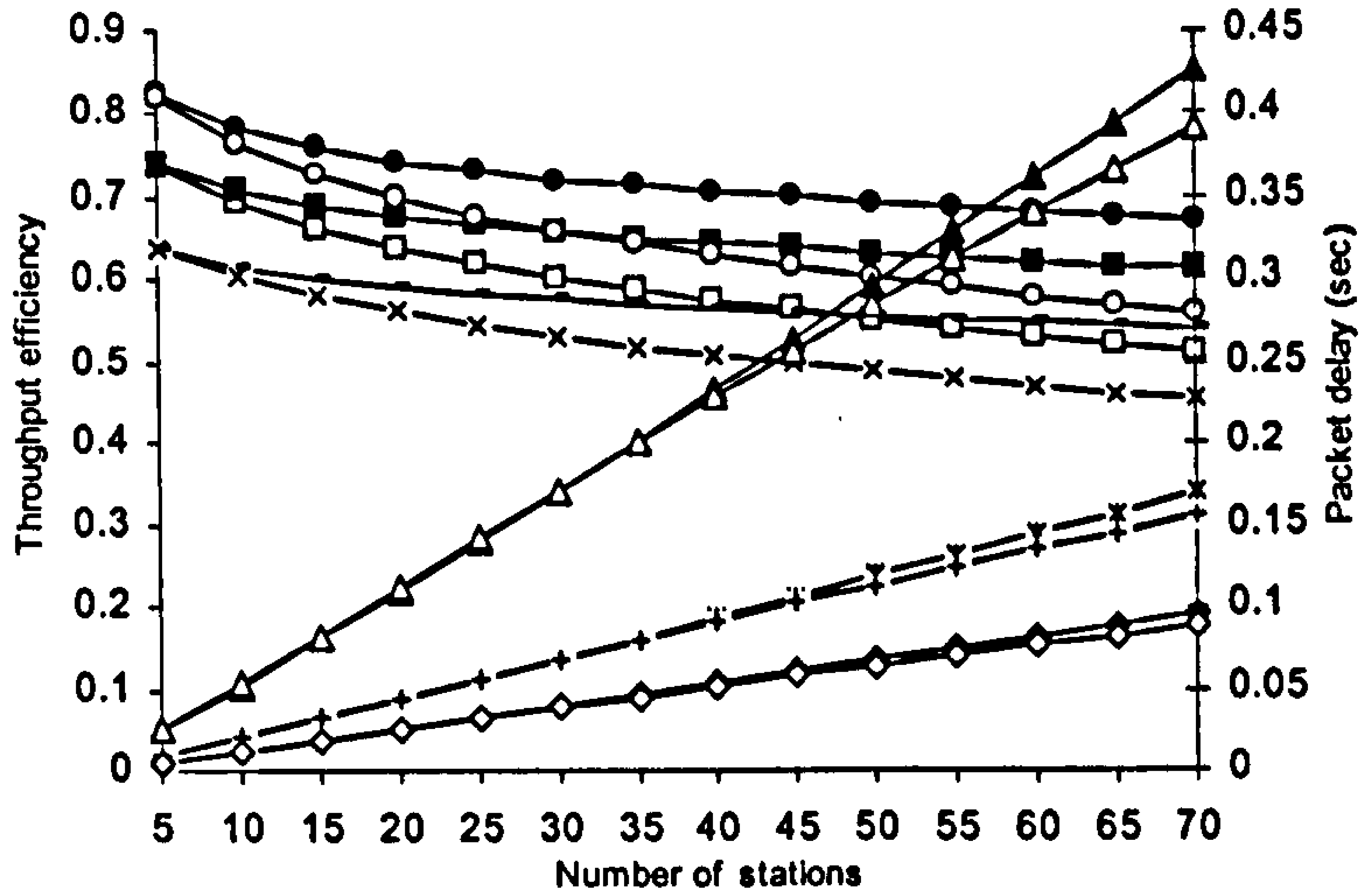
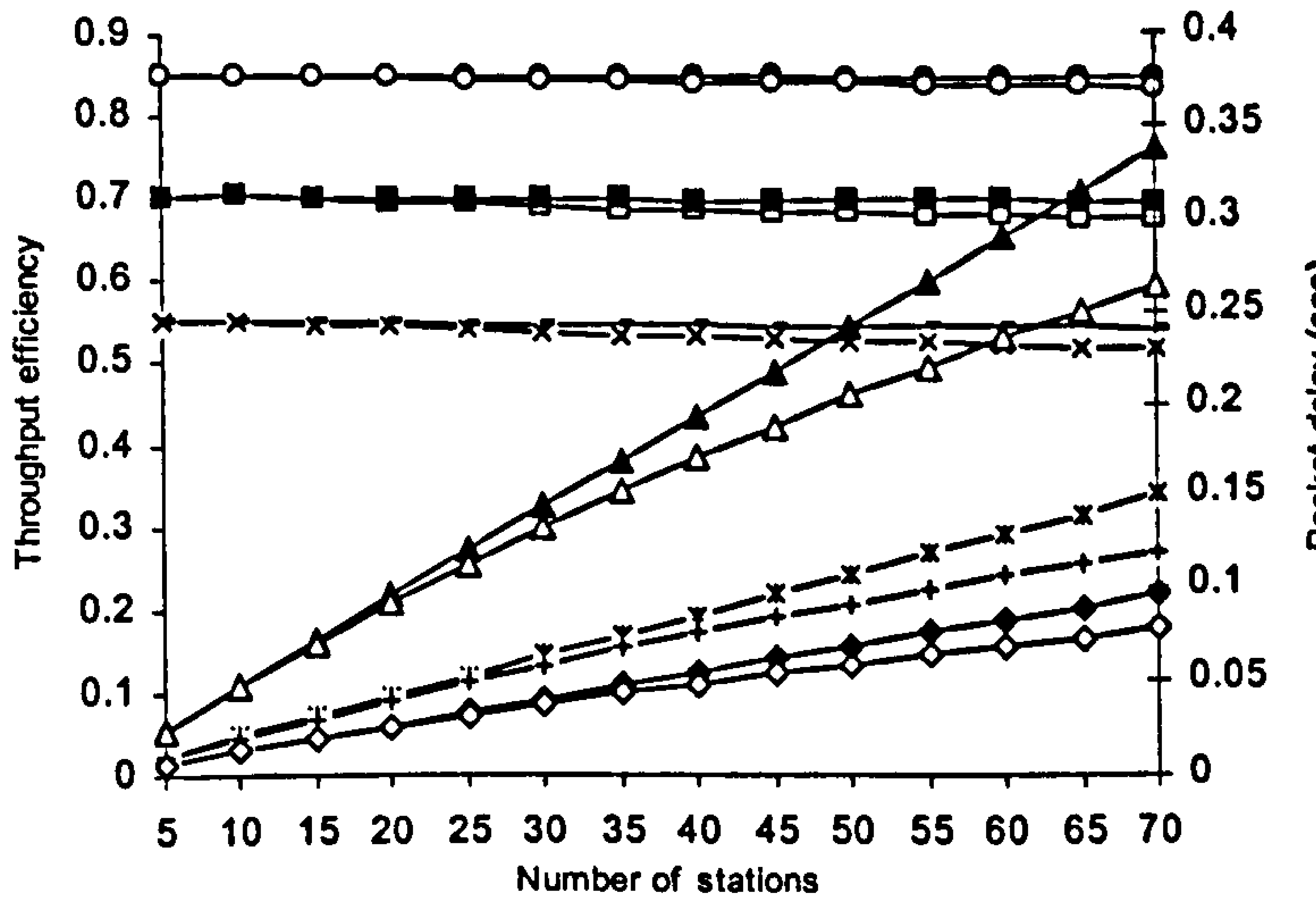


Figure 4.17 Throughput efficiency and packet delay for various CW sizes and backoff stages

DCF, indicating that DIDD is not the best choice under RTS/CTS. This can be explained due to the fact that the RTS/CTS scheme reduces collision duration and itself minimizes the negative impact of packet collisions.



(a) Basic access



(b) RTS/CTS

Throughput

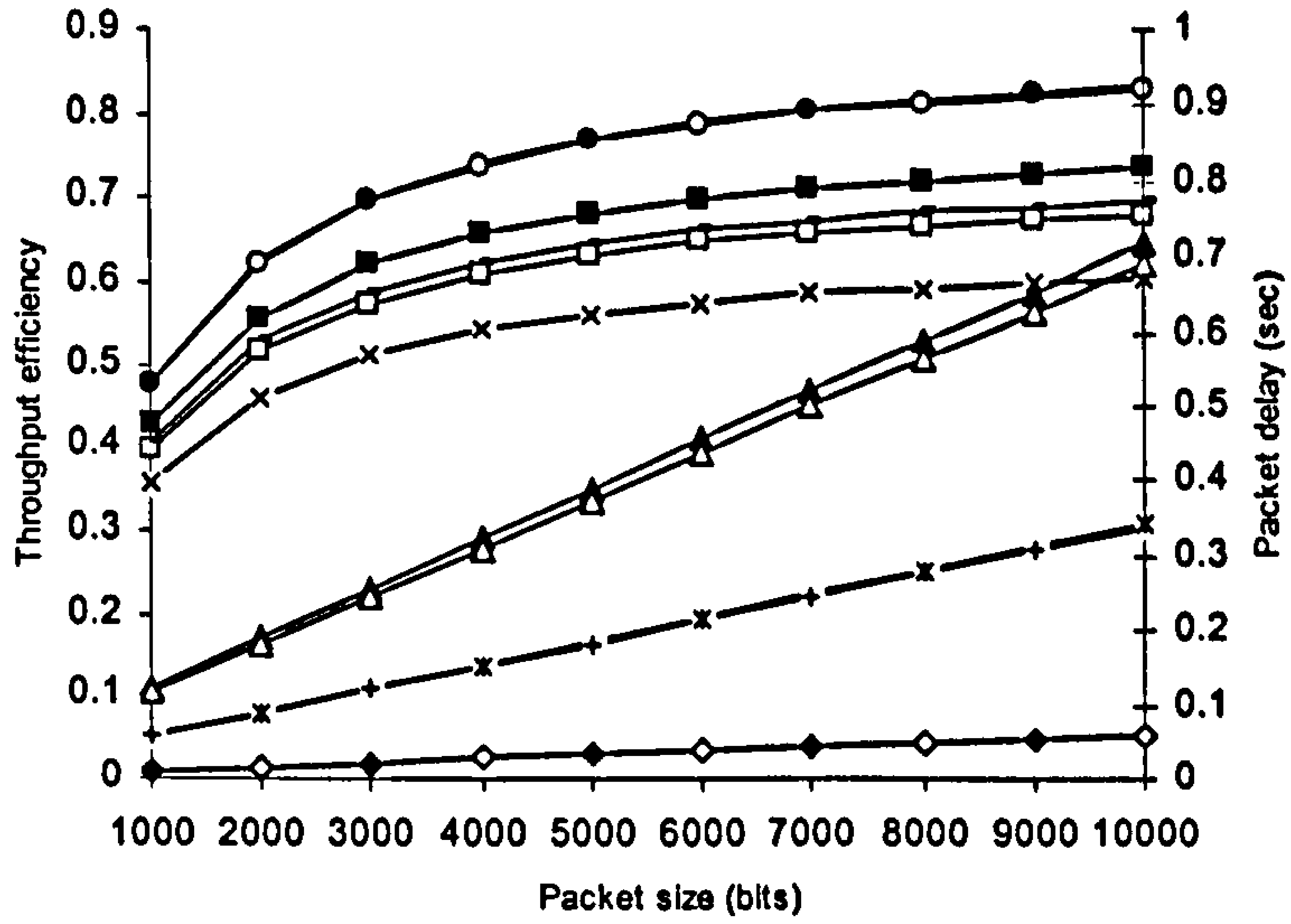
• *DIDD, C=2 Mbit/s*	■ *DIDD, C=5.5 Mbit/s*	- *DIDD, C=11 Mbit/s*
o *802.11, C=2 Mbit/s*	□ *802.11, C=5.5 Mbit/s*	x *802.11, C=11 Mbit/s*

Packet delay

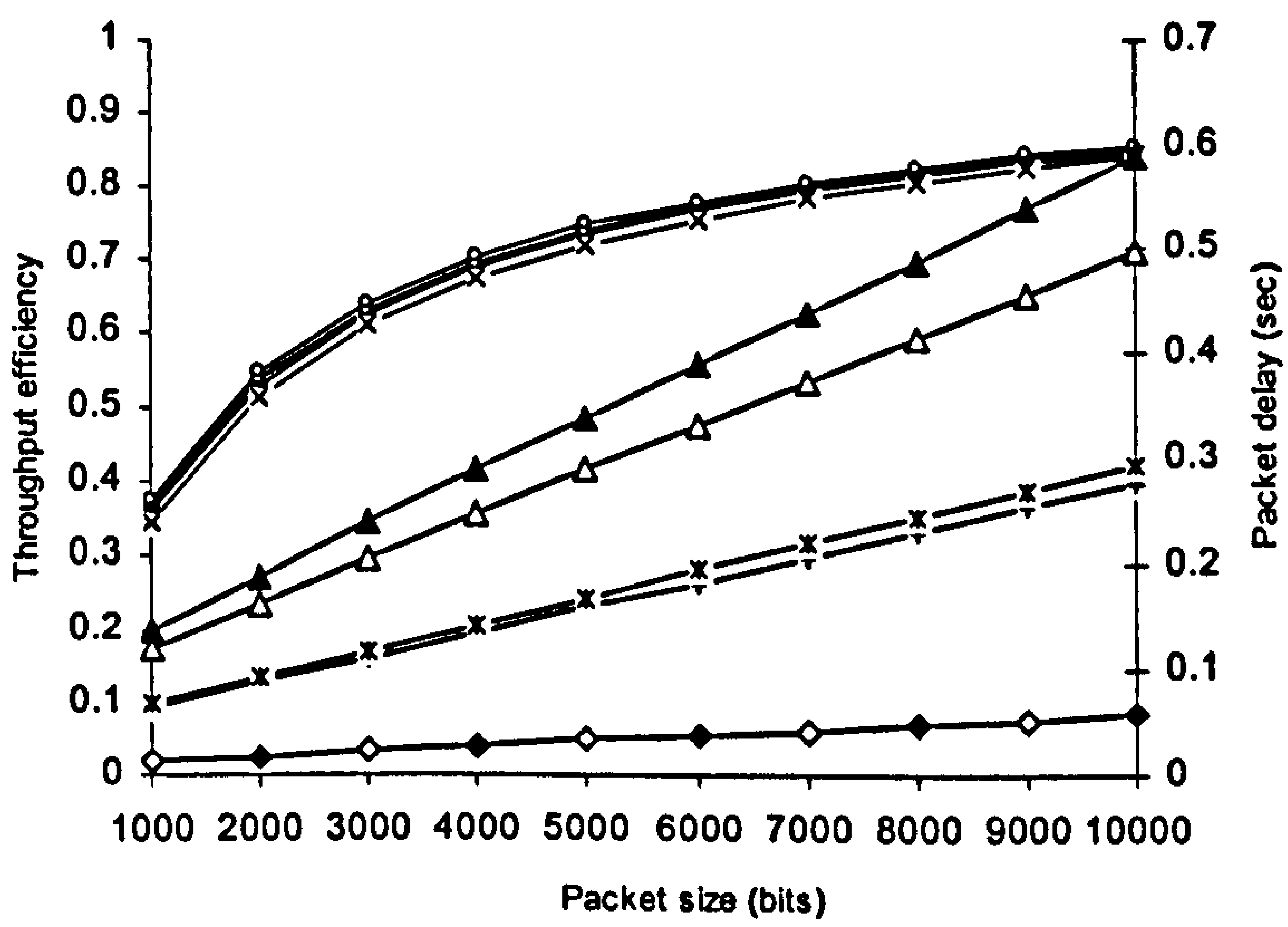
▲ *DIDD, C=2 Mbit/s*	* *DIDD, C=5.5 Mbit/s*	◆ *DIDD, C=11 Mbit/s*
△ *802.11, C=2 Mbit/s*	+ *802.11, C=5.5 Mbit/s*	◇ *802.11, C=11 Mbit/s*

Figure 4.18 Throughput efficiency and packet delay for different data rates

In figure 4.19, we can be easily observe the influence on throughput and packet delay performance resulted from certain factors; medium access mode, packet length and number of stations, in both DIDD and legacy DCF. Firstly, for both DIDD and



(a) Basic access



(b) RTS/CTS

Throughput

● *DIDD, n=5*	■ *DIDD, n=25*	- *DIDD, n=50*
o *802.11, n=5*	□ *802.11, n=25*	x *802.11, n=50*

Packet delay

◆ *DIDD, n=5*	* *DIDD, n=25*	▲ *DIDD, n=50*
◇ *802.11, n=5*	+ *802.11, n=25*	△ *802.11, n=50*

Figure 4.19 Throughput efficiency and packet delay versus packet size

legacy DCF, the RTS/CTS access mode and/or large packet size will bring higher throughput. DIDD can obtain improved throughput performance for both access modes, but the improved performance under basic access is much larger. Under the basic access mode, the improvement in performance as packet length increases. The reason is that because of the effect resulting from lowered collision probability for longer packet sizes. On the contrary, under the RTS/CTS scheme, DIDD marginally enhances throughput performance no matter the packet size values. This is justified since the RTS/CTS reservation scheme by itself can avoid long collision duration and the associated cost on performance when a packet collision occurs. We also observe that the smaller the packet size and the less congested the network are, the lower the packet delay is. Moreover, when the RTS/CTS scheme is utilized, the employment of DIDD, instead of the legacy DCF, causes a considerable increase on packet delay indicating the disadvantage of DIDD under the RTS/CTS case.

CHAPTER 5

Error consideration in IEEE 802.11 and DIDD protocols

In the previous chapters 3 and 4, the performance of the IEEE 802.11 and DIDD protocols was evaluated for error free conditions; the probability of a packet being in error was assumed always zero. Only empty (idle) slots, packet collisions, control packets (i.e. RTS, CTS and ACK) and packet overheads could result in performance degradation. In this chapter, the error free results will be re-examined in the light of realistic link error rate conditions due to fading and/or noise[1].

When the channel is error-prone, performance degradation can be either due to packet collisions or transmission errors. Unsuccessful transmission occurs when more than one user simultaneously transmits packets that collide with each other or unsatisfactory channel conditions corrupt the packet at the receiver even if the packet contends successfully. For both the cases, the behavior of the sender will be the same when either a corrupted packet is received at the receiver or a packet collision occurs; the sender will not be able to receive acknowledgement from the receiver, thus it will backoff and retransmit the packet until its retry limit is reached (the specific packet is discarded) or the packet is transmitted successfully.

A general observation is that often packet loss rates and mean bit error rates are time-variable and follow a bursty behavior. The most popular burst model widely used in the literature is the two-state Gilbert-Elliot model [39][44]. Although, several other burst error models were proposed in the literature [42][130], most of them only gained limited popularity. These alternative models often are hard to parameterize or need an extremely high number of parameters. On the other hand, the Gilbert-Elliot model comprises simplicity, quality in good predictions of performance parameters and parsimonious parameterization.

This chapter is outlined as follows. Section 5.1 provides information about the origin, the effects as well as the variability of transmission errors. Section 5.2 analyses the nature of errors and categorizes them to independent with fixed Bit Error Rate (BER) and time-variable burst errors modelled by the two-state Gilbert-Elliot Markov chain model[2]. Sections 5.3 provides a mathematical analysis for the IEEE 802.11 and DIDD protocols that takes into account independent transmission errors. Section 5.4 studies

[1] Further details about the nature of bit errors are provided in the next section that follows.

[2] Our approach by utilizing the Gilbert-Elliot model is to produce a realistic and accurate burst model that in the same time is not too difficult to implement.

the throughput and packet delay performance of both protocols under an error-prone environment and for the case of independent errors by making use of the previously derived analysis. Section 5.5 models in detail the commonly known Gilbert-Elliot model in order to capture the behavior and model the nature of burst channel errors. Finally, section 5.6 presents performance results in the case of burst errors and discusses various factors that degrade throughput and increase packet delay for both IEEE 802.11 and DIDD protocols.

5.1 Origin, effects and variability of transmission errors

Error characteristics of the multiple access wireless channels differ significantly from that of the wired medium. The distinction between the wired and wireless channels arises for many reasons. Packet losses on the wired medium are very rare and random in nature. In contrast, wireless channels are more prone to bit errors than wired channels. Moreover, the errors on the wireless medium are either random or bursty and the wireless channel is distinct and often time varying for each wireless user. As users move the received signal strength varies significantly, with each user depending on its location with respect to other users or the base station. Generally, packet errors usually occur due to non-ideal channel conditions. Partition loss in the building and multi-path fading, combined with ambient noise, decrease SNR (Signal-to-Noise Ratio) and, therefore, cause packet errors. In addition, there are effects due to fading interference from other users (adjacent channel interferences) and shadowing from objects all of which also cause packet errors and, thus, degrade the channel performance. Wireless device variability is another source of packet errors. Different devices have different output power, receive sensitivity and firmware, which may incur packet errors. If the channel condition varies over time, packet errors would have the corresponding variability. To observe such variability with experimental evidence in real wireless networks, active measurements on wireless traffic were conducted in [36].

It is extremely challenging for wireless network protocol developers to consider the large number of factors that affect the error performance of wireless channels. The error performance of wireless channels is usually modelled by capturing the statistical nature of the interaction among reflected radio waves. The statistical calculation for Bit Error Rate (BER), which is generally used to characterize channel errors at the physical layer, is a well known practice. The *BER* for a communication channel is defined as the probability of a single bit being corrupted in a defined number of transmitted bits. For example, a *BER* of 10^{-4} would mean that, on average, 1 bit in every 10,000 bits would be corrupted.

From the perspective of higher layers, network protocol developers and algorithm designers are interested in packet errors (block errors), since most of the higher-layer applications (running on top of link layers) exchange blocks of data between peers. The Packet Error Rate (PER) is determined by the *BER* as well as the packet payload and header length. *BER* is dependent on the SNR[3], the modulation and coding scheme. For example, bit errors in a link-layer packet may result in the loss of the entire packet; a single packet loss within a message may lead to the loss of the entire message. Packet errors and losses can also have an adverse affect on the perceived quality of wireless communications especially in real-time multimedia transmissions [76]. Classical data applications provide reliability by attempting to recover the corrupted/lost packets through retransmissions. Real-time requirements of emerging delay-sensitive multimedia applications (e.g., internet telephony, video conferencing, multicast audio/video etc.) necessitate a retransmission-less infrastructure (to avoid low-latency and/or implosion of feedback messages). Meanwhile, one positive aspect of such applications (especially those delivering streaming media) is their inherent tolerance to a certain level of errors and losses in the multimedia content. Therefore, it is desirable to have accurate packet-level error models for wireless channels, which can be used by network protocol developers and network system engineers to simulate and analyze the end-to-end performance at the packet level. Design and implementation of such applications stipulates a thorough understanding of the error and loss patterns encountered over the network.

5.2 Characterization and nature of bit errors

A significant amount of previous work [39][44][79][130] has shown that the simple knowledge of the average error rates (independent errors) may not be sufficient to appropriately characterize the error process since over a fading channel, long runs of errors, so-called burst errors, can occur even at a low channel error rate. For example, it has been shown that for the same average error rate, different degrees of correlation between errors correspond to (sometimes considerably) different performance [35]. In this case, we have to consider not only the channel error rate, which describes the

[3] SNR is determined by the path loss and the channel conditions. In wireless communication networks, path loss, fading and interference cause variations in the received SNR, which influences the BER.

average channel condition, but also the length of the burst errors. In fact, modeling the exact structure of burst errors in real digital communication channels is a complex problem. In general, there are no set procedures nor are there exact parameters that can be used to accurately predict the occurrence of such clusters of errors. As discussed previously, data packets are rendered useless if one or more data bits that make up a particular frame change state. Various burst error characterization models have been proposed in the literature [42][79][88][130], including the class of Hidden Markov Models [41] and [116] but their main disadvantage is the implementation complexity resulting in difficulties in easy adoption.

Roughly speaking, there is a tradeoff between the model complexity (as measured in number of parameters) and the models accuracy in matching certain statistics, as they are desired by the model's user or found in traces. Several measurements (e.g. those reported in [133] and [134]) have indicated that the wireless channel often exhibits a quite complex error behavior, often with bursty errors and variations over several timescales. To deal with a complicated channel model, it is sometimes possible to use a less complex one that still reflects the essential (for that particular study) properties of the complicated model. In fact, the use of an approximate model to estimate error performance allows the complex statistics of errors to be reduced to a manageable set of parameters.

The time-varying error characteristic of the wireless channel is been often modelled with a two-state Gilbert-Elliot model where each state represents a Binary Symmetric Channel. Each state is assigned a specific constant BER; in the "*GOOD*" state (G) errors occur with a low probability, while in the "*BAD*" state (B) they happen with a much higher probability. The transition rates between the "*GOOD*" and "*BAD*" states can be chosen according to the statistics of the actual channel being modelled, where the average amount of time spent in the "*BAD*" state equals the average duration of a fade and the average amount of time spent in the "*GOOD*" state equals the average amount of time between fades. As it has been demonstrated in [149] and [150], burst errors in wireless fading channels can be accurately approximated by the Gilbert-Elliot model which is actually widely-used for modeling wireless channels due to its simplicity as well as due to its ability to capture bursty error behavior for both slow and fast fading [150].

5.3 Independent transmission errors

5.3.1 Mathematical analysis of an error-prone channel for IEEE 802.11 protocol

A. Analysis assumptions and parameter definitions

This section presents a mathematical analysis in order to evaluate the impact of an error-prone channel on unsuccessful transmission probability and its impact on the overall performance. By taking into account transmission errors, the derived analytical model will be even more accurate. As we have seen before, in IEEE 802.11 protocol when a packet is corrupted at the receiver, it will be discarded and the sender will try to retransmit the packet. This actually means that, under error prone environments, even when a station contends successfully, it might still need to retransmit since its packet might be corrupted at the receiver. Basically, the behavior after a transmission error is the same as after a collision happens since a station cannot distinguish a packet collision from a transmission error. For this reason, the *CW* will be increased either due to a collision or to an error.

Our mathematical analysis makes use of the same assumptions as in chapter 4; the network consists of n contending stations, each station has always pending packets to transmit (saturation state), utilizing the same channel access mechanism (in our case the basic access) and without the presence of hidden stations. The key assumption of our model is that the collision-error probability of a transmitted packet is constant and independent of the number of packet collisions or transmission errors this packet has suffered in the past. Additionally, we consider the following assumptions for the case of non-ideal channel conditions:

- We assume that the channels between all stations are subject to a constant *BER* value (independent bit errors) or a time-variable *BER* (burst errors).
- We maintain the abstraction of the channel at the *BER* level, rather than consider the Signal-to-Noise Ratio (SNR) at the receiver, since the SNR to *BER* mapping is implementation-dependent.
- Although, *PER* would have the corresponding variability, we assume that the same *PER* is a global parameter and is applicable to all wireless stations.
- Since control packets (RTS, CTS and ACK) are much shorter than a data packet and are transmitted at a more robust rate, the probability of being in error is very small and can be ignored.

Note that the current analysis will be carried out only for the case of basic access mechanism and not for the RTS/CTS reservation scheme which is left for future research due to extreme complexity. This is explained if we take into account the fact that even after a successful RTS/CTS exchange, it is possible for the data packet following the RTS/CTS exchange to experience an error. In order to model the whole system we have to utilize a four-dimensional process $\{s(t), m_1(t), m_2(t), b(t)\}$ where $s(t)$ represents the backoff stage of a station at time t, $m_1(t)$ and $m_2(t)$ are the short and long retry limits and $b(t)$ denotes the backoff counter for the given station. This approach it is exceptionally complex and practically unusable.

B. Calculation of the packet transmission probability

In order to analyse and study the performance of the IEEE 802.11 protocol under an error-prone environment, we utilize the same discrete-time Markov chain model presented in chapter 4 (is illustrated in figure 4.2). The main difference with the case of no transmissions errors (only packet collisions) is that now we denote p_f as the collision-error probability, which is the probability a transmitted packet encounters a collision (at least one of the $n-1$ remaining stations transmit in the same time slot) or is received in error due to channel fading and/or noise. Thus, the values of the transmission probability τ as well as the throughput S will be different from the values computed for an error-free channel in equations (3.10) and (3.14), respectively; both τ and S should be appropriately modified to include transmission errors.

In an error-prone environment, the packet error rate depends on the bit error rate, the packet header and the packet length as we can clearly see in equation (5.1):

$$PER = 1-(1-BER)^{l+H} \tag{5.1}$$

where BER is the link bit error rate, l is the packet payload size and H is the packet header length. When the packet size is larger, the packet transmission is more likely to be corrupted. If the packet size is smaller, the packet is more likely to be received correctly, but the increased overhead ratio will degrade the throughput. Further in this Chapter we will see in detail that a trade-off exists between the desire to reduce the ratio of overhead in the data packet (by adopting larger packet size) and the need to reduce the packet error rate in the error prone channel (by using smaller packet length).

Based on a similar analytical framework and Markov chain model as in chapter 4, the probability τ that a station transmits a packet in a randomly chosen slot time given by:

$$\tau = \sum_{i=o}^{m} b_{i,0} = \sum_{i=o}^{m} p_f^{\,i} \cdot b_{0,0} = b_{0,0} \cdot \frac{1-p_f^{\,m+1}}{(1-p_f)} \tag{5.2}$$

Finally, the probability τ can be acquired from (depending on the values of m and m'):

$$\tau = \begin{cases} \dfrac{2\cdot(1-2p_f)\cdot(1-p_f^{m+1})}{W\cdot(1-(2p_f)^{m+1})\cdot(1-p_f)+(1-2p_f)\cdot(1-p_f^{m+1})} & , \ m \le m' \\ \\ \dfrac{2\cdot(1-2p_f)\cdot(1-p_f^{m+1})}{W\cdot(1-(2p_f)^{m'+1})\cdot(1-p_f)+(1-2p_f)\cdot(1-p_f^{m+1})+W\cdot 2^{m'}\cdot p_f^{m'+1}\cdot(1-2p_f)\cdot(1-p_f^{m-m'})} & , \ m > m' \end{cases} \tag{5.3}$$

where p_f is the probability that a transmitted packet encounters a collision (with probability p) or is received in error (with probability PER):

$$p_f = 1-(1-p)(1-PER) \tag{5.4}$$

By considering that the packet collision probability p is given by equation (3.11) and by using equation (5.1), p_f is finally is equal to:

$$p_f = 1-(1-\tau)^{n-1}(1-BER)^{l+H} \tag{5.5}$$

Since the occurrence of a packet collision and a packet error are independent, the probability p_f can alternatively be given by:

$$\begin{aligned} p_f &= 1-(1-p)(1-PER) \\ &= p + PER - p\,PER \\ &= p+(1-p)PER \\ &= p+(1-p)\left(1-(1-BER)^{l+H}\right) \end{aligned} \tag{5.6}$$

Equations (5.3) and (5.5) form a non-linear system with two unknowns $\tau \in (0,1)$ and $p_f \in (0,1)$. This non-linear system can be solved using numerical methods and has a unique solution (the proof of the uniqueness is similar to the one presented in the Appendix for the error-free case).

C. Saturation throughput

Let P_{tr} be the probability that at least one transmission occurs in a randomly chosen slot time:

$$P_{tr} = 1-(1-\tau)^n \tag{5.7}$$

Moreover, let P_s be the probability that an ongoing transmission is successful:

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n}(1-PER) \tag{5.8}$$

The probability P_c that an occurring transmission collides because two or more stations simultaneously transmit is:

$$P_C = 1-\frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n} \tag{5.9}$$

The probability P_{er} that a packet is received in error is:

$$P_{er} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n} PER \tag{5.10}$$

By considering the above, the probability of a successful transmission in a randomly selected slot is denoted by $P_{tr}\,P_s$, the unsuccessful transmission probability due to simultaneous transmissions in the same slot (packet collision) is $P_{tr}\,P_c$ and the unsuccessful transmission probability due to a transmission error is $P_{tr}\,P_{er}$. Therefore, the throughput efficiency S can be evaluated by dividing the time utilized for transmitting payload information in a slot time by the average duration of a slot time $E[slot]$ (which is different comparing with the one in Chapter 3):

$$S = \frac{P_{tr}\,P_S\,{}^{l}\!/_{C}}{E[slot]} = \frac{P_{tr}\,P_S\,{}^{l}\!/_{C}}{(1-P_{tr})\sigma + P_{tr}\,P_S\,T_S + P_{tr}\,P_c\,T_C + P_{tr}\,P_{er}\,T_{er}} \tag{5.11}$$

where σ is the duration of an empty slot time, T_s , T_c and T_{er} are the average time intervals that the medium is sensed busy due to a successful transmission, a collision or an error respectively. The values of T_s , T_c and T_{er} are equal to:

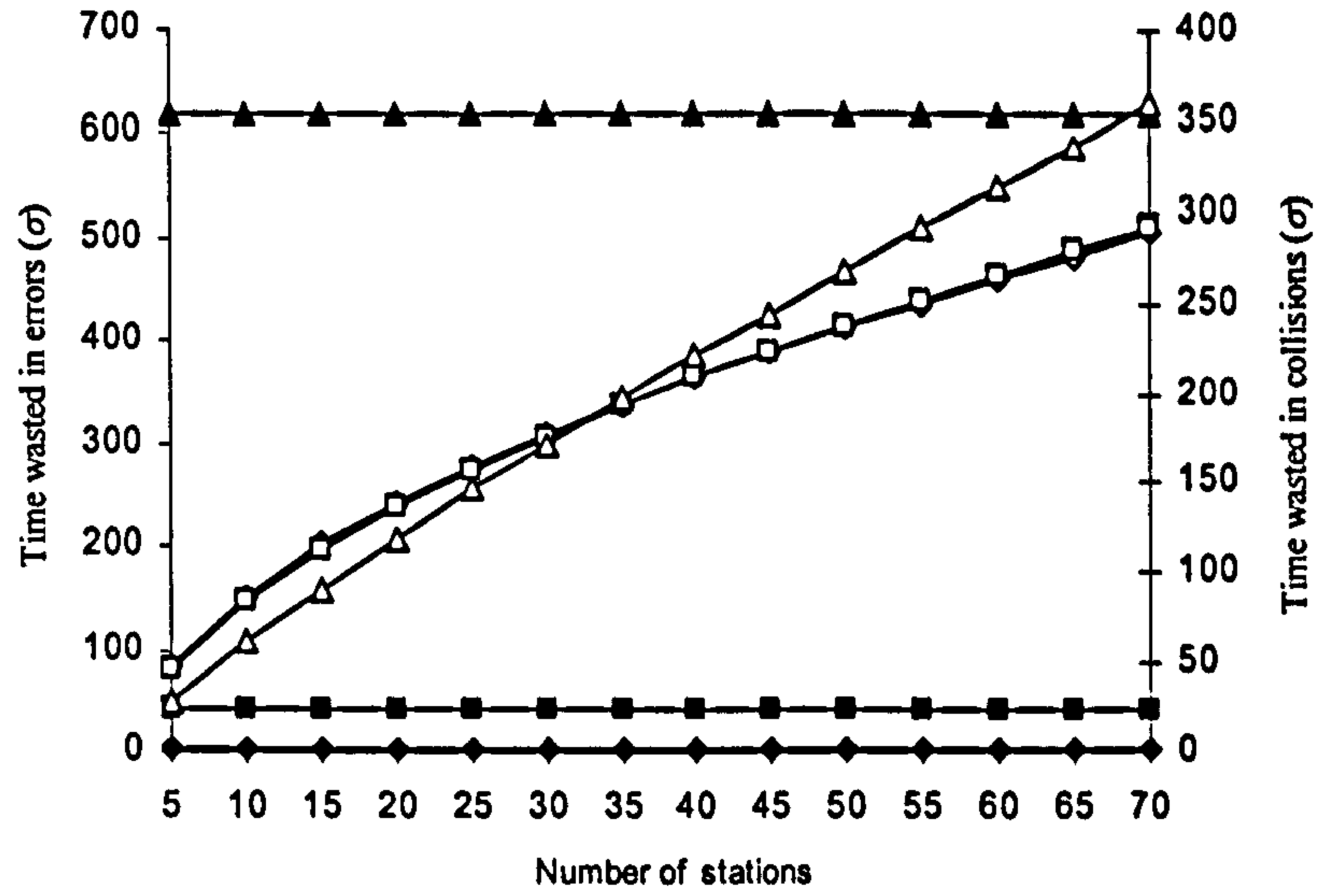
$$T_S = T_C = T_{er} = DIFS + H + l + SIFS + ACK \tag{5.12}$$

Mathematical expressions for the other considered performance metrics i.e. packet delay, packet drop probability, packet drop time and packet inter-arrival time, can be easily acquired from chapter 3 by replacing the collision probability p with the collision-error probability p_f derived in the current chapter.

In order to better understand the impact of transmission errors on performance, we study what occurs in a randomly selected time slot. Dividing numerator and denominator of equation (5.11) by $P_{tr}P_s$, we obtain:

$$S = \frac{l}{\frac{1-P_{tr}}{P_{tr}\cdot P_S}\cdot\sigma + T_S + \frac{P_c}{P_S}\cdot T_C + \frac{P_{er}}{P_S}\cdot T_{er}} \tag{5.13}$$

The denominator of equation (5.13) expresses the average time spent on the channel for a successful transmission. This time is further decomposed into four components. It is important to study the third and fourth terms at the denominator of equation (5.13). The third term represents the time W_{col} wasted due to collisions per successful packet transmission. In fact, $P_c\,/P_s$ is the average number of collided transmissions per successful transmission, which is multiplied by the average duration T_c that the medium is sensed busy due a collision. Following the same approach, the



▲ *Time wasted in errors W_{er}, BER=10^{-4}* △ *Time wasted in collisions W_{col}, BER=10^{-4}*

■ *Time wasted in errors W_{er}, BER=10^{-5}* □ *Time wasted in collisions W_{col}, BER=10^{-5}*

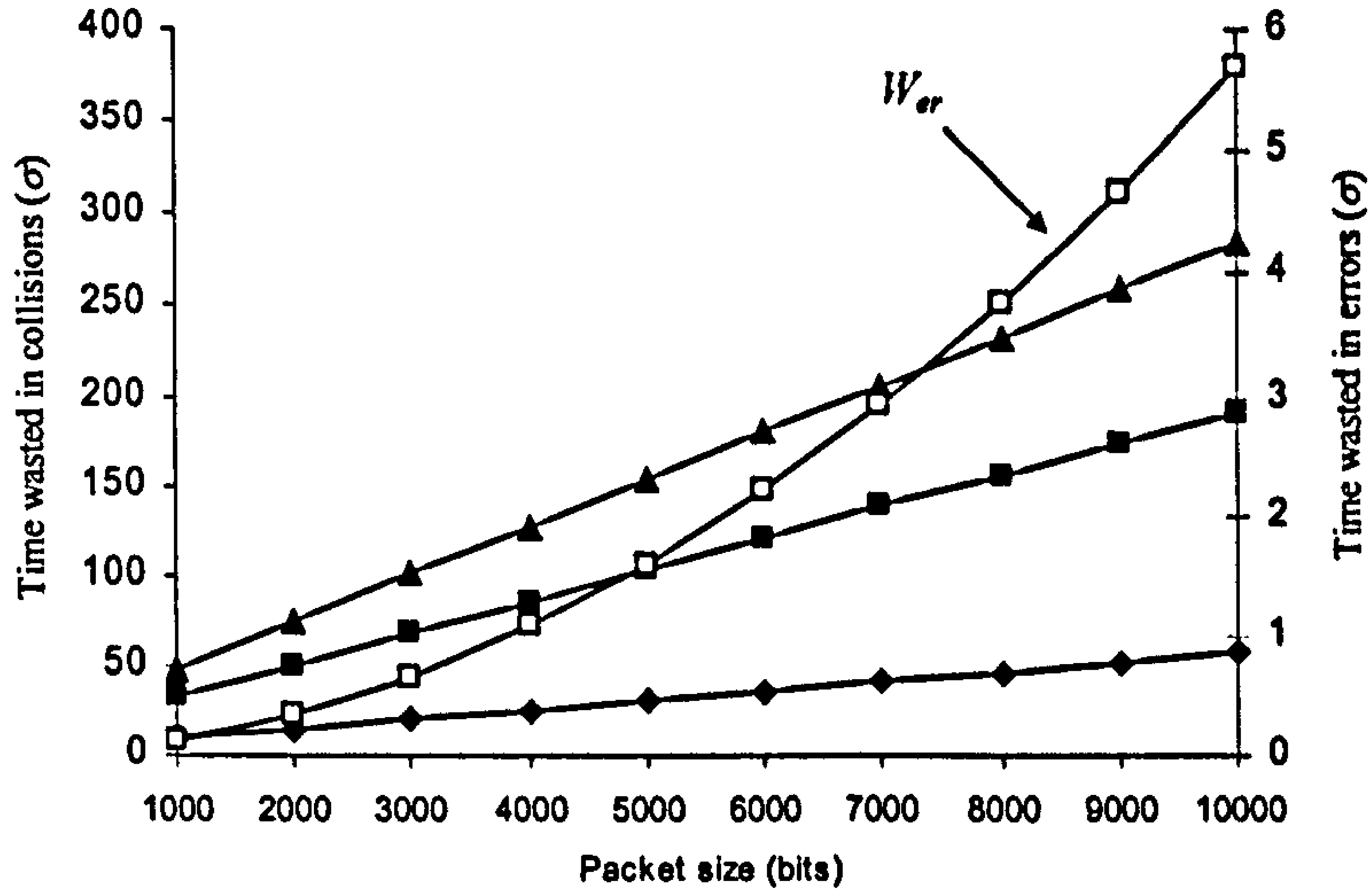
◆ *Time wasted in errors W_{er}, BER=10^{-6}* ◇ *Time wasted in collisions W_{col}, BER=10^{-6}*

Figure 5.1 Average number of slot time units wasted due to errors and packet collisions per successful transmission versus number of stations (802.11b)

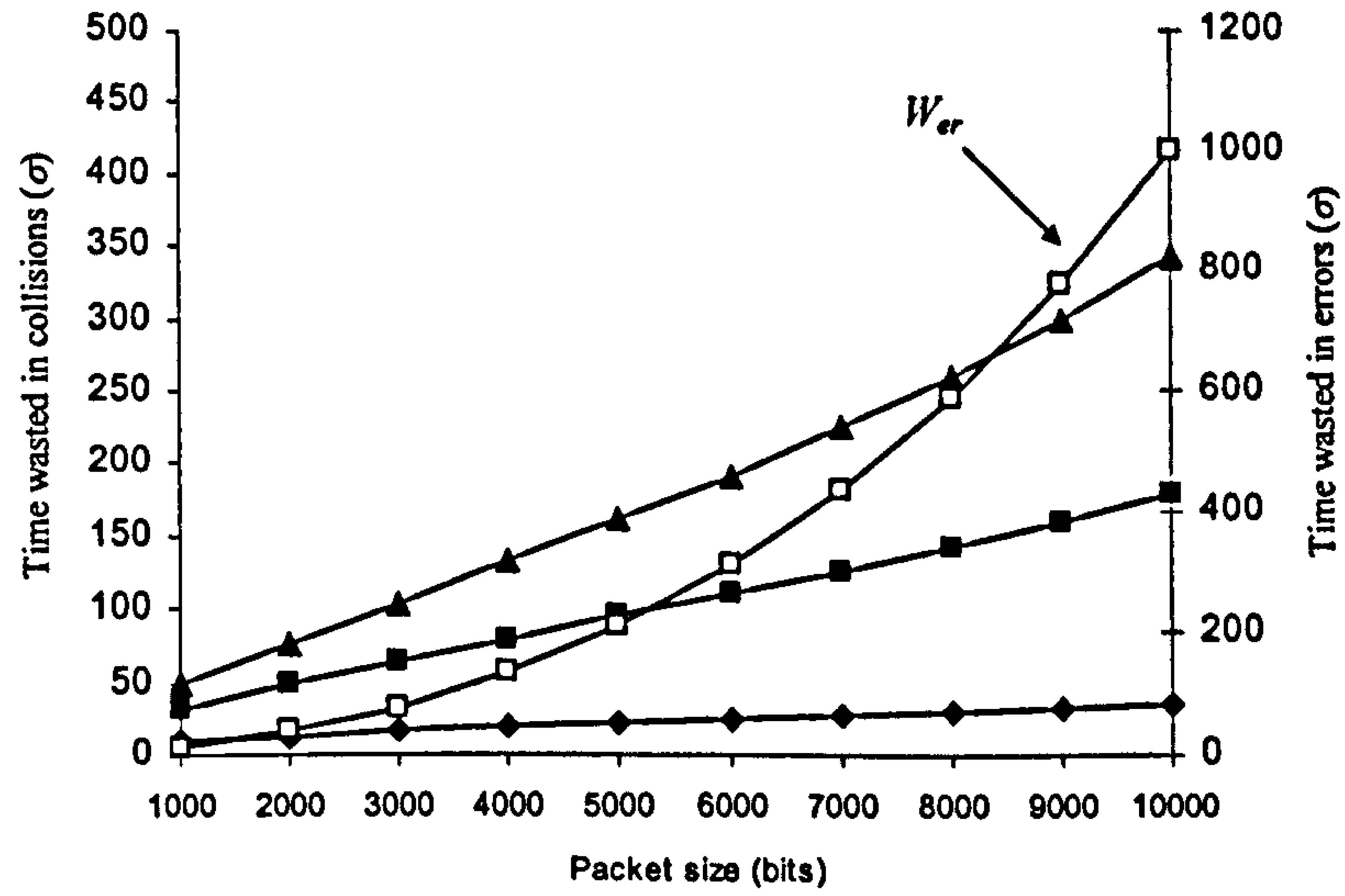
fourth term at the denominator of equation (5.13) denotes the time W_{er} wasted due to transmission errors per successful packet transmission. Note that we can easily prove that the term W_{er} is not affected at all from the number of stations. This is justified by noting that in equation (5.13) the term P_{er}/P_s results to be independent of n.

Figure 5.1 plots the average amount of time spent in transmission errors W_{er} and collisions W_{col} per successful packet transmission, normalized with respect to the slot time σ for a fixed packet size (l=8184 bits) . The figure verifies the fact that the time wasted due transmission errors is not affected by the network size. When the *BER* increases, the time W_{er} increases since more transmission errors take place. In fact, transmission errors slightly affect W_{er} when *BER*=10^{-6} or *BER*=10^{-5} but significantly increase W_{er} for higher *BER* values (*BER*=10^{-4}). Furthermore, the figure shows the significant dependence of the time spent in collisions both from the number of contending stations and transmission errors. More specific, we observe that basic access proves to be sensitive on high values of n and *BER* that significantly penalize overall performance.

Figures 5.2 (a) and (b) plot W_{er} and W_{col} varying packet size for three network sizes (n=5, 25 and 50) and for two different *BER* values (*BER*= 10^{-6} and *BER*= 10^{-4}, respectively). As it has been shown in figure 5.1, the time wasted in transmission errors W_{er} is constantly independent of the network size and, thus, the same W_{er}



(a) BER=10⁻⁶



(b) BER=10⁻⁴

- ◆ *Time wasted in collisions W_{col}, n=5*
- ▲ *Time wasted in collisions W_{col}, n=50*
- ■ *Time wasted in collisions W_{col}, n=25*
- □ *Time wasted in errors W_{er}, n=5, 25 and 50*

Figure 5.2 Average number of slot time units wasted due to errors and packet collisions per successful transmission versus packet size (802.11b)

values are attained for any network size. However, the time W_{er} is highly dependent of the packet size values; larger packets have a higher probability of being in error, and, thus more time is wasted in transmission errors. A remarkable observation from the comparison of the two figures is that the transition of BER from the low value of 10^{-6} to the higher value of 10^{-4} causes a dramatic increase of the time wasted in transmission errors due to the increased packet error probability. Furthermore, the figure illustrates that network size significantly influences the time spent in collisions W_{col} as a result of the fact that many packet collisions are taking place in a highly congested environment.

5.3.2 Mathematical analysis of an error-prone channel for the DIDD protocol

A mathematical analysis in order to evaluate the performance of the DIDD protocol under an error-prone channel can be easily derived by taking into account the previous analysis. The differences with the analysis carried out for IEEE 802.11 protocol are summarized as follows:

- The probability τ that a station transmits a packet in a randomly chosen slot time is different to the one given by equation (5.3) and is equal to:

$$\tau = \frac{2(1-2a)(1-a^{m+1})}{(1-(2a)^{m+1})(1-a)W+(1-2a)(1-a^{m+1})} \tag{5.14}$$

  where $a = \dfrac{p_f}{1-p_f}$ and p_f is the collision-error probability, which is the probability a transmitted packet encounters a collision or is received in error.

- The mathematical modeling of the DIDD protocol can be modeled under the presence of transmission errors for both basic access and RTS/CTS mechanisms since in DIDD there are no retry limits making its analysis easier and less complex. In fact, the T_s, T_c and T_{er} being the average time intervals that the medium is sensed busy due to a successful transmission, a collision or a transmission error respectively, are given for basic access and RTS/CTS mechanisms by:

$$T_S = T_C = T_{er} = DIFS + H + l + SIFS + ACK \tag{5.15}$$

$$\begin{cases} T_S^{RTS} = T_{er}^{RTS} = DIFS + T_{RTS} + SIFS + T_{CTS} + SIFS + T_{DATA} + SIFS + T_{ACK} \\ T_C^{RTS} = DIFS + T_{RTS} + SIFS + T_{CTS} \end{cases} \tag{5.16}$$

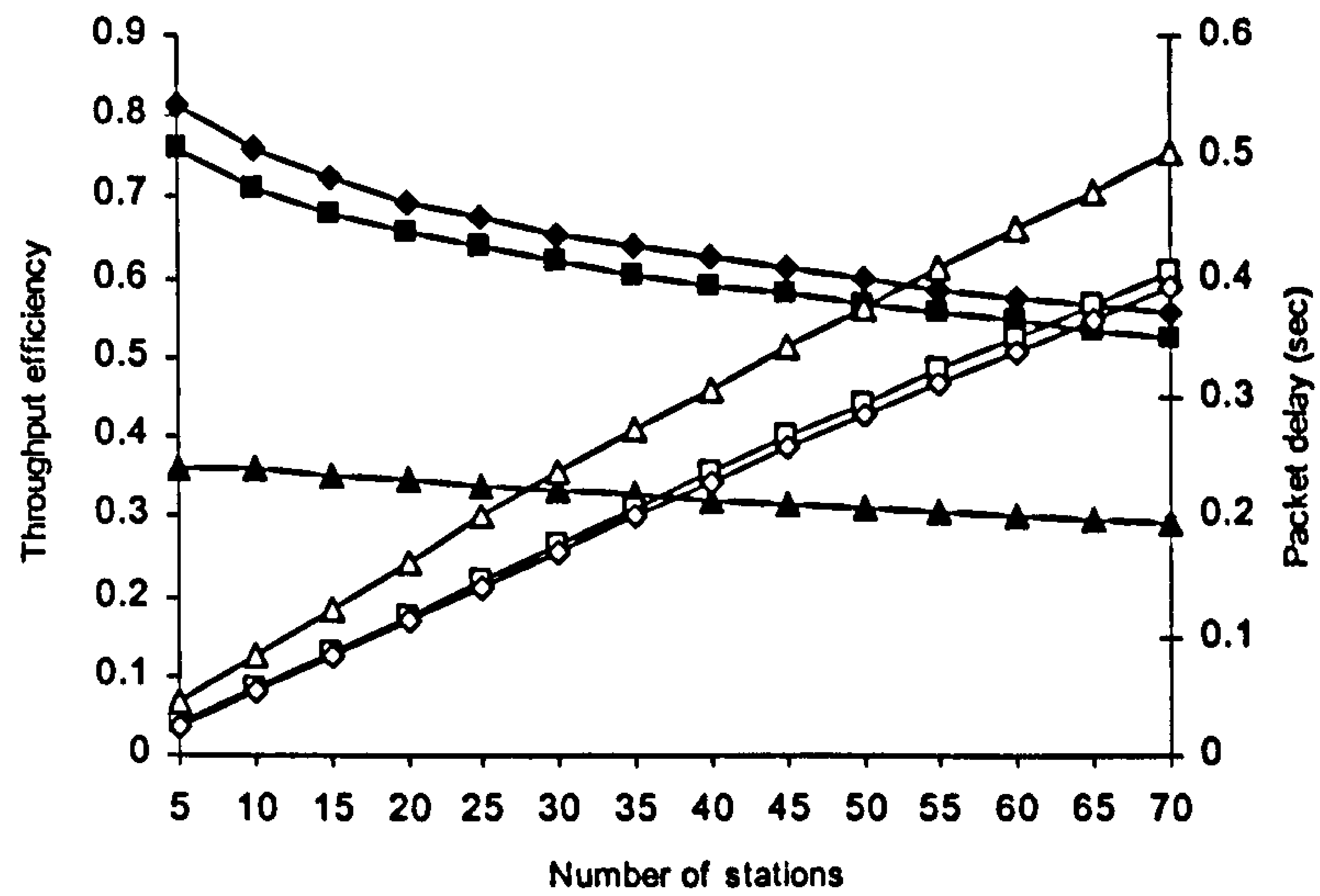
Finally, the mathematical expressions that provide throughput efficiency and packet delay[4] for the DIDD protocol can be easily acquired from chapter 4 by replacing the collision probability p with the collision-error probability p_f derived in the current chapter.

5.4 Performance evaluation under independent transmission errors

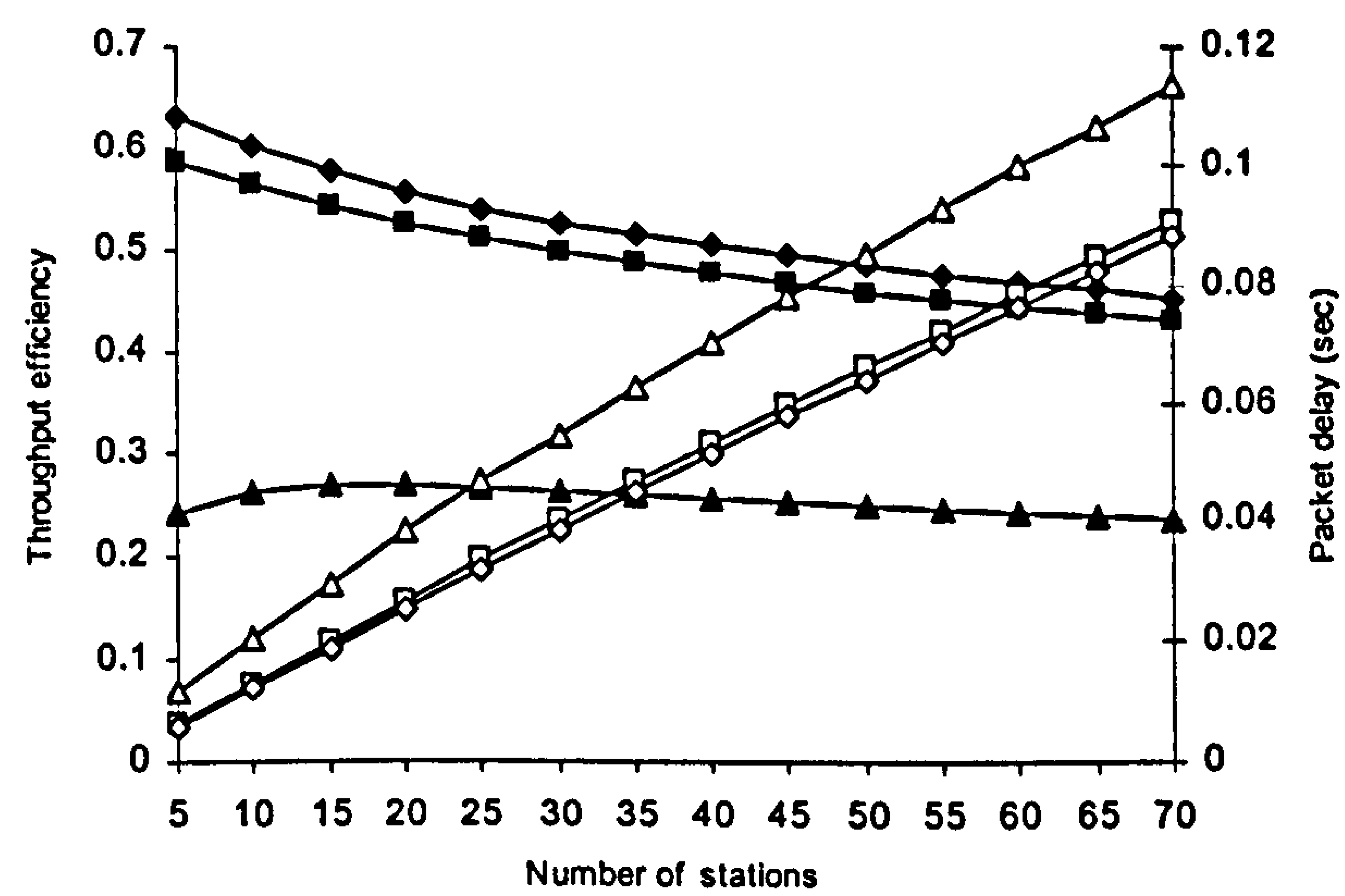
5.4.1 IEEE 802.11 protocol

Figures 5.3 (a) and (b) plot throughput efficiency and packet delay versus network size for fixed packet size (l=8184 bits) and for two data rates (C = 2 and 11 Mbit/s)

[4] Note that we have to use the new values for T_s, T_c and T_{er} derived in equations (5.15) and (5.16).



(a) C= $C_{Control}$= 2 Mbit/s



(b) C= 11 Mbit/s, $C_{Control}$= 2 Mbit/s

◆ *Throughput efficiency, BER=10^{-6}* ◇ *Packet delay, BER=10^{-6}*

■ *Throughput efficiency, BER=10^{-5}* □ *Packet delay, BER=10^{-5}*

▲ *Throughput efficiency, BER=10^{-4}* △ *Packet delay, BER=10^{-4}*

Figure 5.3 Throughput efficiency and packet delay versus network size for different data rates (802.11b)

using the short PHY packet overhead as defined in the IEEE 802.11b standard. The comparison of the two figures shows that when data rate increases from *C*= 2 Mbit/s to 11 Mbit/s, both throughput efficiency and packet delay decrease. Moreover, both figures clearly illustrate that throughput as well as packet delay performance are significantly sensitive to the *BER* values for any data rate. In particular, when *BER* =10^{-4}, performance is degraded to a great extent due to the increased number of errors taking place. On the other hand, lower *BER* values (10^{-6} or 10^{-5}) decrease performance but not considerably for the specific packet size.

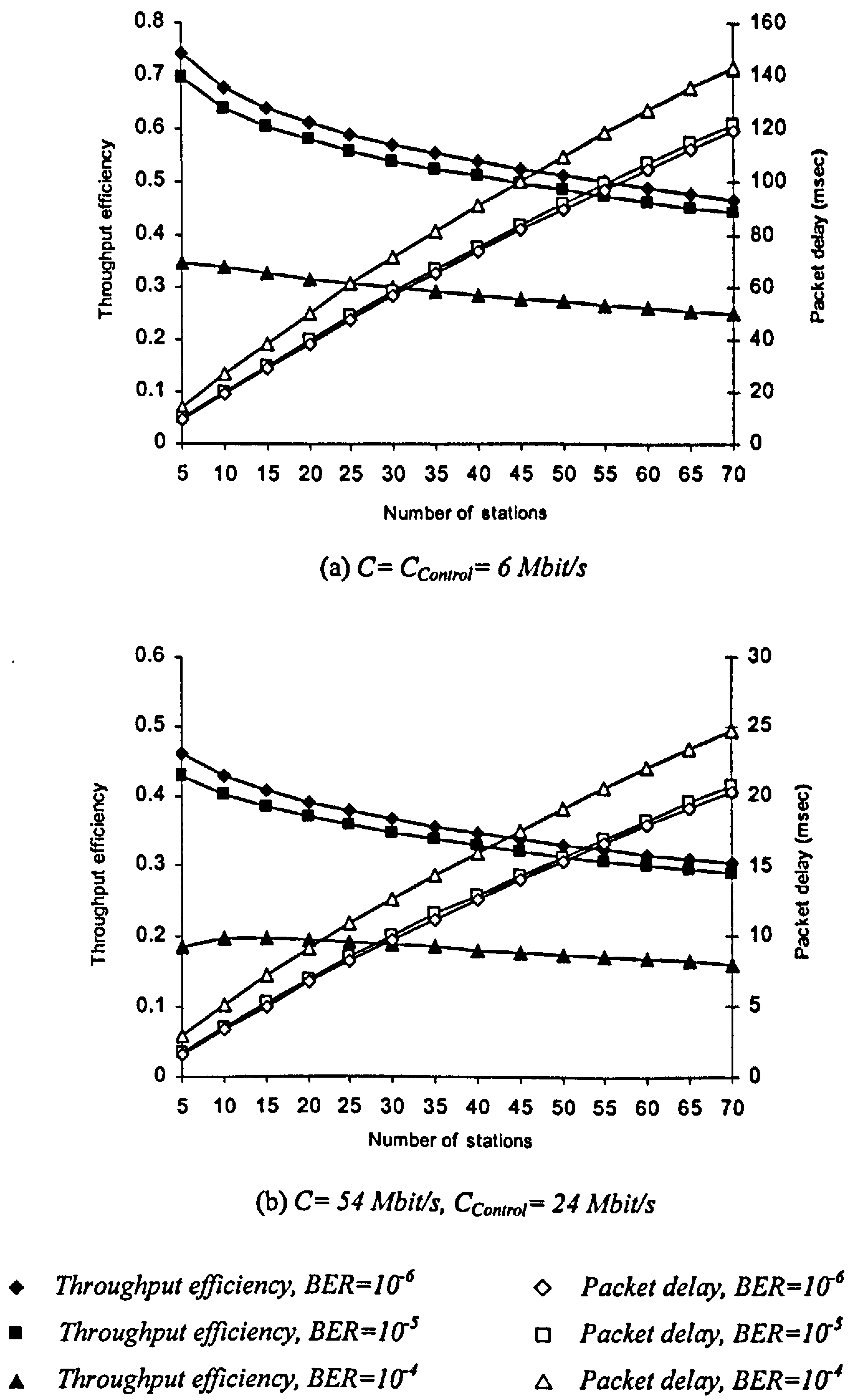


Figure 5.4 Throughput efficiency and packet delay versus network size for different data rates (802.11a)

Figures 5.4 (a) and 5.4 (b) plot packet delay and throughput efficiency against the number of contending stations for the IEEE 802.11a physical layer and for two different pairs of data and control rates (C,$C_{control}$)= (54 Mbit/s, 24 Mbit/s) and (6 Mbit/s, 6 Mbit/s). We clearly observe that similar conclusions are derived with the case of the IEEE 802.11b as it has been shown in figures 5.3(a) and 5.3(b); high *BER* values considerably affect both throughput and packet delay performance. As expected, the IEEE 802.11a PHY achieves a better overall performance comparing with the IEEE 802.11b due to the higher data and control rates. Note that in both figures 5.3 and 5.4

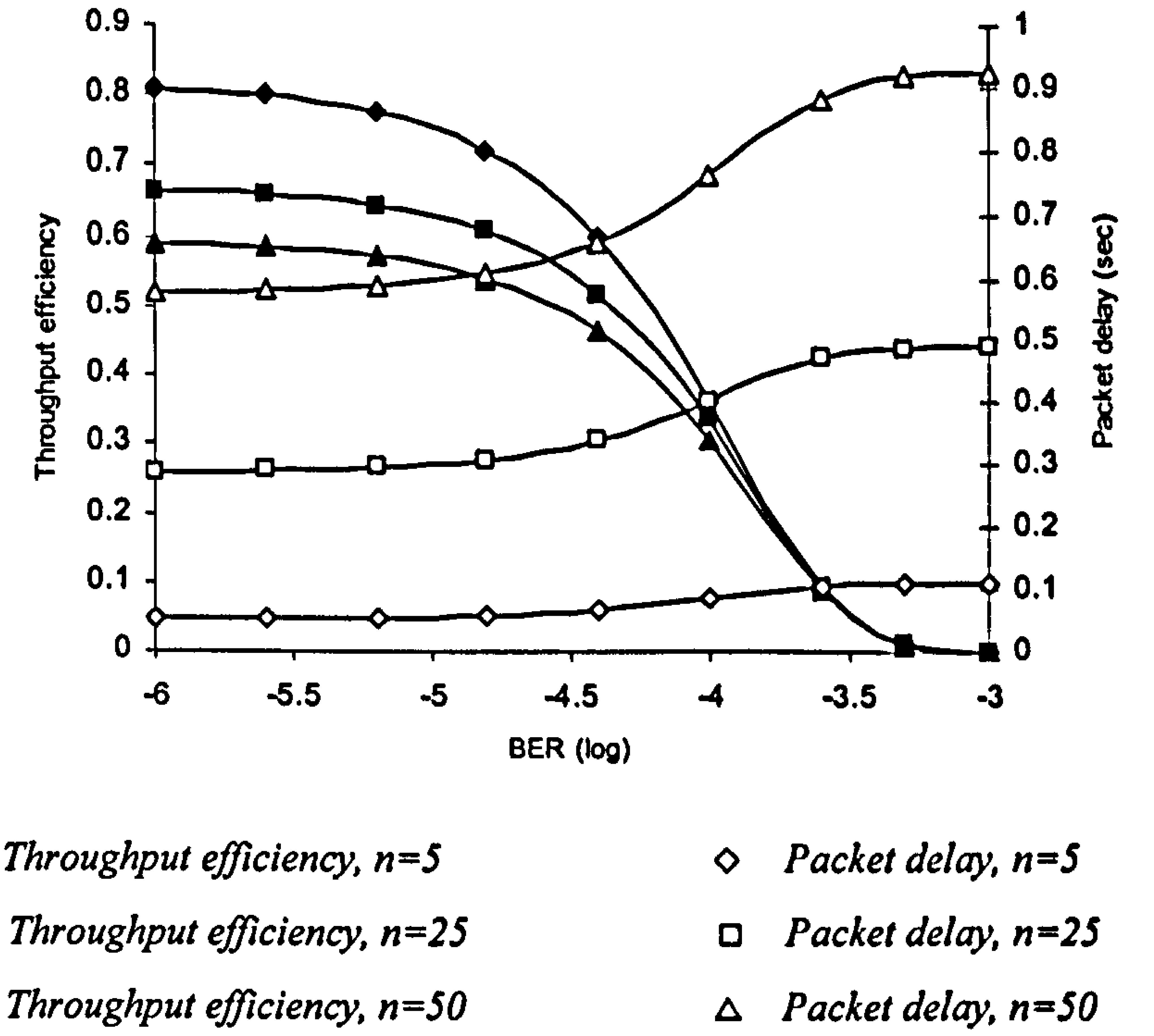


Figure 5.5 Throughput efficiency and packet delay versus BER for various network size values

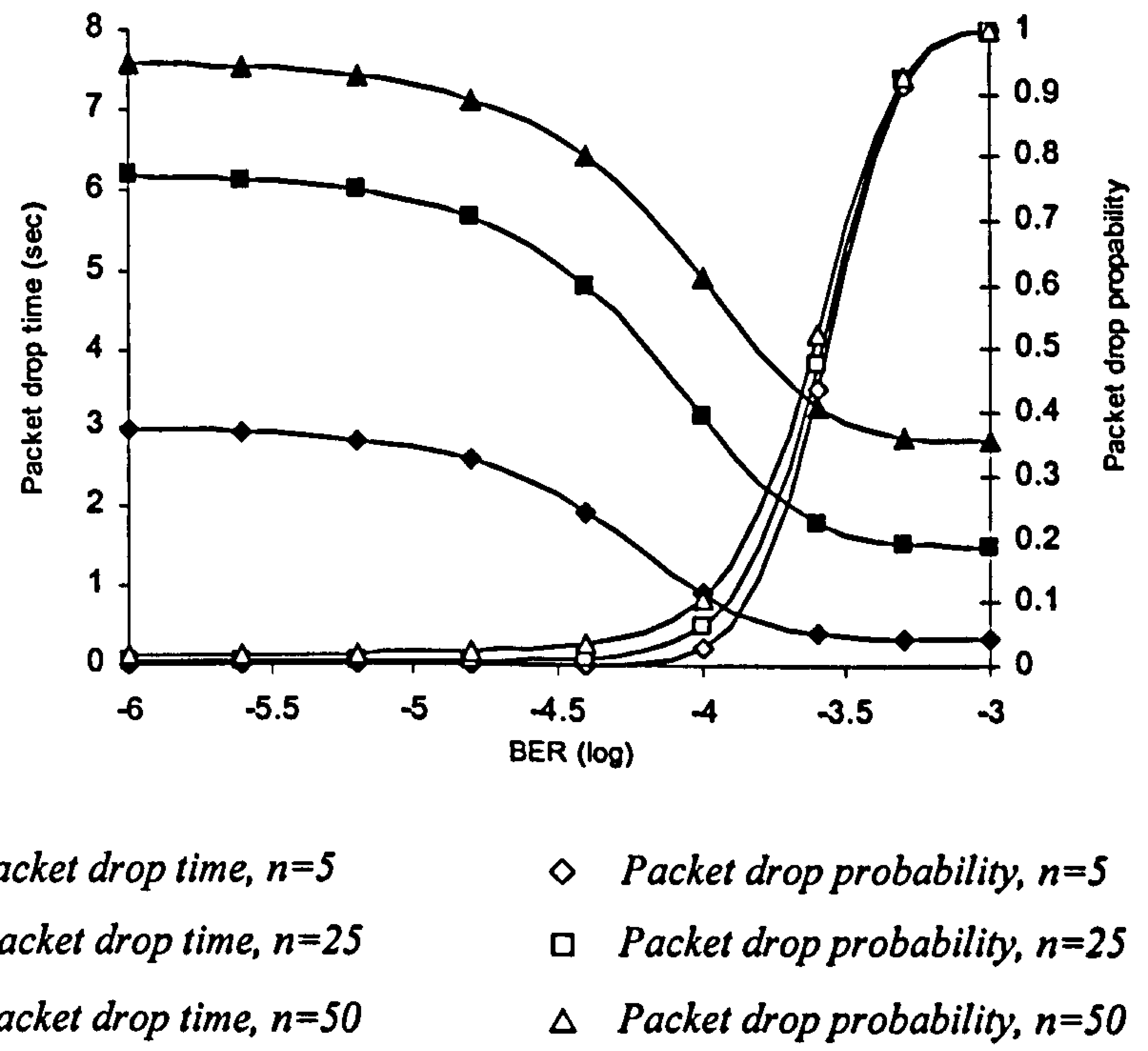


Figure 5.6 Packet drop time and packet drop probability versus BER for various network size values

when the network size increases (more packet collisions), performance degrades regardless the combination of data and control rates or the employed physical layer.

Figures 5.5, 5.6 and 5.7 study in a more detailed approach the effect of transmission errors by plotting throughput efficiency, average packet delay, average packet drop time, packet drop probability and packet inter-arrival time versus *BER*, for three representative network sizes (n = 5, 25 and 50) and a fixed packet size of l= 8184 bits.

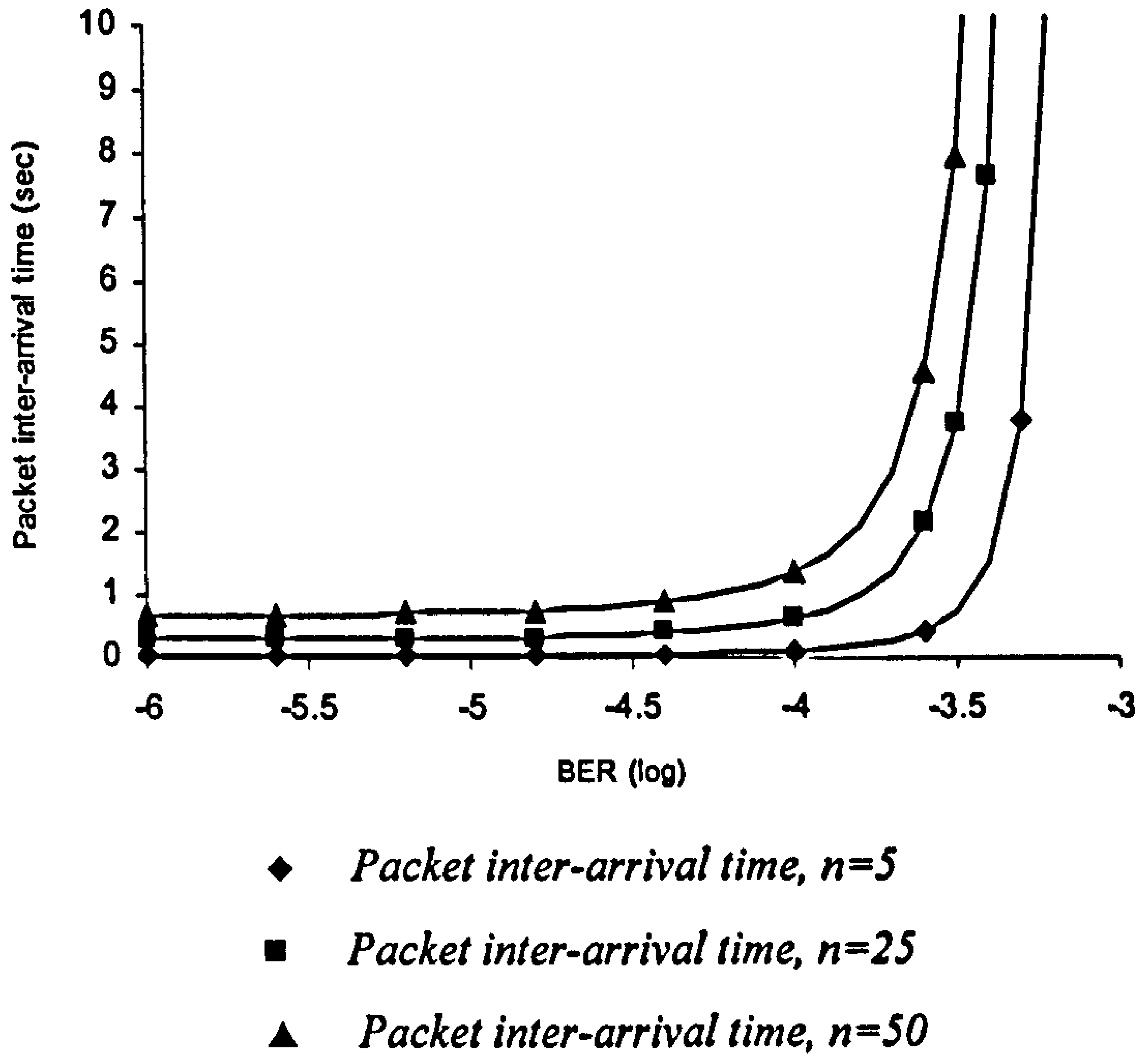


Figure 5.7 Packet inter-arrival time versus BER for various network size values

Figure 5.5 illustrates that when *BER* increases, throughput always gradually degrades and finally drops to 0 (the quality of the wireless medium is very poor and almost no data packets are successfully received). As expected, throughput performance is also sensitive on the network size. The figure also shows that packet delay gradually increases and finally (for high *BER* values) attains roughly unvarying values. More specifically, increasing *BER* results in packet delay growth due to increased number of packet retransmissions which highly delay the successful packet reception. At high *BER* values, although high backoff stages with large *CW* sizes are more often used, the long delay of the dropped packets do not contribute to the average packet delay. Moreover, successfully transmitted packets are less delayed by transmissions of other stations that utilize high *CW* sizes. In view of the fact that the packet delay values at high *BER* concern only a small number of successfully received packets due to high drop probability (see figure 3) and, therefore, have a very small significance.

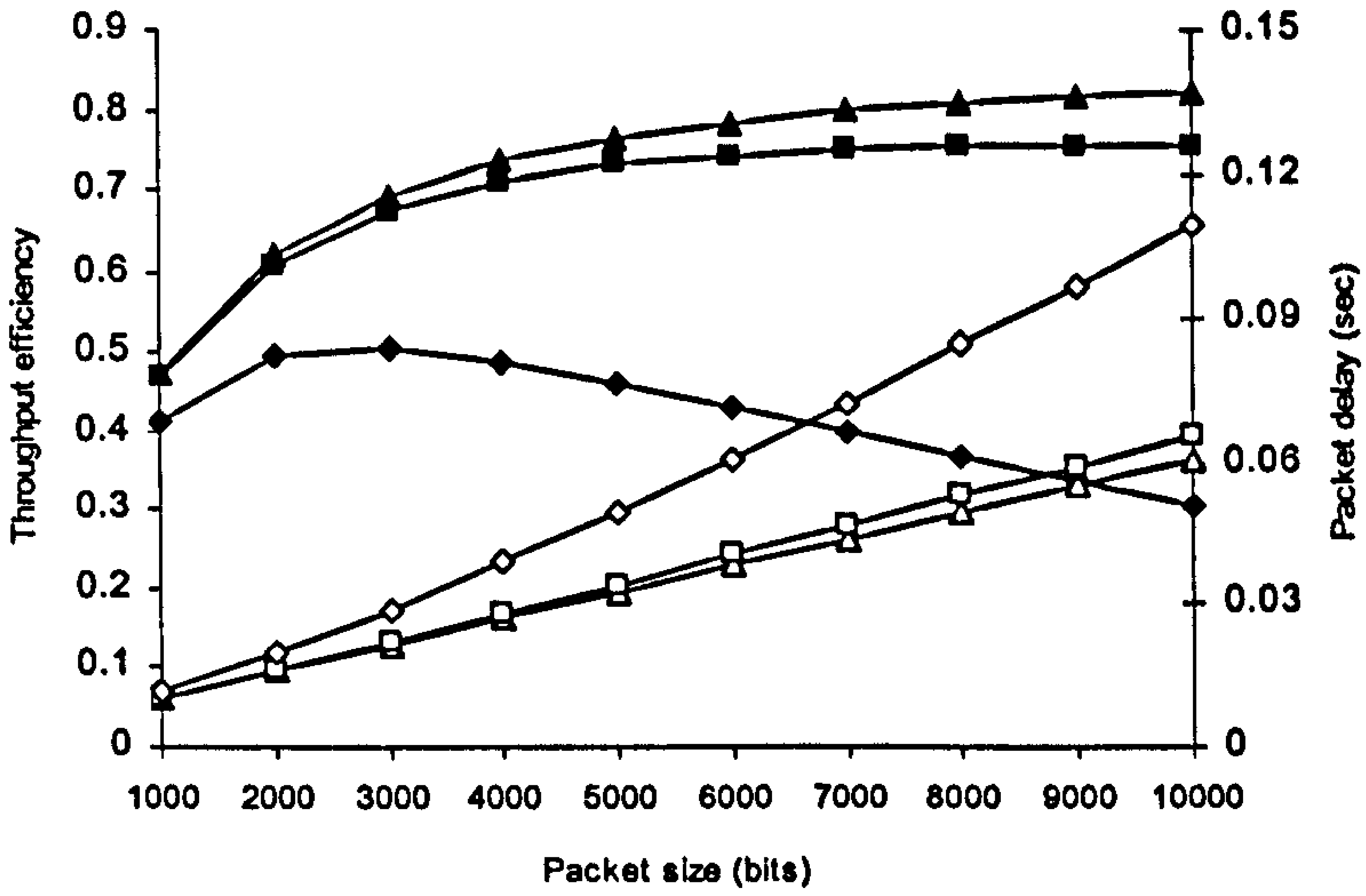
Figure 5.6, which plots packet drop time and packet drop probability versus *BER*, depicts that packet drop time is highly sensitive on the number of contending stations and increases when the network size grows. Increasing *BER* results in packet drop time decrease regardless the network size. In fact, the level of decrease grows with *BER* increase but, when packet drop probability increases rapidly ($BER>10^{-4}$), the decrease level is reduced again and finally packet drop time stays at constant levels. The figure also illustrates that *BER* has a substantial influence on the packet drop probability; when $BER>10^{-4}$ packet drop probability significantly increases due to the increased number of packet transmissions in error. On the contrary, network size also affects packet drop probability but in a less significant level than *BER*.

Finally, figure 5.7 depicts that both network size and transmission errors have a considerable effect on packet inter-arrival time; large network sizes and especially high *BER* values significantly increase packet inter-arrival time, which is the time interval between two successful packet receptions at the receiver. The extremely high values of packet inter-arrival time at high *BER* can be justified by taking into account that when *BER*>10-4, a large amount of packets are being discarded (figure 5.6) and their long delays contribute to the packet inter-arrival time.

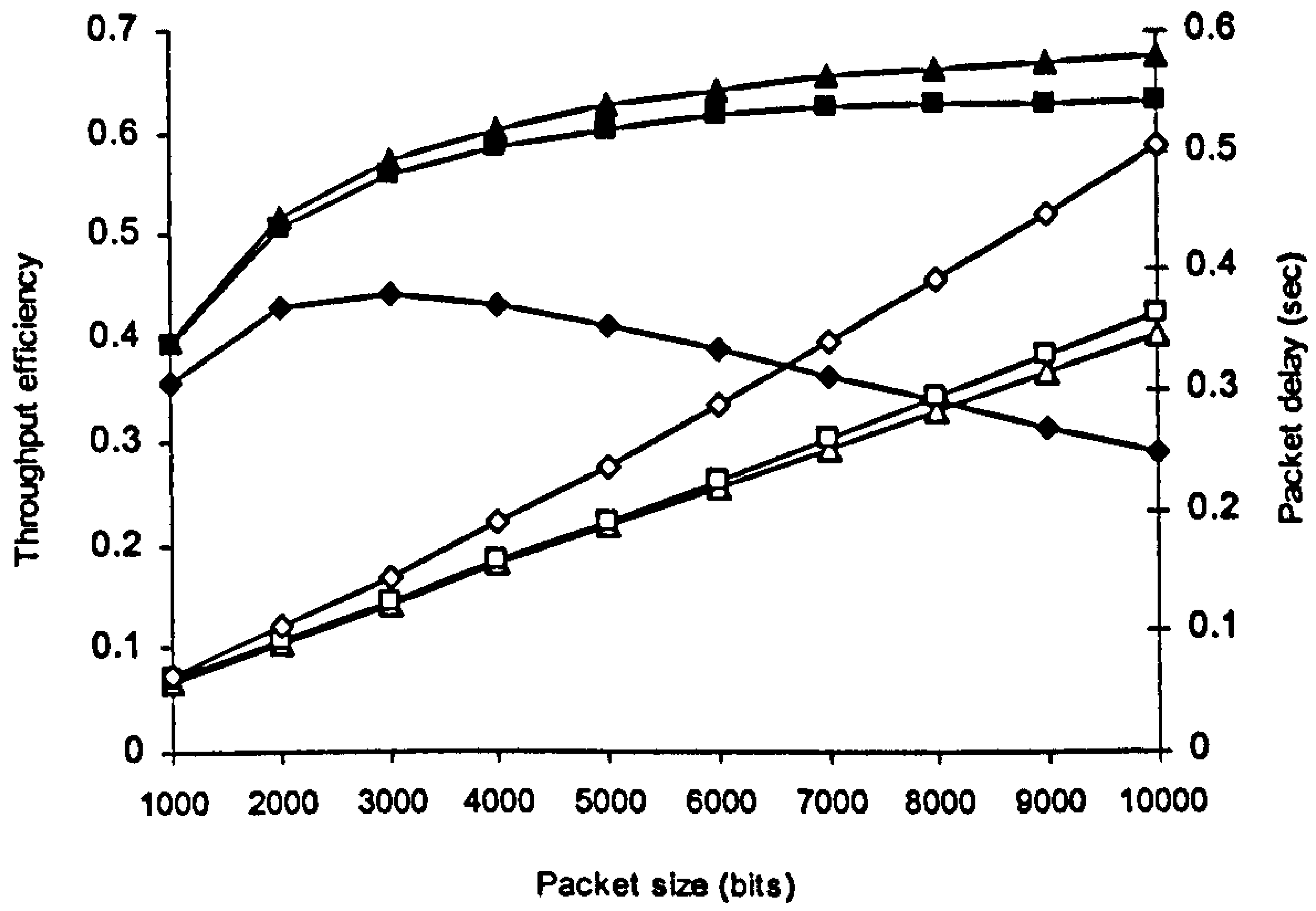
In all the previously presented figures we assumed a fixed packet size of l= 8184 bits. As we have seen before in equation 5.1, the probability of a packet being in error highly depends on packet size apart from the Bit Error Rate. For this reason, figures 5.8 and 5.9 examine the dependency of performance from the packet size by plotting all the considered performance metrics versus l, for three different network sizes (n = 5, 25 and 50) and three *BER* values (*BER*=10^{-4} , 10^{-5} and 10^{-6}).

As we have seen earlier in chapter 3, throughput increases with increasing packet length in an ideal channel (*BER*=0). On the other hand, figure 5.8 illustrates that in an error-prone environment there is a trade-off exists between the desire to reduce the overhead by adopting a larger packet size and the need to reduce packet error rates by using smaller packet length. The figure clearly shows that there is a packet size that maximizes throughput performance in a heavily error-prone channel. This optimal packet length does not vary with the change of the number of contending stations but significantly depends on the *BER*. More specifically, in the case of good or medium quality channel (*BER*<10^{-4}), excessive overhead in each packet actually limits the throughput; larger packet sizes improve throughput performance. As channel conditions deteriorate (*BER*=10^{-4}), it is better to employ a smaller packet size instead of the not effective selection of a large one; the optimal packet length is approximately equal to 3000 bits for any network size. Conversely, we see that for large packet and network size values, packet delay considerably increases especially for high *BER* values.

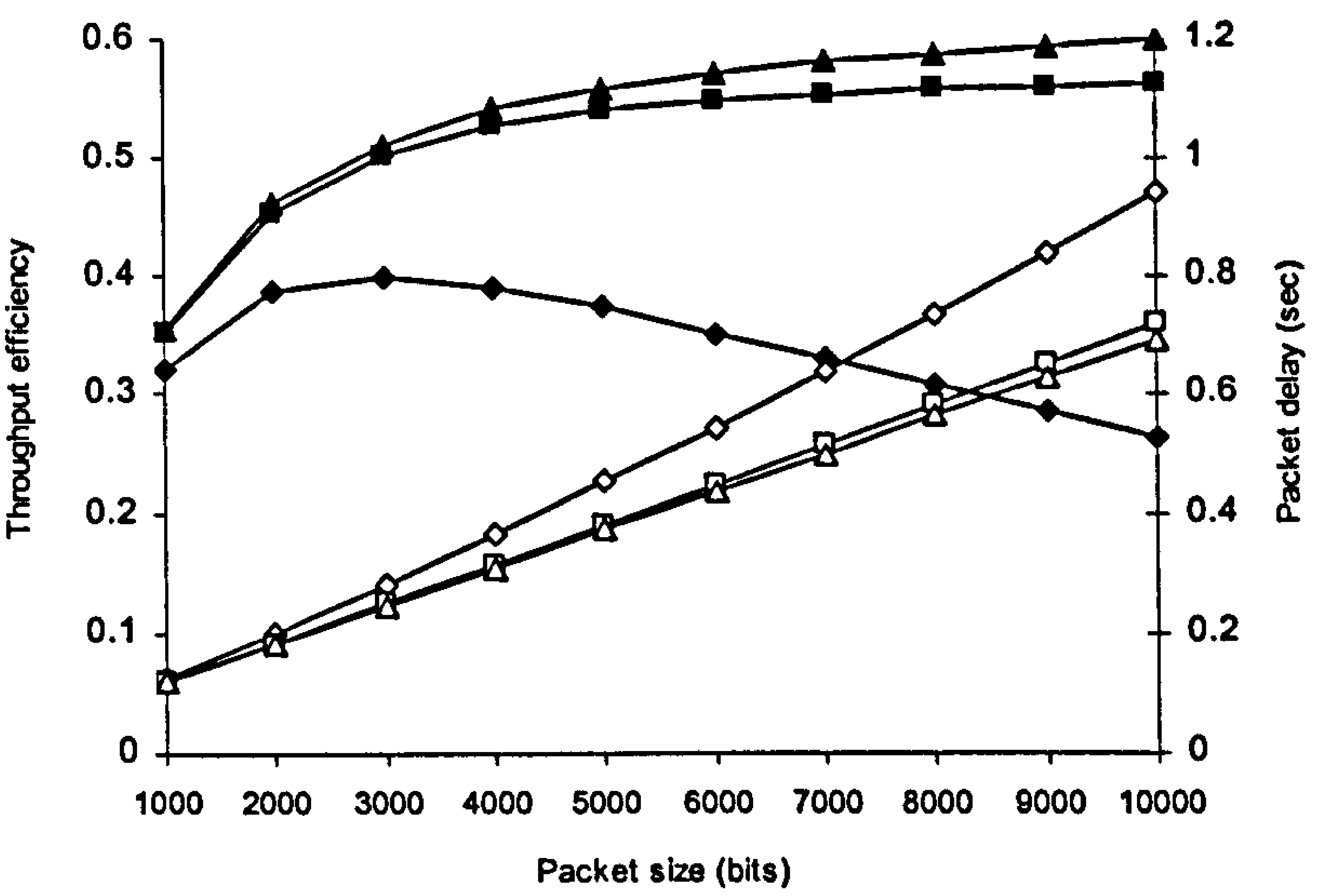
In figure 5.9 we can see that increasing packet size has a similar effect to both packet drop time and inter-arrival time; when packet size increases both performance metrics attain significant higher values. Moreover, both performance metrics appear to be considerably sensitive to network size as well as to *BER*. In particular, high *BER* values significantly decrease (increase) packet drop time (packet inter-arrival time). Note that packet drop probability is not plotted here, being independent of either the packet size or the data rate, and can be obtained for a range of different *BER* values from figure 5.6.



(a) n=5



(b) n=25



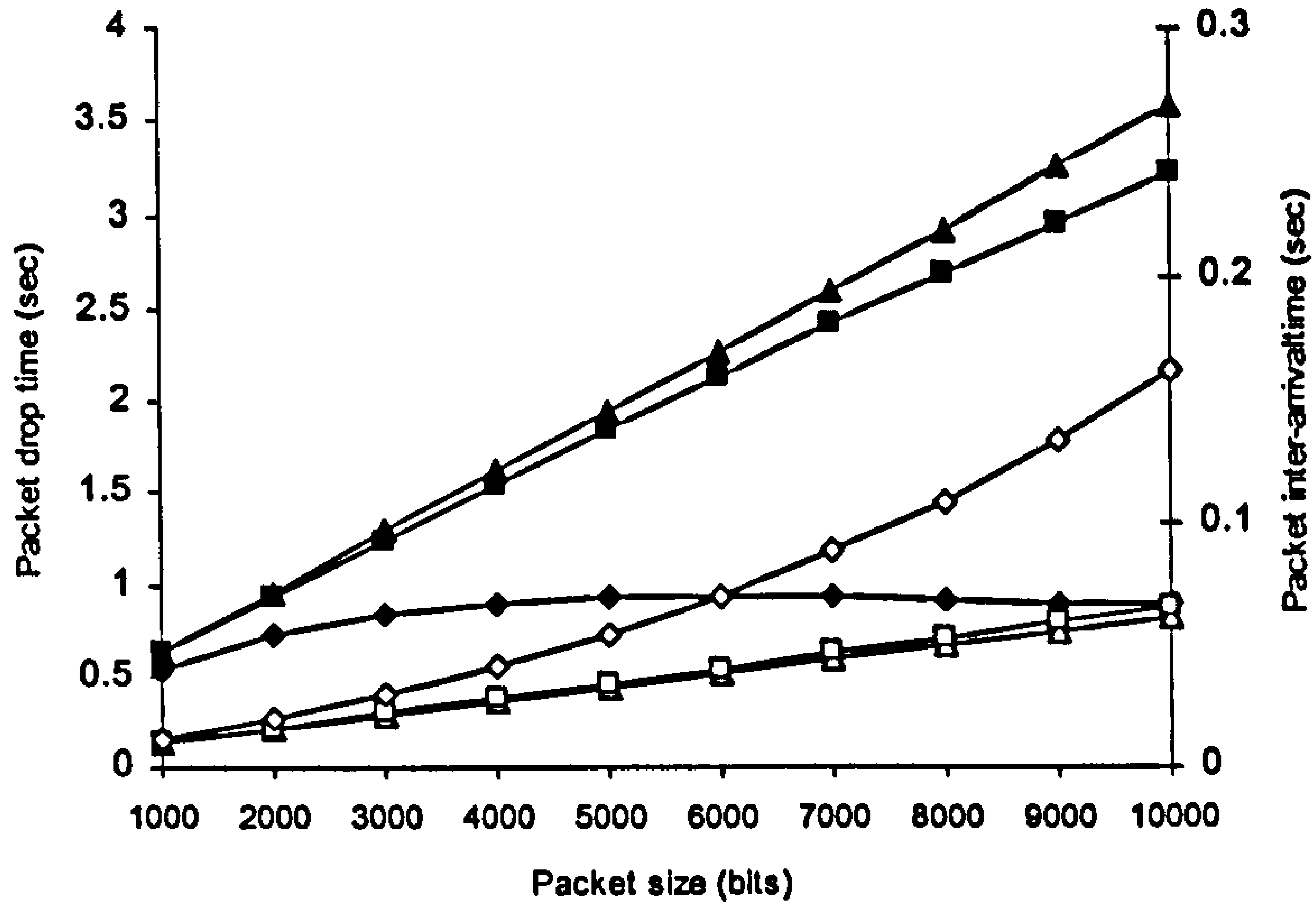
(c) n=50

◆ *Throughput efficiency, BER=10^{-4}* ◇ *Packet delay, BER=10^{-4}*

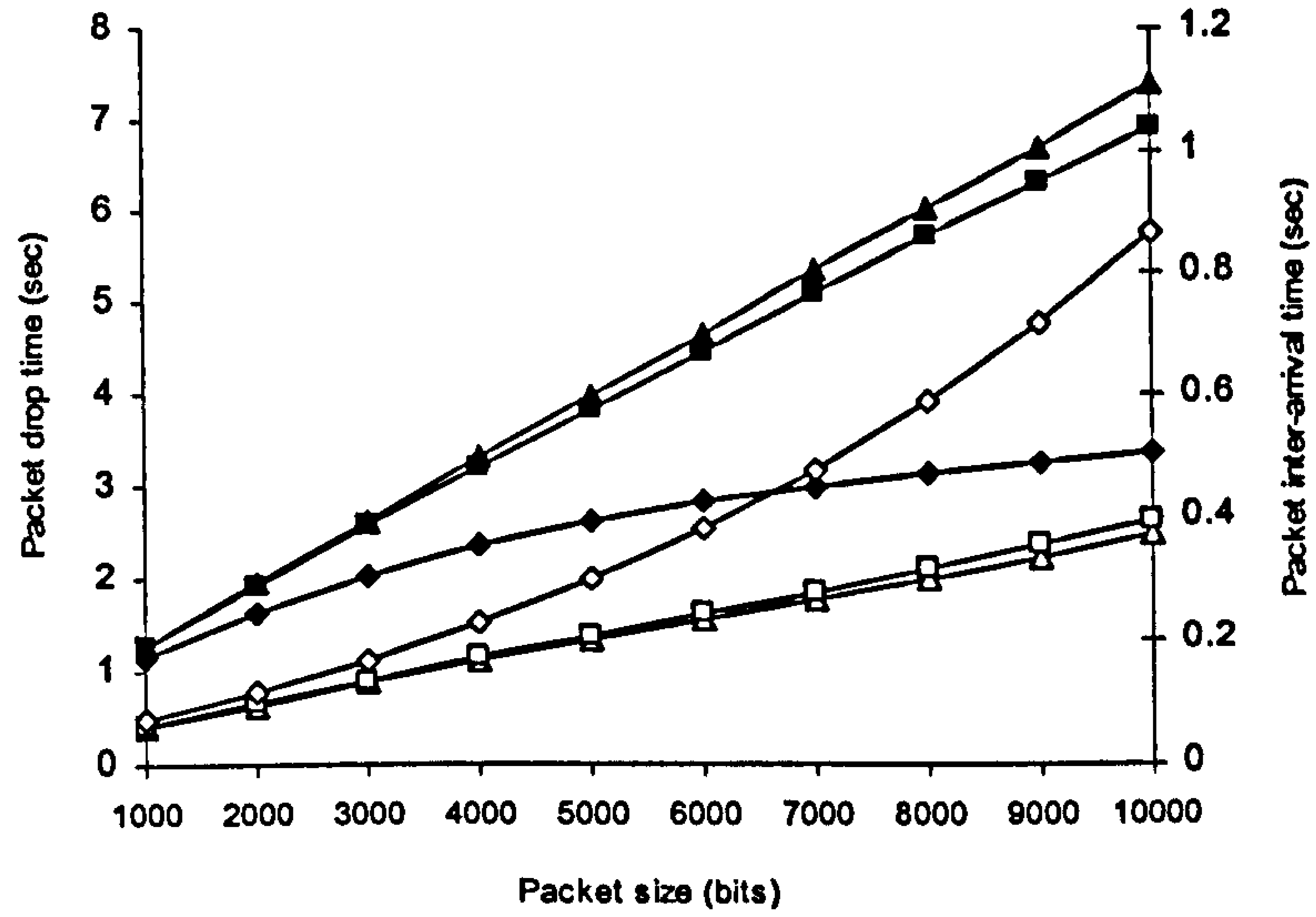
■ *Throughput efficiency, BER=10^{-5}* □ *Packet delay, BER=10^{-5}*

▲ *Throughput efficiency, BER=10^{-6}* △ *Packet delay, BER=10^{-6}*

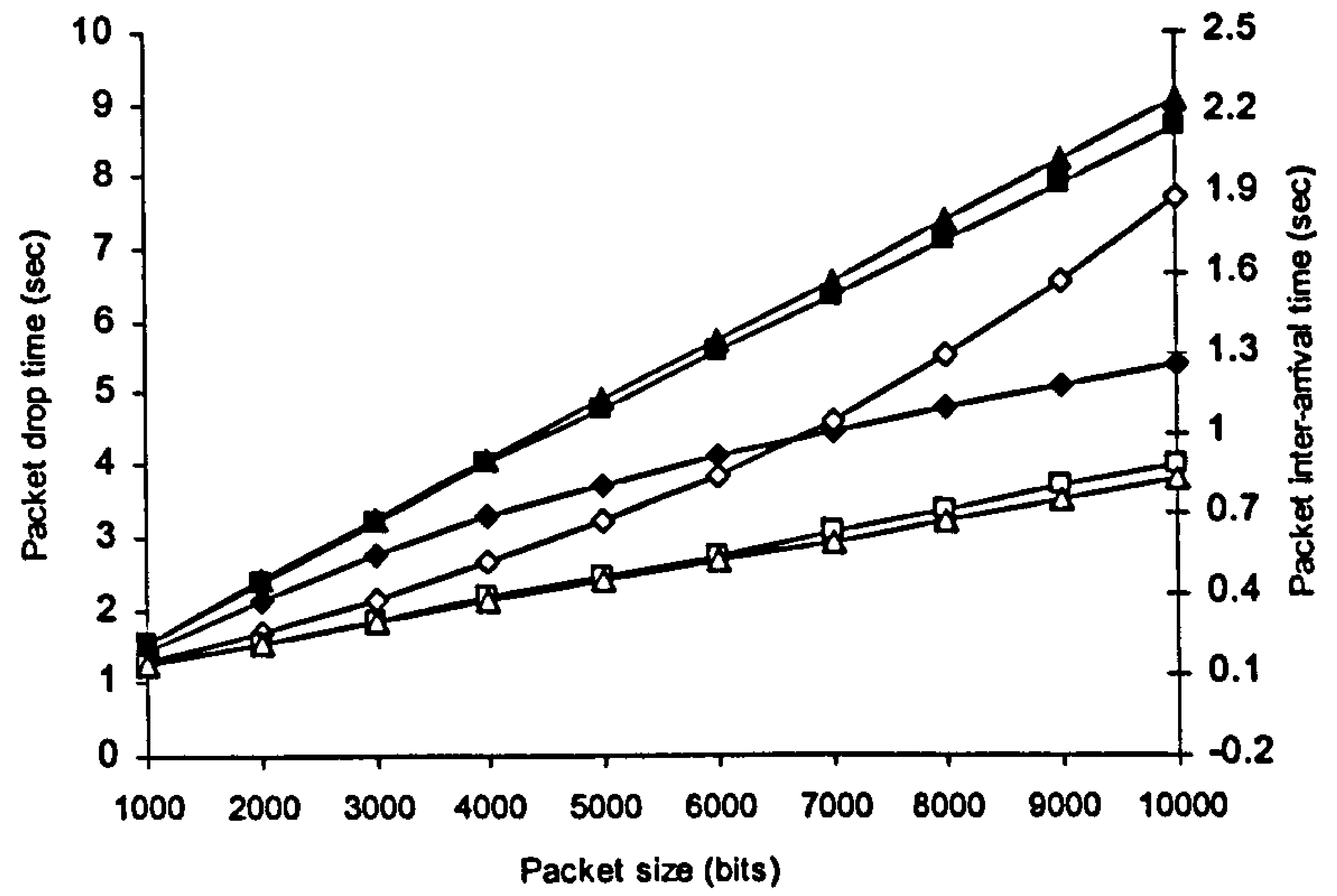
Figure 5.8 Throughput efficiency and packet delay versus packet size for various BER values



(a) n=5



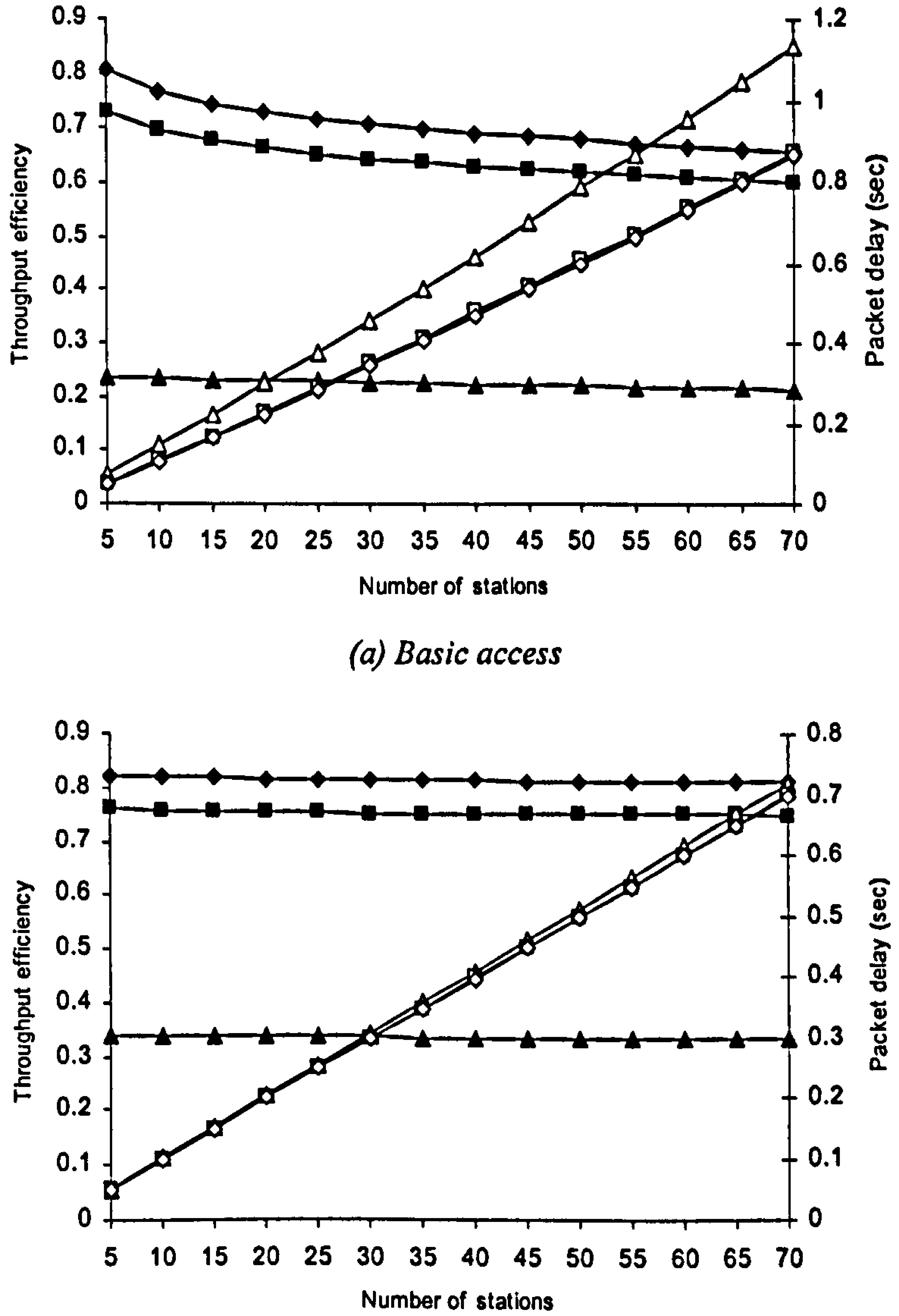
(b) n=25



(c) n=50

◆ *Packet drop time, BER=10^{-4}*	◇ *Packet inter-arrival time, BER=10^{-4}*
■ *Packet drop time, BER=10^{-5}*	□ *Packet inter-arrival time, BER=10^{-5}*
▲ *Packet drop time, BER=10^{-6}*	△ *Packet inter-arrival time, BER=10^{-6}*

Figure 5.9 Packet drop time and packet inter-arrival time versus packet size for various BER values



(a) Basic access

(b) RTS/CTS

▲ *Throughput efficiency, BER=10^{-4}* △ *Packet delay, BER=10^{-4}*

■ *Throughput efficiency, BER=10^{-5}* □ *Packet delay, BER=10^{-5}*

◆ *Throughput efficiency, BER=10^{-6}* ◇ *Packet delay, BER=10^{-6}*

Figure 5.10 Throughput efficiency and packet delay versus network size for the DIDD protocol

5.4.2 DIDD protocol

Figures 5.10 (a) and 5.10 (b) study the effect of transmission errors and network size for the DIDD protocol by plotting throughput efficiency and packet delay versus n, for three BER values (BER=10^{-4}, 10^{-5} and 10^{-6}) for both the basic access and the RTS/CTS schemes, respectively. Both figures illustrate as expected that when the number of contending stations increases, throughput drops off and packet delay increases in both basic access and RTS/CTS schemes as a result of more packet collisions. However, it appears that throughput performance of RTS/CTS scheme is less sensitive on network

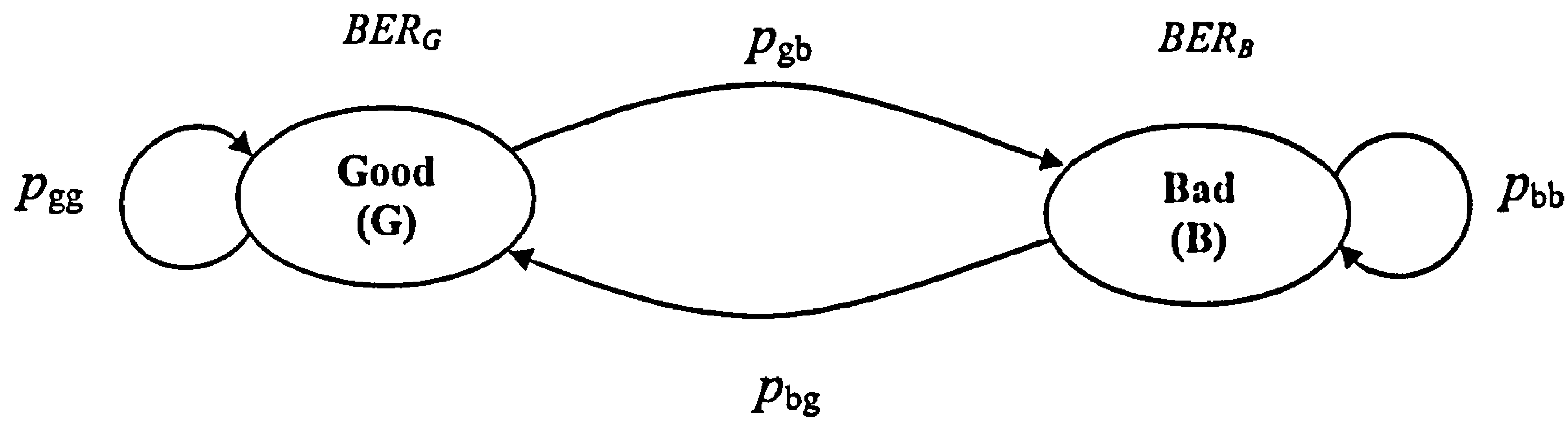


Figure 5.11 Gilbert-Elliot burst model of a wireless channel

size than the basic access scheme. Furthermore, we can observe that although a transmission error penalizes performance when the RTS/CTS is utilized compared to the basic access (note T_{er} values in equation (5.12)), the overall performance achievable by the RTS/CTS is still much higher to that attainable by the basic access scheme. The explanation is mainly because the RTS/CTS scheme achieves a better performance in a congested environment due to shorter packet collision duration.

5.5 Burst transmission errors

5.5.1 Mathematical analysis of the Gilber-Elliot burst error model

The performance analysis of the previous section assumed that the channel bit error rate remained constant and that bit errors were independent. However, over a wireless medium channel, conditions are likely to be time varying. We model the time-varying channel utilizing the well known Gilbert-Elliott model shown in figure 5.11. The wireless channel is modeled as a discrete time Markov chain and is assumed as having two states; one is the *GOOD* state and the other is the *BAD* state. Within each state, bit errors occur according to the independent model with rates BER_G and BER_B, respectively ($BER_G \ll BER_B$). At any time, the probability of the next channel state is determined by only the current state and it has no relationship with any previous state. No matter which state the channel is in, errors occur according to an independent and identical distribution (IID) model. This means that the bits sent over the wireless channel are facing a certain bit error rate to be corrupted, where the value of *BER* is determined by the channel state.

We denote with $X(i)$, i=1,2,.... the channel status sampled for each bit sent over the wireless channel. The event set for X is $\{\text{GOOD}, \text{BAD}\}$ representing the *GOOD* channel status and the *BAD* one, respectively. Note that the bits sent over the wireless channel may or may not belong to the same data packet. The transition probabilities of this model (with p_{xy} being the probability that the current state is x and the next state is y) are given as:

$$\begin{cases} P\big(X(i)=\text{GOOD}\,|\,X(i-1)=\text{GOOD}\big)=p_{gg} \\ P\big(X(i)=\text{BAD}\,|\,X(i-1)=\text{GOOD}\big)=p_{gb} \\ P\big(X(i)=\text{GOOD}\,|\,X(i-1)=\text{BAD}\big)=p_{bg} \\ P\big(X(i)=\text{BAD}\,|\,X(i-1)=\text{BAD}\big)=p_{bb} \end{cases} \tag{5.17}$$

where p_{gb} and p_{bg} represent the transition probabilities from the *GOOD* to the *BAD* state and from the *BAD* to the *GOOD* state respectively, the probabilities p_{bb} and p_{gg} of staying in the same state (*BAD* and *GOOD*, respectively) are[5]:

$$p_{bb}=1-p_{bg} \tag{5.18}$$

$$p_{gg}=1-p_{gb} \tag{5.19}$$

Since the channel is modeled as discrete and memory-less, the next channel state actually can be determined after every bit according to a discrete two-state Markov chain with transition matrix[6]:

$$P=\begin{bmatrix} p_{gg} & 1-p_{gg} \\ 1-p_{bb} & p_{bb} \end{bmatrix} \tag{5.20}$$

The two states of the Markov model in figure 5.11, are termed *GOOD* (with low error probability P_G) and *BAD* (with high error probability P_B). The *BAD* state represents a situation in which is extremely difficult to achieve the successful transmission and correct reception of data packets. In general is $P_B \gg P_G$. Actually, a high-error rate while in the *BAD* state is used to represent a channel fade and a lower error rate in the *GOOD* state represents the channel under normal conditions.

We denote T_{BAD} and T_{GOOD} as the mean sojourn time intervals in the two states i.e. the average time of transmitting bits in *BAD* (error burst) and *GOOD* (error-free burst) states respectively. The transition probabilities p_{gb} and p_{bg} are related to T_{BAD} and T_{GOOD} by the following equations:

$$T_{BAD}=\frac{1}{p_{bg}}=\frac{1}{1-p_{bb}} \quad \text{and} \quad T_{GOOD}=\frac{1}{p_{gb}}=\frac{1}{1-p_{gg}} \tag{5.21}$$

---

[5] The channel state transition probabilities p_{gb}, p_{gb}, p_{bb} and p_{bg} are actually conditional probabilities such that $p_{gg}+p_{gb}=1$ and $p_{bb}+p_{bg}=1$.

[6] The value of the transition matrix can be calculated according to the channel features or from the real world tracing results.

The steady state probability of being in the *GOOD* state is given by:

$$\pi_G = \frac{1-p_{bb}}{1-p_{gg}+1-p_{bb}} = \frac{1-p_{bb}}{2-\left(p_{gg}+p_{bb}\right)} \tag{5.22}$$

and the steady state probability of being in the *BAD* state is given by:

$$\pi_B = \frac{1-p_{gg}}{1-p_{gg}+1-p_{bb}} = \frac{1-p_{gg}}{2-\left(p_{gg}+p_{bb}\right)} \tag{5.23}$$

Finally, the average Bit Error Rate (*BER*) can be calculated by the following equation:

$$BER = P_G\,\pi_G + P_B\,\pi_B$$

$$BER = P_G \frac{p_{bg}}{p_{bg}+p_{gb}} + P_B \frac{p_{gb}}{p_{bg}+p_{gb}}$$

$$BER = P_G \frac{p_{bg}}{p_{bg}+p_{gb}} + P_B \frac{p_{gb}}{p_{bg}+p_{gb}}$$

$$BER = \frac{P_G\,p_{bg}+P_B\,p_{gb}}{p_{bg}+p_{gb}} \tag{5.24}$$

By considering equations (5.17)–(5.18) as well that $P_G = BER_G$ and $P_B = BER_B$, finally the average *BER* is given by:

$$BER = \frac{BER_G\,(1-p_{bb})+BER_B\,(1-p_{gg})}{1-p_{bb}+1-p_{gg}} \tag{5.25}$$

The average BER gives an accurate estimate of the link Bit Error Rate and will be utilized to study the performance of IEEE 802.11 in the presence of burst errors.

5.6 Performance evaluation under burst transmission errors

The current analysis considers two different burst error models. The parameters for the two models are given in table 5.1. The first burst error model used to characterize fading is known as Gilbert-Elliot model [44]. The second model is more realistic with higher *BER* values whereas the holding times in each state are rough estimates obtained by simulations.

Model	BER_G	BER_B	T_{GOOD}	T_{BAD}
Gilbert-Elliot	10^{-10}	10^{-5}	33.333	10
Our model	10^{-6}	10^{-4}	20	2

Table 5.1 Parameters for the two employed burst error models

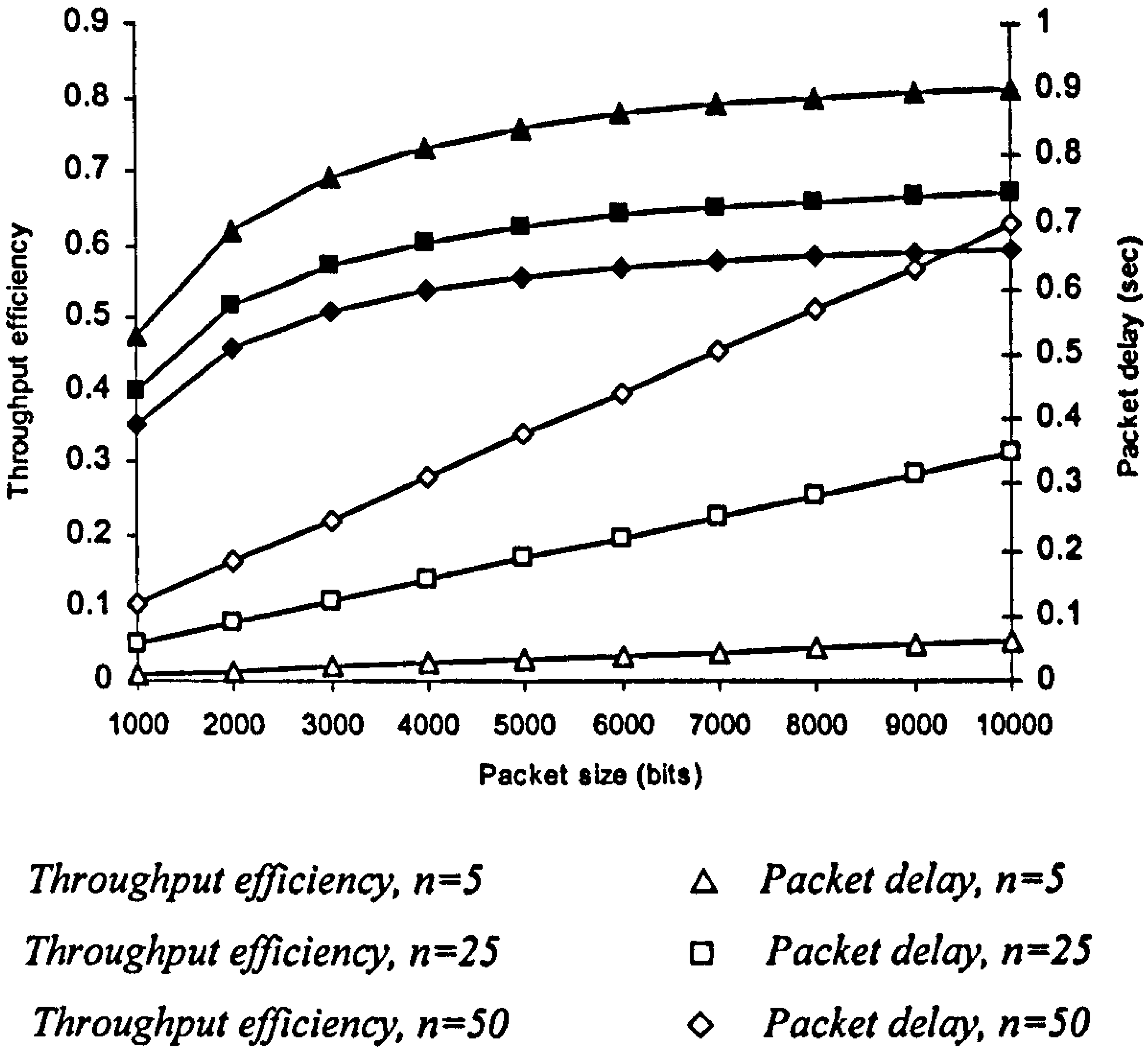


Figure 5.12 Throughput efficiency and packet delay against packet size varying network size ($BER_G=10^{-10}$, l=8184 bits, p_{gg}= 0.97, p_{bb}= 0.9) Gilbert – Elliot model

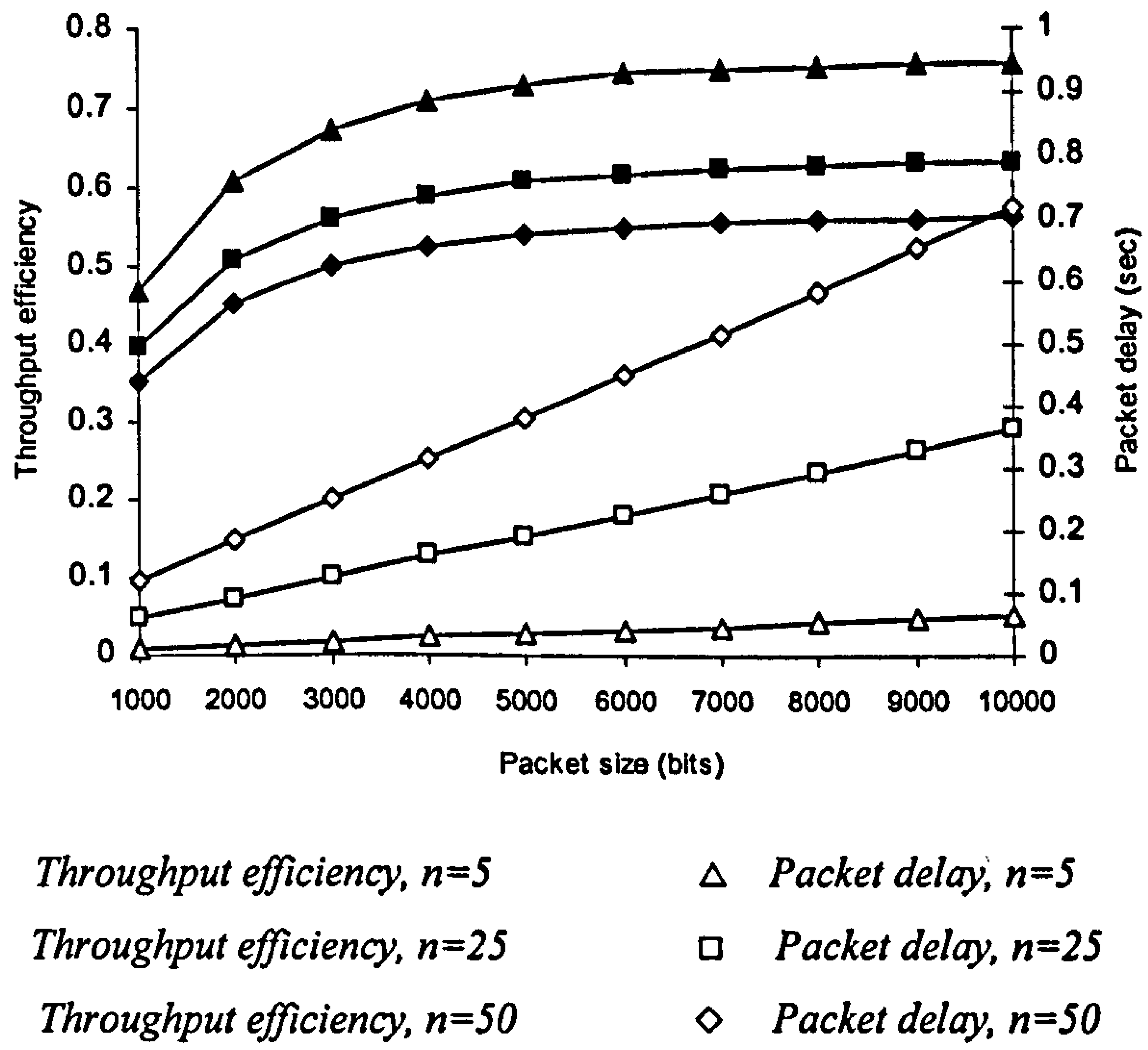


Figure 5.13 Throughput efficiency and packet delay against packet size varying network size ($BER_G=10^{-6}$, l=8184 bits, p_{gg}= 0.95, p_{bb}= 0.5) Proposed model

5.6.1 IEEE 802.11 protocol

Figures 5.12 and 5.13 illustrate the effect of packet payload size (l) on performance for the two employed burst error models, the Gilbert-Elliot and the proposed model, respectively. We can observe that the two error models achieve a similar throughput efficiency and packet delay performance. This can be explained by considering the fact in both models the average BER is almost the same. In fact, table 5.1 easily explains the main difference between the two error models; in the Gilbert-Elliot model the BER in both the $GOOD$ and BAD states is much lower (10^{-10} and 10^{-5}, respectively) than the case of the proposed model (10^{-6} and 10^{-4}, respectively). On the other hand, in the proposed model the BER is greatly higher but the average time spent on the in the two states is much lower than the one in the Gilbert-Elliot model. The figures also show that the performance of IEEE 802.11 is significantly sensitive to the utilized packet size in an error-prone environment. Clearly there is a tradeoff between the throughput constraint (favoring larger packets) and the delay constraint (choosing smaller packets).

Figure 5.14 depicts how throughput and packet delay performance is affected by different values of BER_B ; we consider 3 different values (10^{-5}, 10^{-4} and 10^{-3}) for the proposed error model (BER_G=10-6, T_{BAD}=2, T_{GOOD}=20). The figure shows that performance degrades when the number of stations increases due to the increased packet collision probability regardless the value of BER_B. We can also observe that when BER_B is equal to 10^{-5} or 10^{-4}, performance is not greatly affected. On the other hand, a higher BER_B value ($BER_B = 10^{-3}$) significantly decreases performance, indicating a poor quality link where retransmissions due to errors are taking place more often.

Figures 5.15 and 5.16 study the effect of the time p_{bb} that will be spent in the BAD state for 3 different network sizes (n = 5, 25 and 50) as well as 3 different packet sizes (l = 1000, 8184 and 10000 bits). The figures show that when p_{bb} increases, we note that performance becomes significantly worse (especially for $p_{bb} > 0.8$). In fact, when p_{bb} is equal to 1, the average BER gets equal to BER_B (this can be easily verified from equation (5.25) if we substitute p_{bb}=1). That actually means that the average BER experienced by any packet becomes equal to the encountering BER in the BAD state. Another interesting observation is that throughput performance is better for large packets (l=10000 bits) and for small values of p_{bb} . However, when p_{bb} attains higher values, throughput performance appears to be sensitive to the utilized packet size; this is explained since more time will be spent in the BAD state than in the $GOOD$ state (more transmission errors will take place) and longer packets are more susceptible to errors. On the other hand, packet delay always increases for either larger packet or network sizes or higher p_{bb} values.

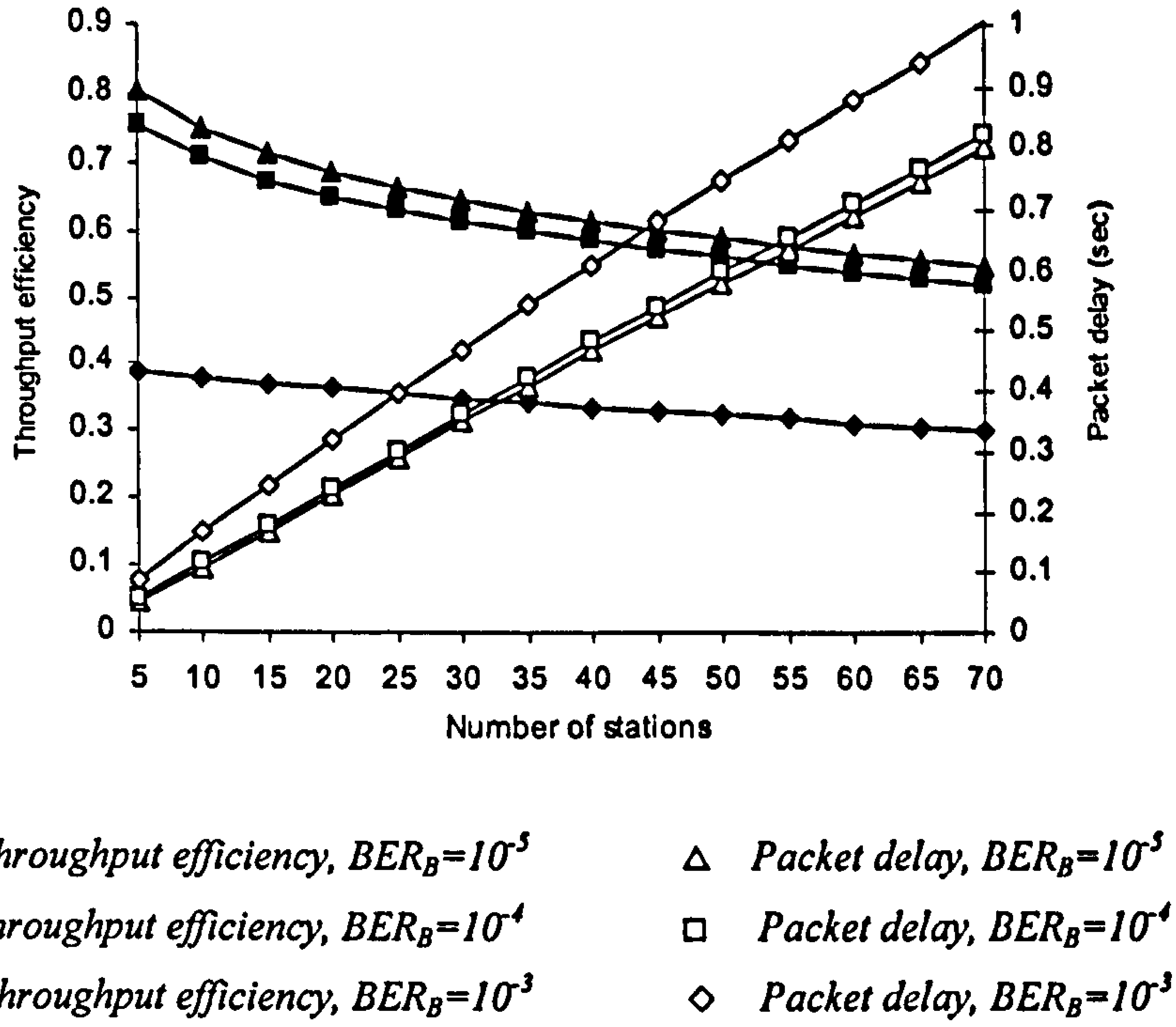
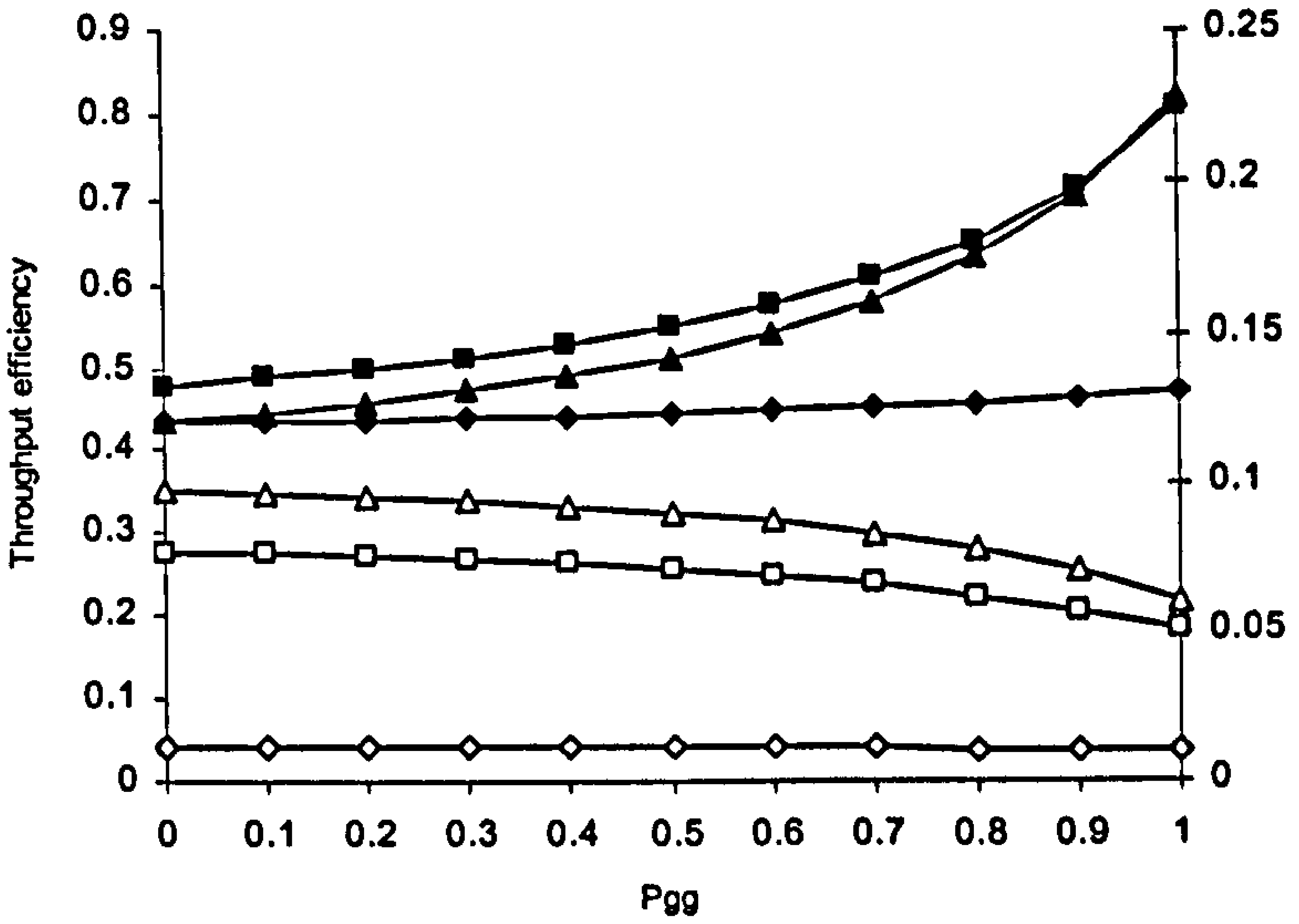
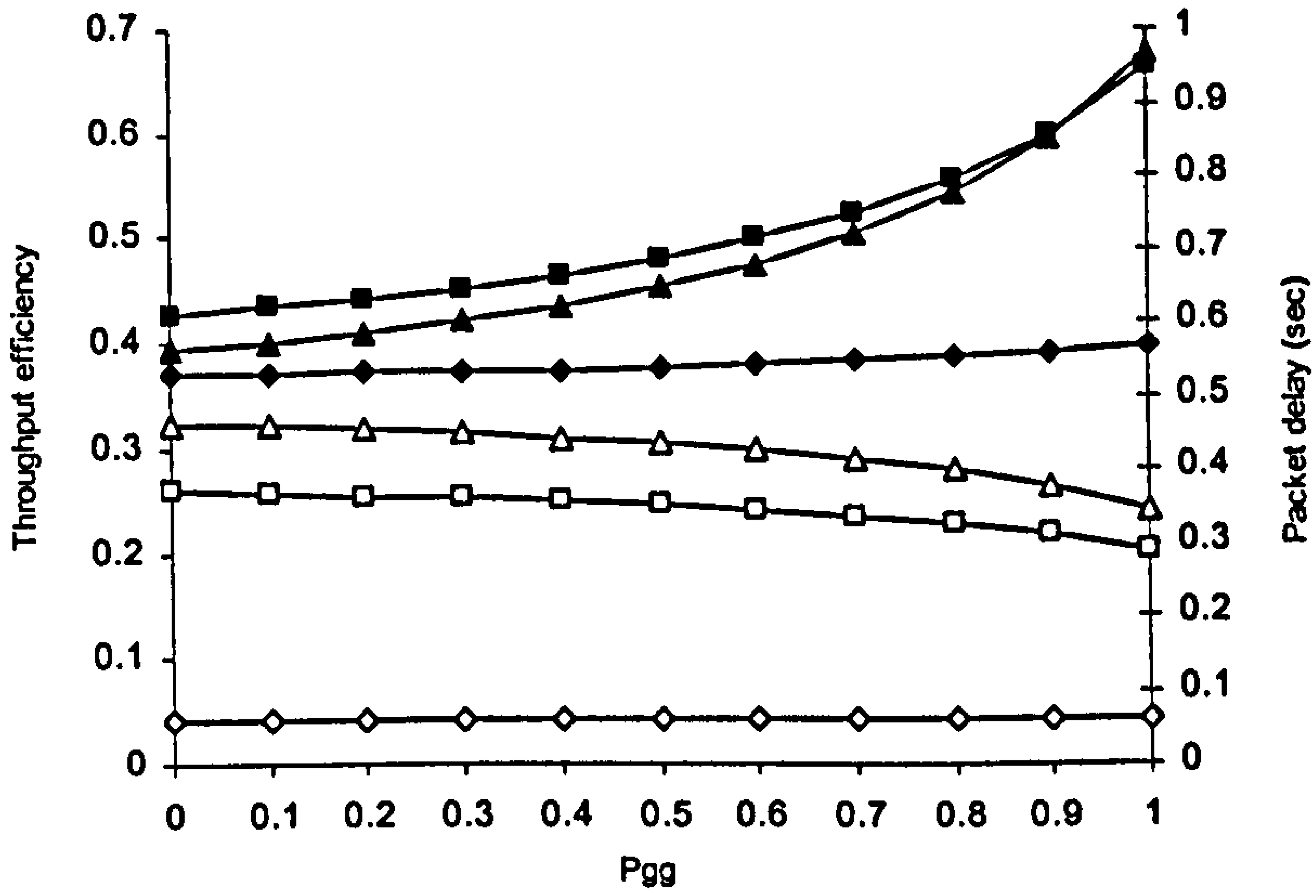


Figure 5.14 Throughput efficiency and packet delay against network size for various values of BER_B ($BER_G=10^{-6}$, l=8184 bits, p_{gg}= 0.95, p_{bb}= 0.5)

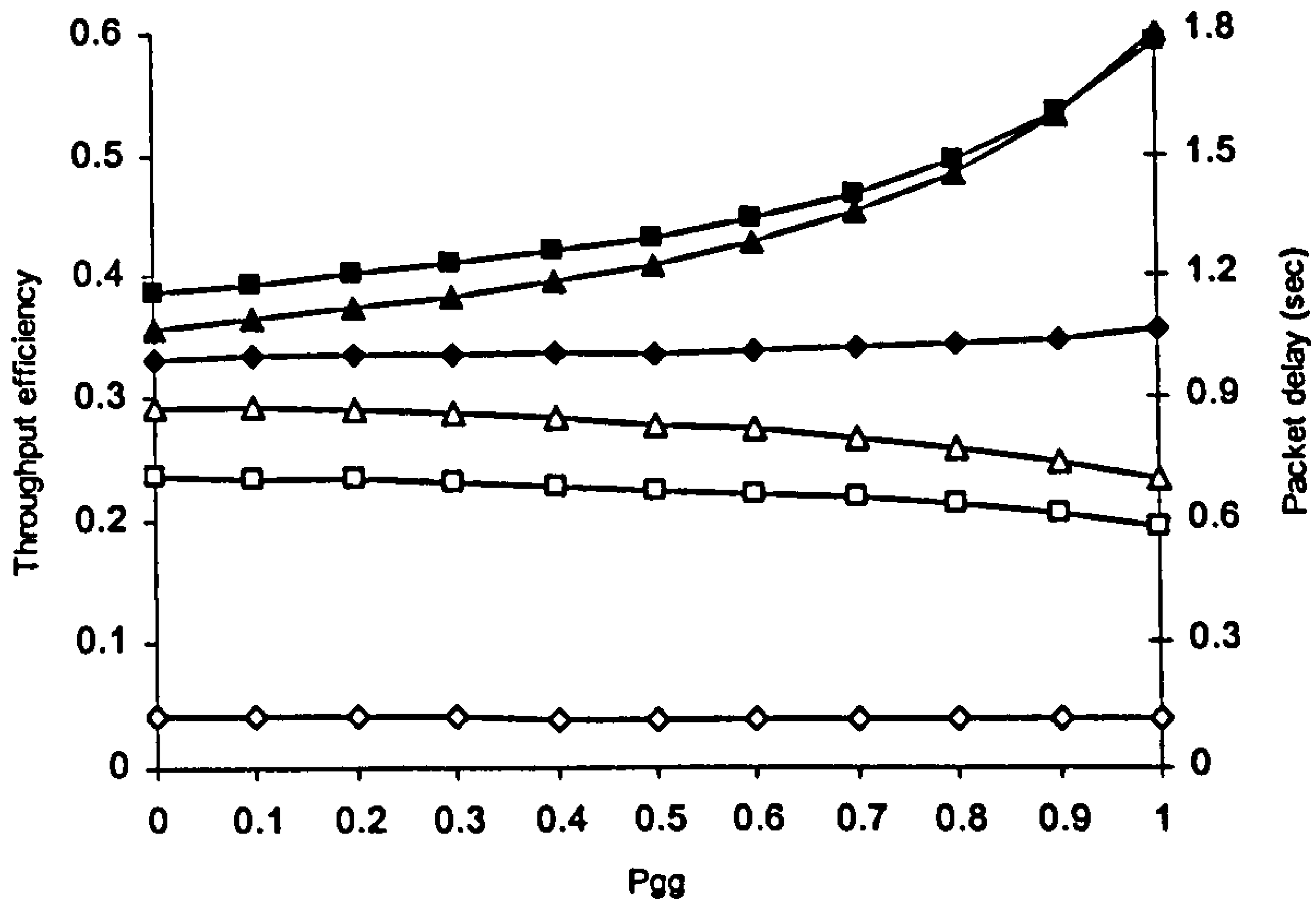
Figure 5.15 illustrates throughput efficiency and packet delay with varying the packet size as well as the time p_{gg} that will be spent in the *GOOD* state. We vary the value of p_{gg} from 0 to 1, whereas the number of contending stations is either 5 or 25 or 50. As the value of p_{gg} increases, throughput efficiency increases and packet delay decreases. The reason is that a larger p_{gg} value means less error-prone slots, leading to more successful transmissions and higher performance. The figures also show that a trade-off exists between a desire to reduce the overhead by adopting larger packet size, and the need to reduce packet error rates in the error-prone environments by using smaller packet length. In fact, when $p_{gg} < 0.9$, the best throughput performance is achieved for packets which their size is l = 8184 bits. However, when p_{gg} becomes greater than 0.9, a larger packet size value (l = 10000 bits) maximizes throughput perfomance. In any case, a small packet size value (l = 1000 bits), limits throughput due to excessive overhead in each packet but also keeps a low packet delay value. A short packet is only preferred for more error-prone channels (small p_{gg} values). Actually, from the figures we can see that, the optimal packet length that maximizes throughput performance becomes smaller when the channel is more error-prone. Moreover, it is observed that the optimal packet length does not vary with the change of the number of contending nodes.



(a) n=5



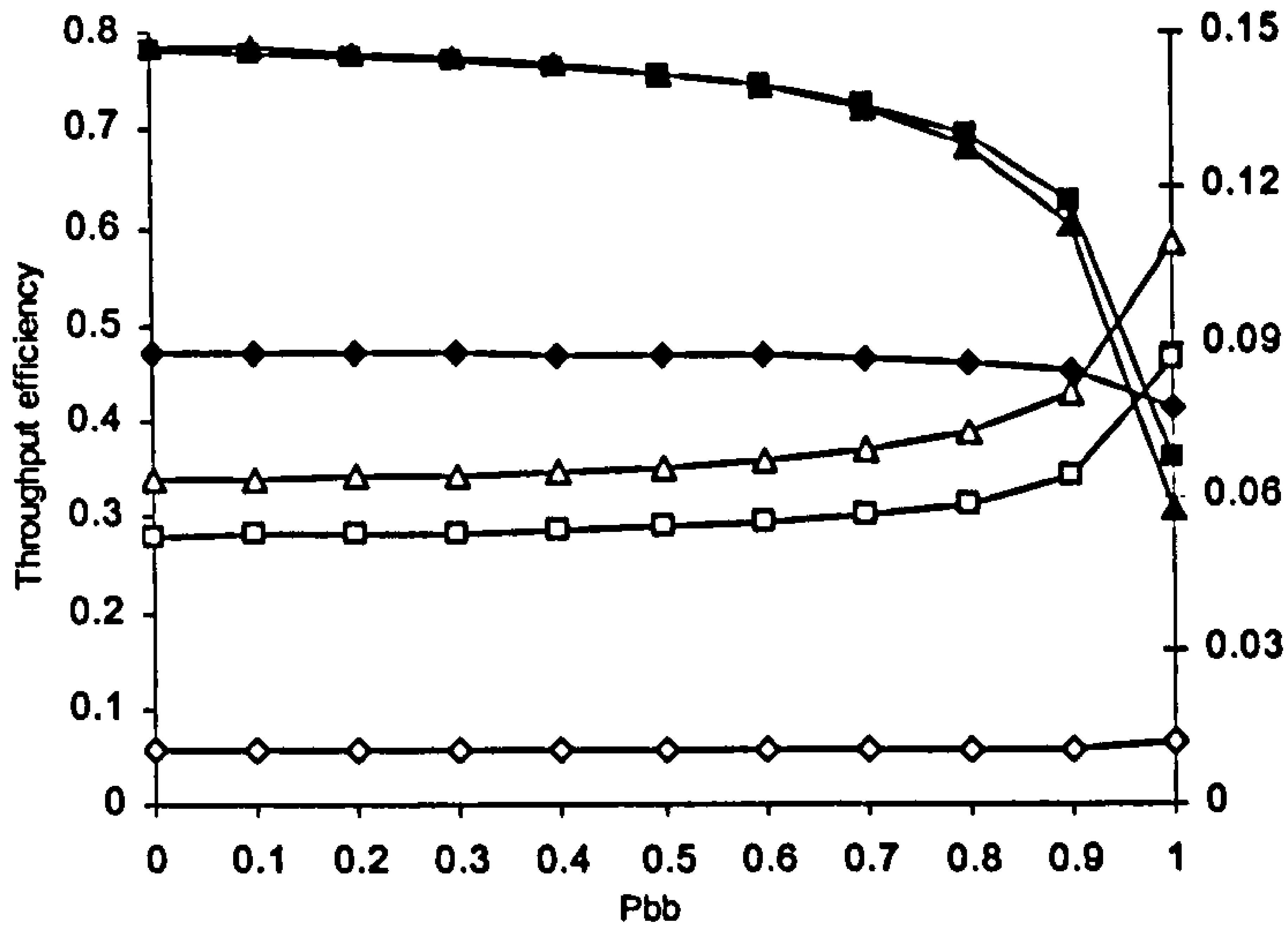
(b) n=25



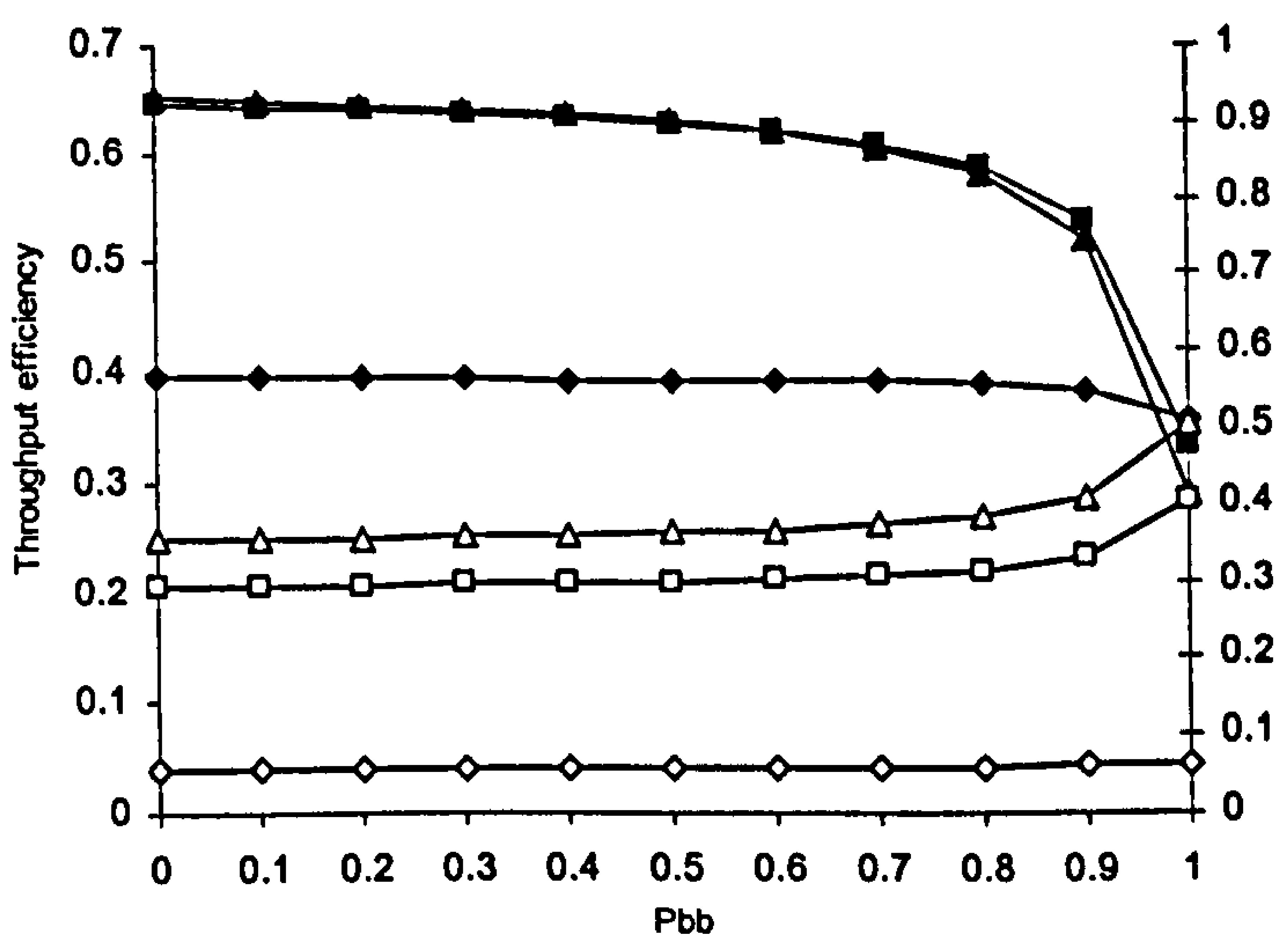
(c) n=50

◆ *Throughput efficiency, l=1000 bits* ◇ *Packet delay, l=1000 bits*
■ *Throughput efficiency, l=8184 bits* □ *Packet delay, l=8184 bits*
▲ *Throughput efficiency, l=10000 bits* △ *Packet delay, l=10000 bits*

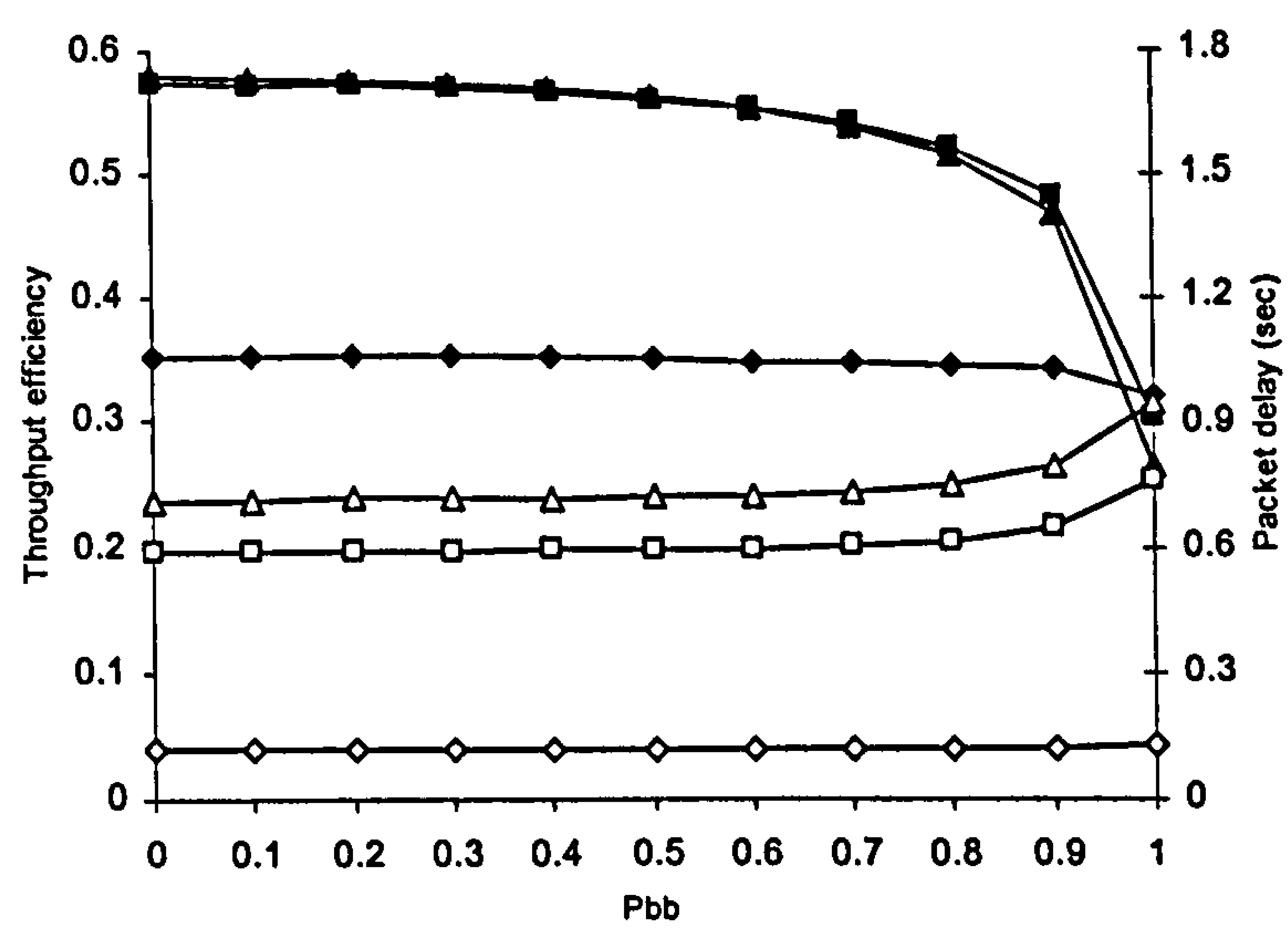
Figure 5.15 Throughput efficiency and packet delay against P_{gg} varying packet size ($BER_G=10^{-6}$, $p_{bb}= 0.5$) Proposed model



(a) n=5



(b) n=25



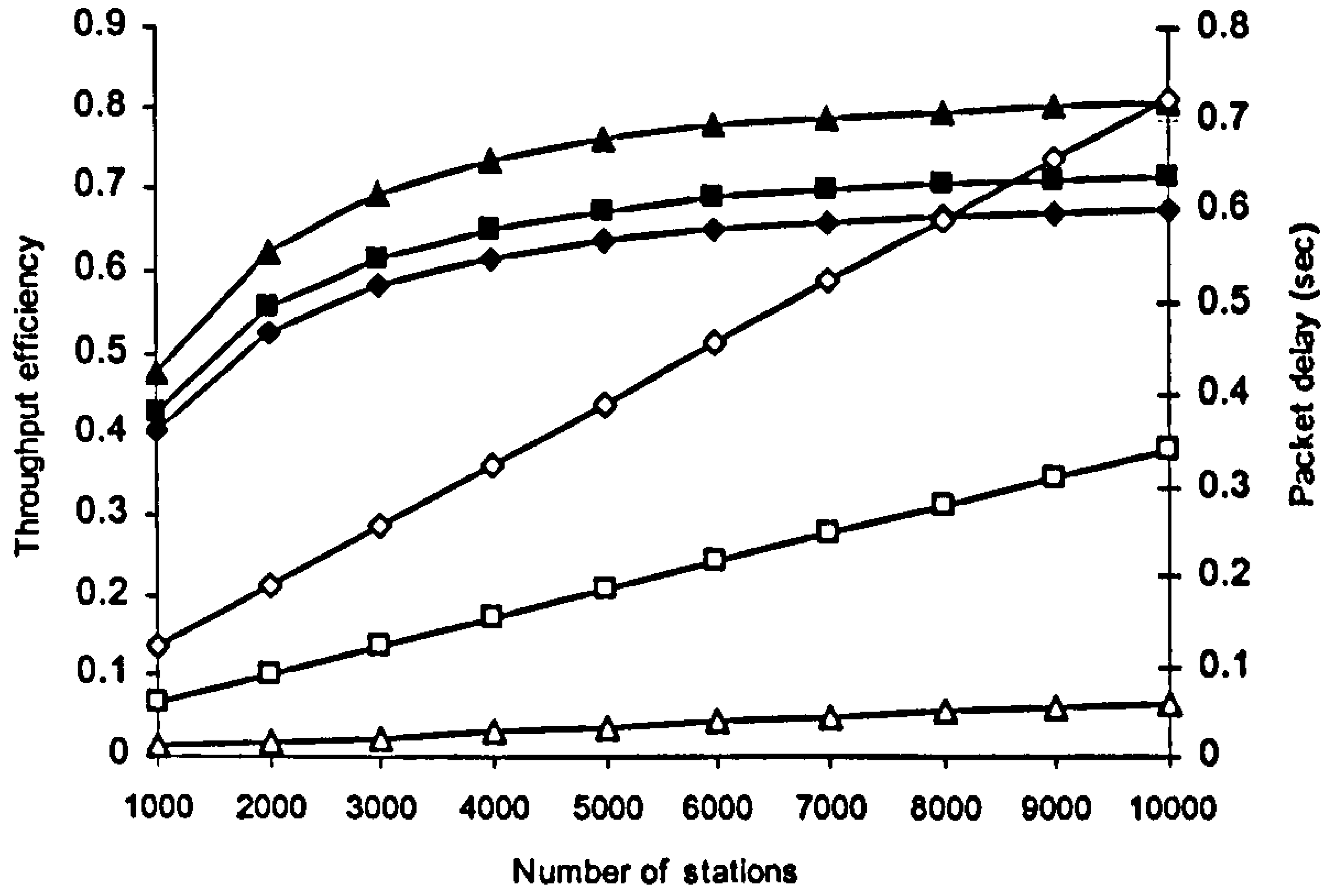
(c) n=50

◆ *Throughput efficiency, l=1000 bits* ◇ *Packet delay, l=1000 bits*

■ *Throughput efficiency, l=8184 bits* □ *Packet delay, l=8184 bits*

▲ *Throughput efficiency, l=10000 bits* △ *Packet delay, l=10000 bits*

Figure 5.16 Throughput efficiency and packet delay against P_{bb} varying packet size ($BER_G=10^{-6}$, $p_{bb}= 0.5$) Proposed model

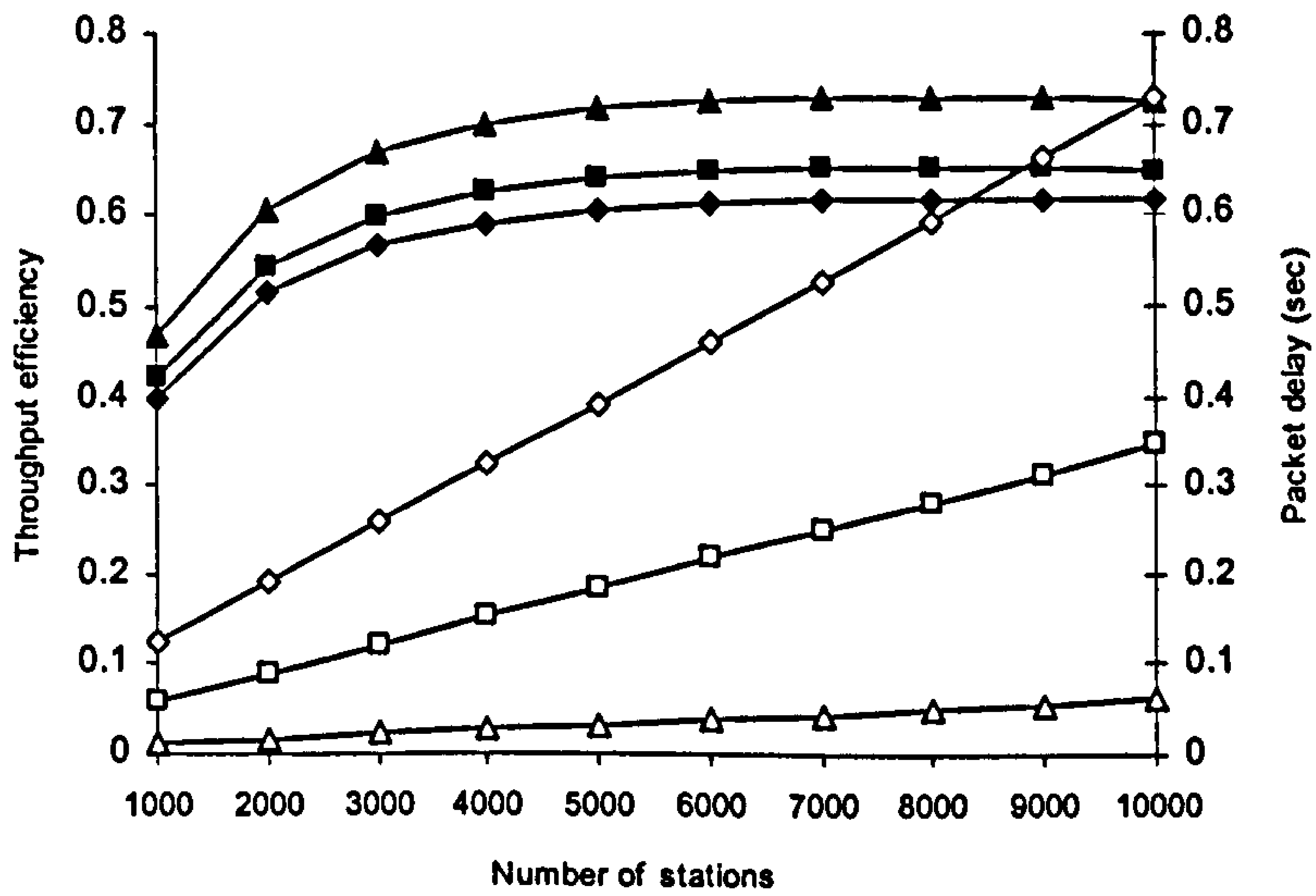


▲ *Throughput efficiency, n=5* △ *Packet delay, n=5*

■ *Throughput efficiency, n=25* □ *Packet delay, n=25*

◆ *Throughput efficiency, n=50* ◇ *Packet delay, n=50*

Figure 5.17 Throughput efficiency and packet delay against packet size varying network size ($BER_G=10^{-6}$, l=8184 bits, p_{gg}= 0.95, p_{bb}= 0.5) Gilbert – Elliot model (DIDD)



▲ *Throughput efficiency, n=5* △ *Packet delay, n=5*

■ *Throughput efficiency, n=25* □ *Packet delay, n=25*

◆ *Throughput efficiency, n=50* ◇ *Packet delay, n=50*

Figure 5.18 Throughput efficiency and packet delay against packet size varying network size ($BER_G=10^{-6}$, l=8184 bits, p_{gg}= 0.95, p_{bb}= 0.5) Proposed model for DIDD

5.6.2 DIDD protocol

Figures 5.18 and 5.19 depict how the employed packet payload size affects performance for the two employed burst error models, the Gilbert-Elliot and the proposed model, respectively. In fact, the figures illustrate that both burst error models a similar throughput efficiency and packet delay performance since the average *BER* attains a comparable value. As it was also expected, the performance of DIDD protocol is significantly sensitive to burst errors as well as to the utilized packet size in a bursty error-prone environment. The latter can be justified since when a too large packet size is employed, there is an increased need for retransmissions (since more errors are taking place), while a too small packet size is inefficient because of the fixed overhead required per packet.

CHAPTER 6

Advanced Infrared Collision Avoidance Procedures

IrDA 1.x protocol has been proven very popular and millions of devices are equipped with an IrDA 1.x infrared port. However, IrDA 1.x specifications are addressing the 'point and shoot' user model; only one pair of devices can communicate in the same infrared space and the link range is limited. The significant increase on the number of mobile devices on the market today and recent advances in infrared technology have led to the decision to address the communication requirements of a pool of users. IrDA proposed the Advanced Infrared (AIr) protocol specifications for indoor, high-speed, low cost and multipoint wireless communications.

AIr Medium Access Control (MAC) coordinates medium access by employing Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) techniques. The AIr Collision Avoidance (CA) procedures are summarized as follows. A station wishing to transmit first contends for medium access. When the medium is busy, the competing station for medium access first waits for the transmitting station to finish and for the beginning of the next contention period. The contention period is slotted and a contending station can only transmit at the beginning of a slot. The slot period is referred to as Collision Avoidance Slot (CAS). To minimize collision probability with other contenting stations, a station first selects a random number of CAS to wait before transmitting and assigns the selected value to its CAS timer (CT). This integer random number is uniformly chosen in the range (0, CW-1), where CW is the current Contention Window (CW) size. The CW size represents the range the random number is picked from and competing stations may utilize different CW values during the contention period. If during the station's deferral period another transmission is observed, the station freezes its CAS timer and decreases it again when the on-going transmission is over and the next contention period is started. The station transmits when its CAS timer reaches zero. Based on AIr CA procedures, the current model considers a CW increase by 4 after a collision and a CW decrease by 4 after a successful reservation [66][125]. Moreover, the AIr standard specifies a CW lower limit of 8 and upper limit of 256.

This chapter presents an alternative and intuitive derivation of the AIr performance analysis previously based on a 2-dimensional Markov chain model [125]. By assuming that the probability of collision is constant, a 1-dimensional Markov chain model can be constructed instead of the 2-dimensional model. This new approach considerably simplifies previous analyses and can be applied to any CSMA/CA MAC protocol. We also extend the throughput performance analysis presented in [122] in order to calculate the average packet delay for the AIr protocol by deriving simple mathematical equations. The key approximation of our model is the assumption that a reservation attempt collides with a constant probability, which is independent of the number of collisions and successful reservations the station has experienced in the past[1]. By comparing analytical with simulation results we present evidence that the derived mathematical model provides extremely accurate results for AIr packet delay performance. Utilizing the proposed analysis, an extensive AIr packet delay evaluation is carried out by taking into account all the factors and parameters that affect protocol performance. Finally, we propose suitable values for both backoff and protocol parameters that reduce average packet delay and, thus, maximise performance.

The outline of this chapter is as follows. Section 6.1 introduces the AIr protocol architecture and section 6.2 presents the frame formats supported by the AIr MAC layer. Section 6.3 provides information about the AIr MAC transfer schemes including the Reserved and Unreserved transfer modes of the protocol. Section 6.4 discusses the collision avoidance procedures of the AIr protocol. Section 6.5 presents the assumptions and protocol parameters utilized in our analysis. Moreover, in this section a simple and accurate analytical model is derived to calculate the performance of the CA procedures of the AIr protocol assuming a finite number of stations. The model is based on the *CW* adjustment procedures proposed in AIr LM specification [66]. The model is validated by comparing analytical with simulation results in section 6.6. Section 6.7 employs the previously derived mathematical analysis and provides an extensive performance evaluation of AIr performance for SDATA frame employment assuming an error free channel and no Repetition Rate implementation (*RR*=1).

---

[1] An analytical model based on the same assumptions for the exponential backoff adjustment algorithm of the IEEE 802.11 protocol was presented in chapter 3.

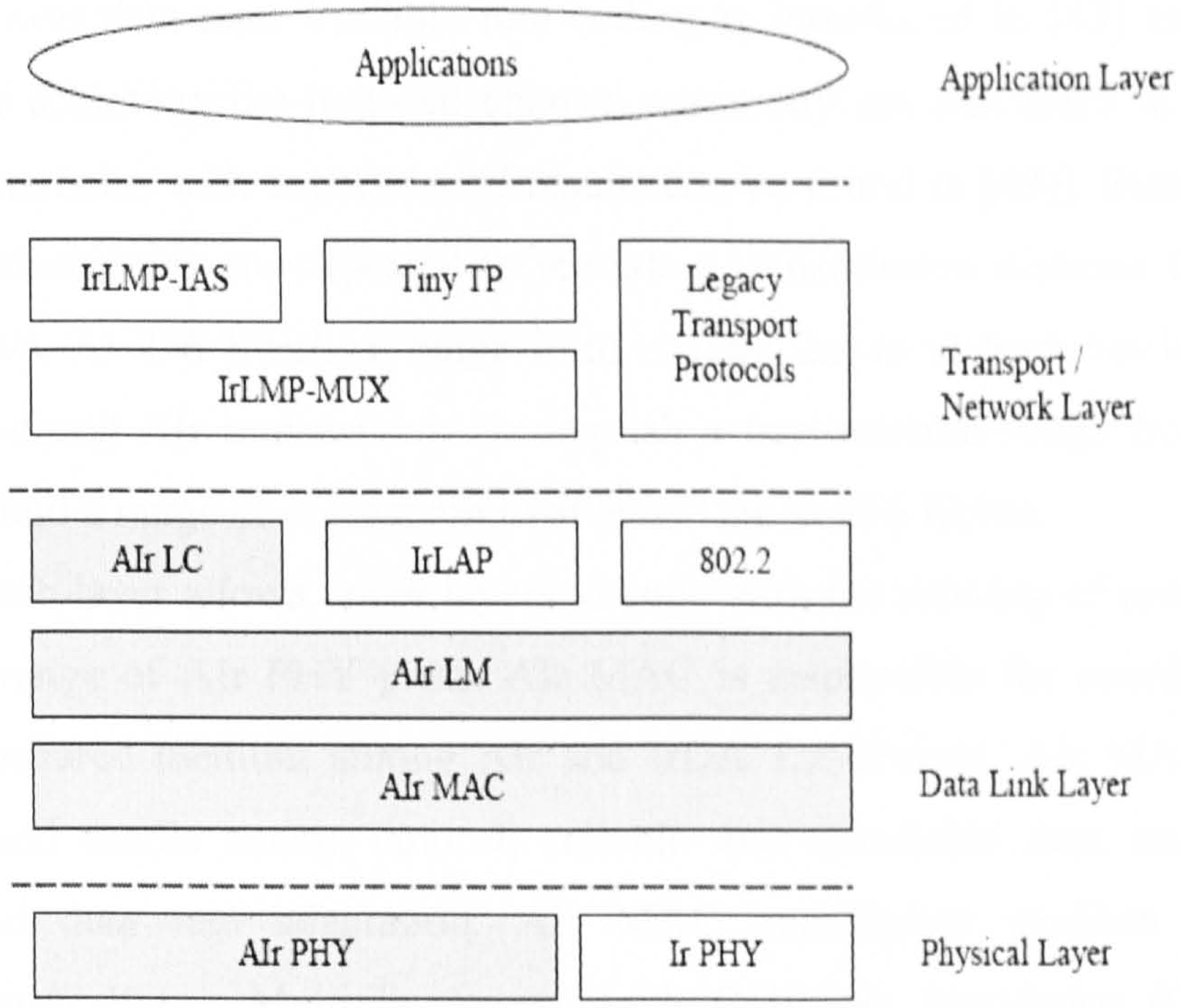


Figure 6.1 The AIr Protocol stack

6.1 Architecture overview

The primary goal in developing AIr specifications was to introduce multipoint connectivity and, at the same time, preserve the investment in upper layer applications by making certain that existing IrDA 1.x applications will be able to utilize the proposed extensions in lower layers. Figure 6.1 presents the overview of the AIr protocol stack. A new physical layer, AIr PHY [63], is introduced and the IrDA 1.x IrLAP layer is split into three sub-layers, the AIr Medium Access Control (MAC) [65], the AIr Link Manager (LM) [66] and the AIr Link Control (LC) [67] sub-layers. IrDA 1.x and AIr LC procedures for establishing device-to-device connections is transparent to network layer entities.

AIr PHY is developed employing line of sight infrared ports operating at a wide-angle of ±60 to ±75 degrees, compared to narrow-angle ±15 to ±30 degrees for the IrDA 1.x, in order to achieve multipoint connectivity. AIr utilizes one common modulation format for all supported data rates. This format is defined as four-slot Pulse Position Modulation with Variable Repetition Rate (RR) encoding (4PPM/VR) with a base data rate of 4 Mbit/s. Lower data rates (up to 256Kbit/s) may be utilized by repeating the transmitted symbols *RR* times at the base rate. The introduced redundancy improves link signal to noise ratio (SNR) and provides additional transmission range at the expense of a lower data rate. Variable rate coding is introduced in [43] and physical layer issues for achieving the required channel symmetry are discussed in [106]. AIr physical characteristics with experimental results can be found in [43]]. Standard range (S-class) AIr transceivers are expected to provide a transmission distance from 1m to 2.5m at 4 Mbit/s. At 256 Kbit/s, a range from at least 2m to at least 5m is achieved. Long-range (L-class) AIr transceivers accomplish a transmission range from 2.5m to 6m at 4 Mbit/s and a range of at least 5m to at least 12m at 256 Kbit/s.

AIr MAC sub-layer allows upper layers to cope with the relaxing of restrictions on the angle and range of AIr PHY ports. AIr MAC is responsible for coordinating the access to the infrared medium among AIr and IrDA 1.x devices. AIr MAC supports reservation based media access control, reliable and unreliable data transfer, data sequencing and data rate adaptation. AIr MAC coordinates medium access by employing Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) techniques. AIr LM is the upper layer client for the AIr MAC and allows multiplexing of multiple different LC clients. It also provides dynamic addressing, station grouping and priority and non-priority data channels. Dynamic addressing is used to cope with MAC address conflicts and station grouping is utilized to enable multicast transmissions. In requesting a reservation from the AIr MAC it is the responsibility of the LM to define the set of CAS (Collision Avoidance Slot) values to be used by the MAC. AIr LC supports connections to multiple devices. AIr LC is a derivative of the widely used HDLC protocol operating at the Asynchronous Balanced Mode of the protocol. AIr LC uses balanced stations (i.e. does not assign primary and secondary roles to communicating devices). It supports error detection and recovery services, address conflict resolution procedures and guaranteed data delivery services.

6.2 AIr MAC frame formats

The variable Repetition Rate (*RR*) values that are supported by AIr PHY and MAC layers are presented in Table 6.1. The receiver monitors channel quality and advises the transmitter to implement a suitable *RR*. The transmitter repeats the symbols it transmits *RR* times to increase the symbol capture probability at the receiver side. *RR* coding is very suitable for Pulse Position Modulation (PPM) that AIr protocol utilizes. A higher *RR* is used to achieve a better SNR as well as to reach a station that is far away from the transmitter (by trading speed for range).

AIr MAC utilizes 12 packet types in total, which are given in Table 6.2. Two general classes are defined; the reservation control packets (used to contend, initiate and terminate reservations) and the data transfer packets (used to transfer payload data). Figure 6.2 illustrates the general format of AIr packet. The AIr MAC packet definitions are shown in detail in figure 6.3. The Preamble (PA) field is transmitted for carrier sensing, symbol clock synchronisation and chip clock phase acquisition by the phase locked loop (PLL). The Synchronisation (SYNC) field qualifies the carrier detection and allows exact identification of the beginning of the robust header element. Both PA and SYNC fields actually allow the receiver to detect the beginning of an incoming packet. The Robust Header (RH) field contains AIr PHY and AIr MAC information and is always transmitted using the maximum allowable Repetition Rate encoding (RR=16) to providemaximum detection sensitivity. Thus, all stations capable of interfering with the current transmission refrain from transmitting. The Main Body (MBR) field

Repetition Rate	**Data rate**
RR=1	4 Mbit/s
RR=2	2 Mbit/s
RR=4	1 Mbit/s
RR=8	512 Kbit/s
RR=16	256 Kbit/s

Table 6.1 AIr Repetition Rate (RR) values

Type	**Description**
RTS	Request To Send (reservation)
CTS	Clear To Send (reservation)
SOD	Start Of Data (reservation)
EOB	End Of Burst (reservation)
EOBC	End Of Burst Confirmed (reservation)
UDATA	Unreserved data packet (data transfer)
DATA	Reserved data packet (data transfer)
SDATA	Reserved data packet with sequencing (data transfer)
ADATA	Reserved data packet with acknowledgment (data transfer)
ACK	Acknowledgment packet (data transfer)
SPOLL	Sequenced poll packet (data transfer)
SACK	Sequenced acknowledgment (data transfer)

Table 6.2 AIr MAC packet format types

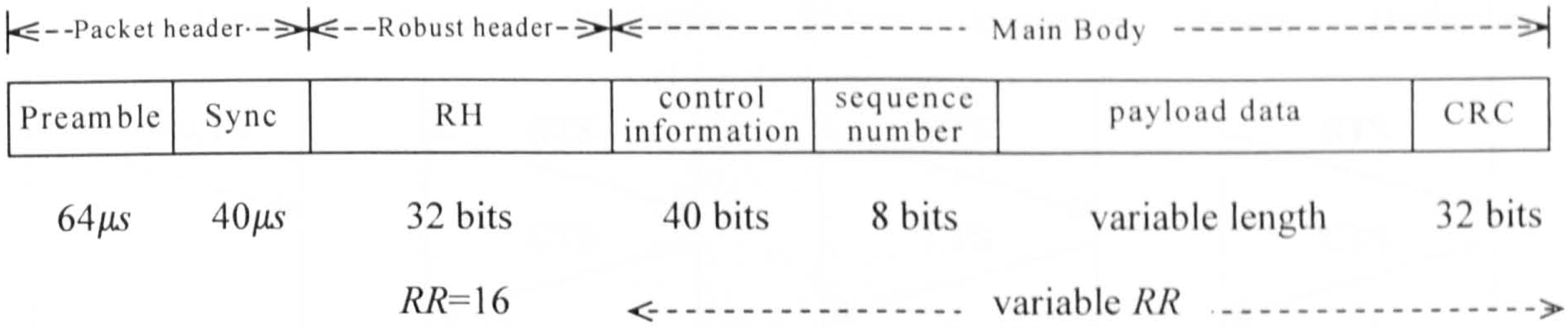


Figure 6.2 AIr general packet format

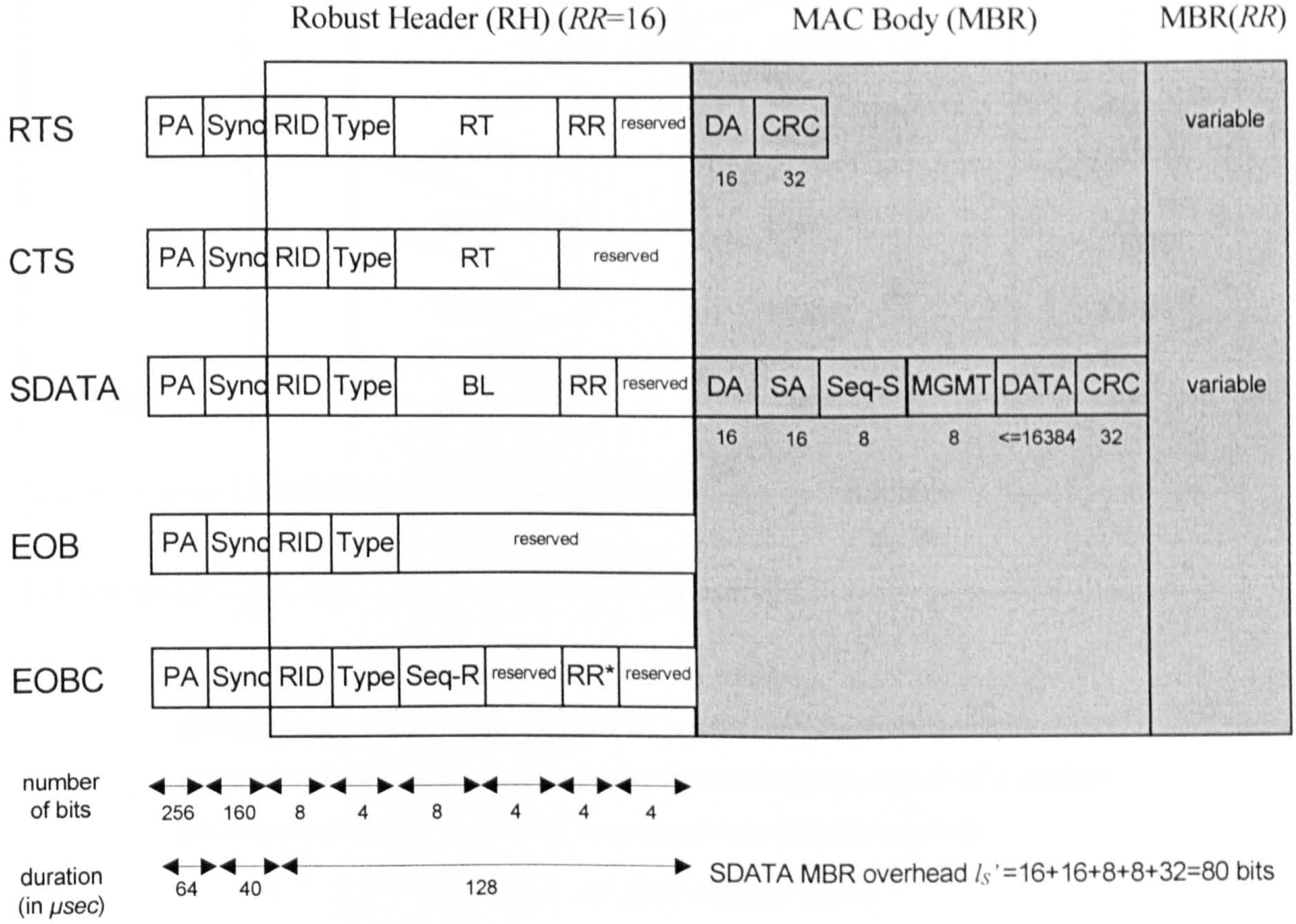
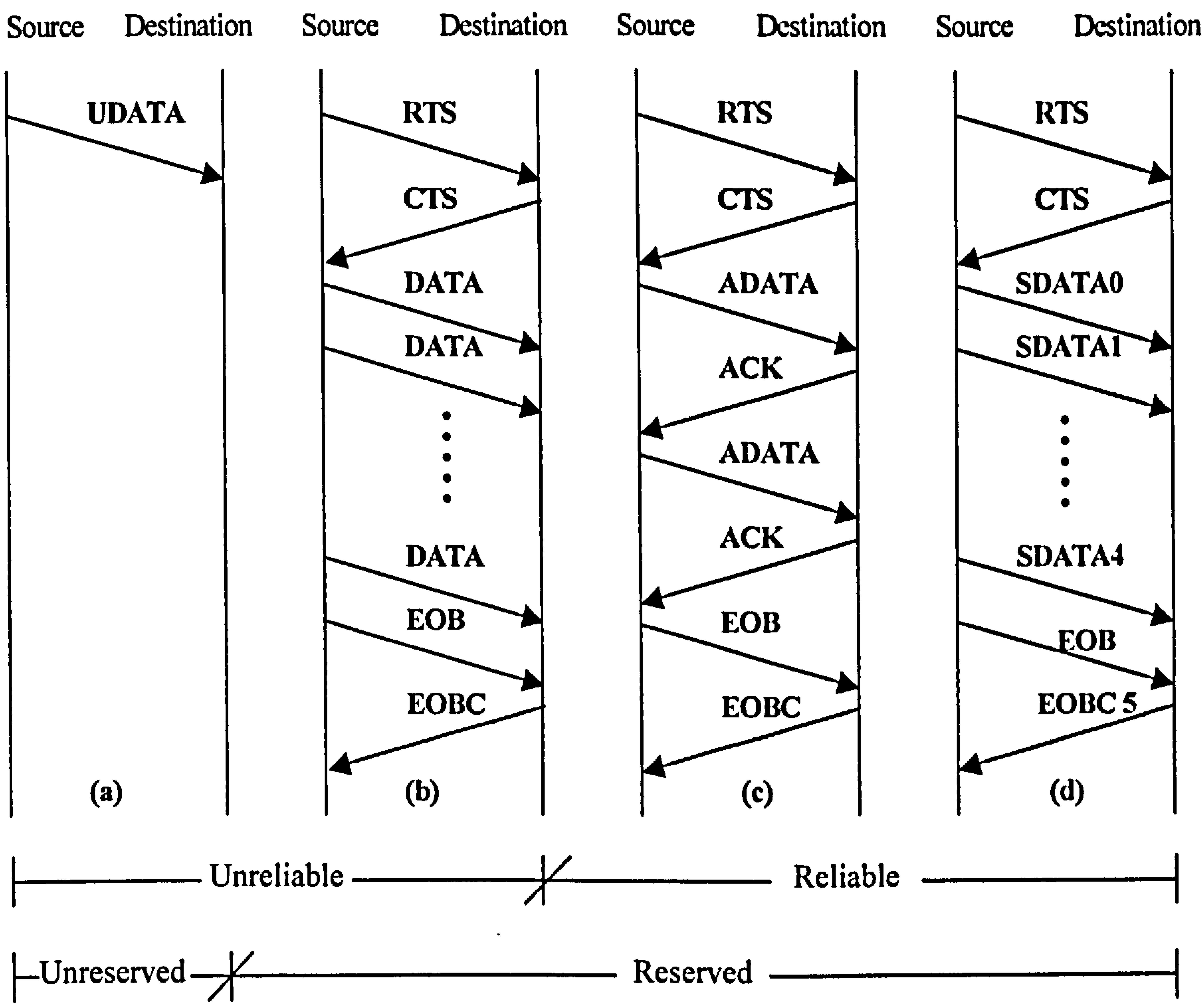


Figure 6.3 AIr MAC packet definitions

contains upper layer and AIr MAC information and is transmitted using variable *RR* shown in table 6.1. MBR contains payload data and has a variable length. PA, SYNC and RH fields are present in all AIr MAC packet types. MBR is not present in some packet types; in this case the RH field is not protected by a CRC because it is transmitted using *RR*=16. When present, it is followed by a Cyclic Redundancy Check (CRC) field protecting both the RH and MB fields. The transmitter decides the suitable *RR* for specific transmission according to its evaluation the link quality to the receiving station. A receiving station also recommends RR values to the transmitter based on its evaluation of link quality.



(a) Unreserved transfer mode with UDATA packet

(b) Reserved transfer mode with DATA packet (no acknowledgement)

(c) Reserved transfer mode with packet Acknowledgement (ADATA packet)

(d) Reserved transfer mode with Sequenced data (SDATA packet)

Figure 6.4 AIr MAC transfer modes

6.3 AIr MAC transfer modes

While a transmitting station is sending a packet, it 'blinds' its own receiver such that it cannot receive remote infrared pulses. The receiving station waits a minimum Turn-Around Time (TAT) to allow for the transmitter's receive circuitry to recover and responds with a CTS packet. According to AIr MAC specification, TAT duration is 200μs being significantly higher than the TAT period of similar WLAN protocols, such as the IEEE 802.11 [135].

AIr MAC defines reliable and unreliable data transfer modes as well as reserved and unreserved media access shown in figure 6.4. Unreliable transfer modes (figures 6.4(a) and 6.4(b)) guarantee the transmission of user data but not the delivery as no acknowledgement is provided to indicate correct packet reception. Reliable modes guarantee correct packet reception as an acknowledgement is provided for every data packet (figure 6.4(c)) or for a packet burst (figure 6.4(d)).

In reservation media access schemes (figures 6.4(b),(c),(d)), the transmitting station reserves the medium for the duration contained in the Reservation Time (RT) field of the RTS packet it transmits. After a TAT delay, the receiving station responds with a CTS packet and echoes the reservation period in the RT field of the CTS packet. As the RT field is contained in RH, it is always transmitted using maximum RR=16 to ensure maximum coverage. Thus, even stations being able to hear only the RTS or only the CTS packet defer transmission for the entire reservation period. Moreover, the RTS/CTS scheme is employed to address the hidden station (a station not being able to hear the transmitter or the receiver) problem at the expense of the time required for transmitting the RTS and CTS packets. When the transmitter receives the CTS packet, waits for a TAT delay and initiates a window packet transmission. After the last data packet is transmitted and before the reservation time expires, the transmitter requests termination of current reservation by transmitting an End Of Burst (EOB) packet. The receiver waits a TAT period and confirms termination of current reservation by responding with an End Of Burst Confirm (EOBC) packet. As with RTS/CTS exchange, a station receiving the EOBC or the EOB packet realizes that the current reservation is over and that it is able to contend for the medium again.

6.3.1 Unreserved transfer mode

Unreserved data transfer mode (figure 6.4(a)) transmits only one UDATA data packet to a multicast or broadcast (i.e. all devices) address using RR =16 to ensure maximum coverage. Note that the Unreserved mode incurs the least overhead since does not reserve the infrared medium by employing the RTS/CTS packet exchange and is unreliable because no acknowledgment is received. If a packet collision occurs due to two or more stations choosing the same time slot, it will be the responsibility of upper protocol layers to perform error recovery.

6.3.2 Reserved transfer modes

6.3.2.1 Reserved transfer mode with DATA frame (no acknowledgment)

Reserved transfer mode with no acknowledgement (figure 6.4(b)) uses the RTS/CTS reservation scheme to reserve the medium, transmits a window of DATA packets in a successful reservation, terminates the reservation using the EOB/EOBC packet exchange but it is still unreliable since no acknowledgement is exchanged. When one of the previous two data transfer modes is used (figures 6.4(a) and 6.4(b)), a MAC successful transmission indication to LM layer means that the packets are sent and not that the packets are correctly received. In this case, the AIr LC layer implements a retransmission scheme to handle frame errors.

6.3.2.2 Reserved transfer mode with acknowledgment and Reserved transfer mode with Sequenced data

Reserved transfer mode with acknowledgement (figure 6.4(c)) and Reserved transfer mode with Sequenced data (figure 6.4(d)) also employ the RTS/CTS reservation scheme. In the first transfer mode ADATA frames carry the payload data and a window of frames is transmitted after a successful reservation. Successful ADATA frame reception is based on an immediate acknowledgement (ACK) packet ACK (acknowledgement) frame transmitted by the receiver. Finally, the reservation is terminated by using the EOB/EOBC control frame exchange. In the second transfer mode payload data is transmitted data utilizing SDATA frames. An SDATA frame contains its sequence number in the Seq-S field. This mode terminates a reservation by using the EOB/EOBC frame exchange but, in this case, the EOBC frame contains the next frame sequence number expected by the receiver in the Seq-R field (EOBC 5 in figure 6.4(d)). Thus, the transmitter is informed of the correctly in-sequence received frames when the reservation terminates. For these two data transfer modes, a MAC successful transmission indication to LM layer means that the packets are correctly received. Since this work studies the performance of the reserved access reliable sequenced transfer mode (SDATA), the format definition of the AIr MAC packets used in SDATA was shown in figure 6.2.

6.4 Collision avoidance procedures

AIr MAC employs Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) techniques to minimize collision probability. A station wishing to transmit and regardless of the transfer mode it employs, it first invokes the Collision Avoidance (CA) procedures in an effort to minimize collisions with other stations. In the SDATA transfer mode, which is presented in figure 6.5, a contending station always invokes the

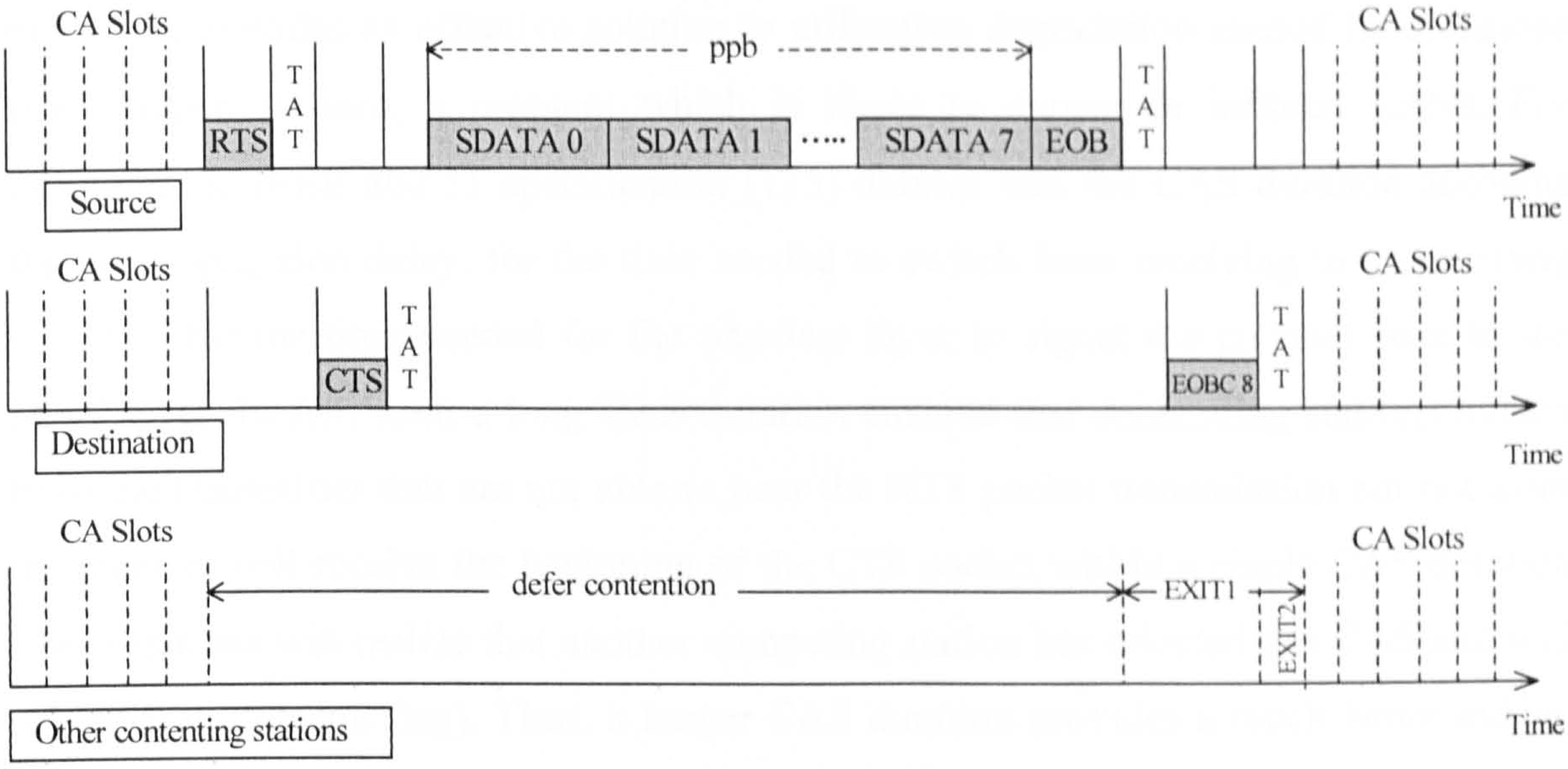


Figure 6.5 Reserved access scheme with Sequenced transfer mode (SDATA packets)

CA procedures before an RTS packet transmission. Contention periods start a TAT period[2] after the EOBC frame that terminates a reservation. Contention periods end when a non-colliding RTS frame is transmitted. AIr MAC specification ensures that even stations hidden from the transmitter or the receiver are synchronized in contending for the medium at the beginning of the next contention period after a successful RTS/CTS medium reservation. Synchronization is accomplished by means of the EXIT1 and EXIT2 timers. EXIT1 is started when the EOB frame is received. EXIT1 expires at the beginning of the next contention period synchronizing stations not receiving the EOBC frame. EXIT2 is started when the EOBC frame is received and, as it accounts for the TAT delay, it expires when the next contention period starts. Thus, stations hearing only the EOB or the EOBC frame will contend for medium access at exactly the correct time.

The contention period is slotted (see figure 6.5) and a station is allowed to transmit only at the beginning of a Collision Avoidance Slot (CAS). The CAS duration (σ) is defined as being greater than the transmission time of the RTS packet plus the TAT delay plus the time need to detect the beginning portion (PA and SYNC fields) of the responding CTS packet ($\sigma > T_{RTS}$+TAT+TT_{PA}+TT_{SYNC}). AIr MAC defines that σ=800 μs. The CAS duration, which is significantly longer than the one of similar CSMA/CA

[2] The TAT delay is required because the station that transmitted the EOBC packet should be able to receive the next packet.

protocols, provides an effective solution to utilization degradation caused by collisions from hidden stations, a problem which is likely to appear in infrared LANs. For example, the IEEE 802.11 specification [135] defines that the CAS duration accounts for the propagation delay, for the time needed to switch from receiving to transmitting state and for the time needed for the physical layer to signal the channel state to the MAC layer. In AIr, such a long CAS duration ensures that contending stations hidden from the transmitter that are not able to hear the RTS packet transmission but not from the receiver will receive the beginning of the CTS packet within a single CAS duration (these stations will realize that another competing station has selected this CAS and will refrain from transmitting). Thus, a longer CAS duration provides a much better hidden station approach but at the expense of possible performance degradation if the number of empty and colliding CAS during the contention periods is high. This number depends on the number of the competing stations and on the *CW* values used by these stations.

A competing station for medium access first senses the medium; if the medium is busy, it waits for the transmitting station to finish and for the beginning of the next contention period. The contending station then initialises its backoff counter by selecting an integer random number of CAS to defer transmission in order to minimize the collision probability with other transmissions. This backoff counter is uniformly selected in the range (0,*CW*-1) where *CW* is the current Contention Window (*CW*) size and the backoff interval is assigned to CAS timer (CT). The *CW* size values depend on the number of successful reservations and collisions that the transmitting station has experienced in the past. If during the station's deferral period another transmission is observed, the station freezes its CAS timer and restarts it again when the ongoing transmission is finished (the medium becomes free again) and the next contention period is started. When the CAS timer reaches zero, the station attempts to reserve the channel by transmitting an RTS packet. The receiving station waits a minimum Turn Around Time (TAT) to allow for the transmitter's receive circuitry to recover and responds with a CTS packet (figure 6.5). After the successful RTS/CTS exchange, the transmitting station, after a TAT delay, transmits a burst of data packets and requests termination of current reservation by transmitting an End-Of-Burst (EOB) packet. The receiving station responds with an End-Of-Burst-Confirm (EOBC) packet confirming reservation termination. The reservation time duration is echoed in the Reservation

Packet/packet element	Duration	Time (μs)
T_{PA} *(packet element)*		64
T_{SYNC} *(packet element)*		40
T_{RH} *(packet element)*		128
TT_{RH} *(packet)*	$T_{PA}+T_{SYNC}+T_{RH}$	232
T_{RTS} *(packet)*	$TT_{RH}+48/C$	244
T_{CTS} *(packet)*	TT_{RH}	232
T_{EOB} *(packet)*	TT_{RH}	232
T_{EOBC} *(packet)*	TT_{RH}	232
T_{ACK} *(packet)*	TT_{RH}	232
Turn Around Time (*TAT*)		200
CAS slot (σ)		800
EXIT1 Timer	$TAT+T_{EOBC}+TAT$	632
EXIT2 Timer	TAT	200
WFCTS timer	$\sigma - T_{RTS}$	556

Table 6.3 AIr timer durations, packet and packet element transmission times for C=4 Mbit/s

Time (RT) field of both the RTS and CTS packets. Thus, stations being able to hear only the RTS or only the CTS packet refrain from transmitting for the entire reservation period.

As it was stated earlier, AIr MAC considers synchronizing all stations contending for the medium at exactly the same time after a successful RTS/CTS medium reservation, even for stations hearing only the EOB or the EOBC packet. Synchronization is accomplished by implementing two timers (EXIT1 and EXIT2). EXIT1 time duration is defined as the TAT after the EOB plus the transmission time of the EOBC packet plus the TAT after the EOBC packet and EXIT2 is defined as the TAT delay (figure 6.5). Moreover, a contending station, after transmitting the RTS packet, starts the Wait For CTS (WFCTS) timer. If another (one or more) stations has selected the same CAS slot, it transmits an RTS packet at the same time and the reservation attempt is unsuccessful. The transmitting stations determine the resulting collision by the WFCTS timer expiration. This timer value ($T_{WFCTS} >= TAT + TT_{RH}$) expresses the amount of time a station that has transmitted an RTS packet will wait for the corresponding CTS packet. If a valid CTS has not been received within the WFCTS period, the transmitter assumes that a collision occurred and contends again for medium

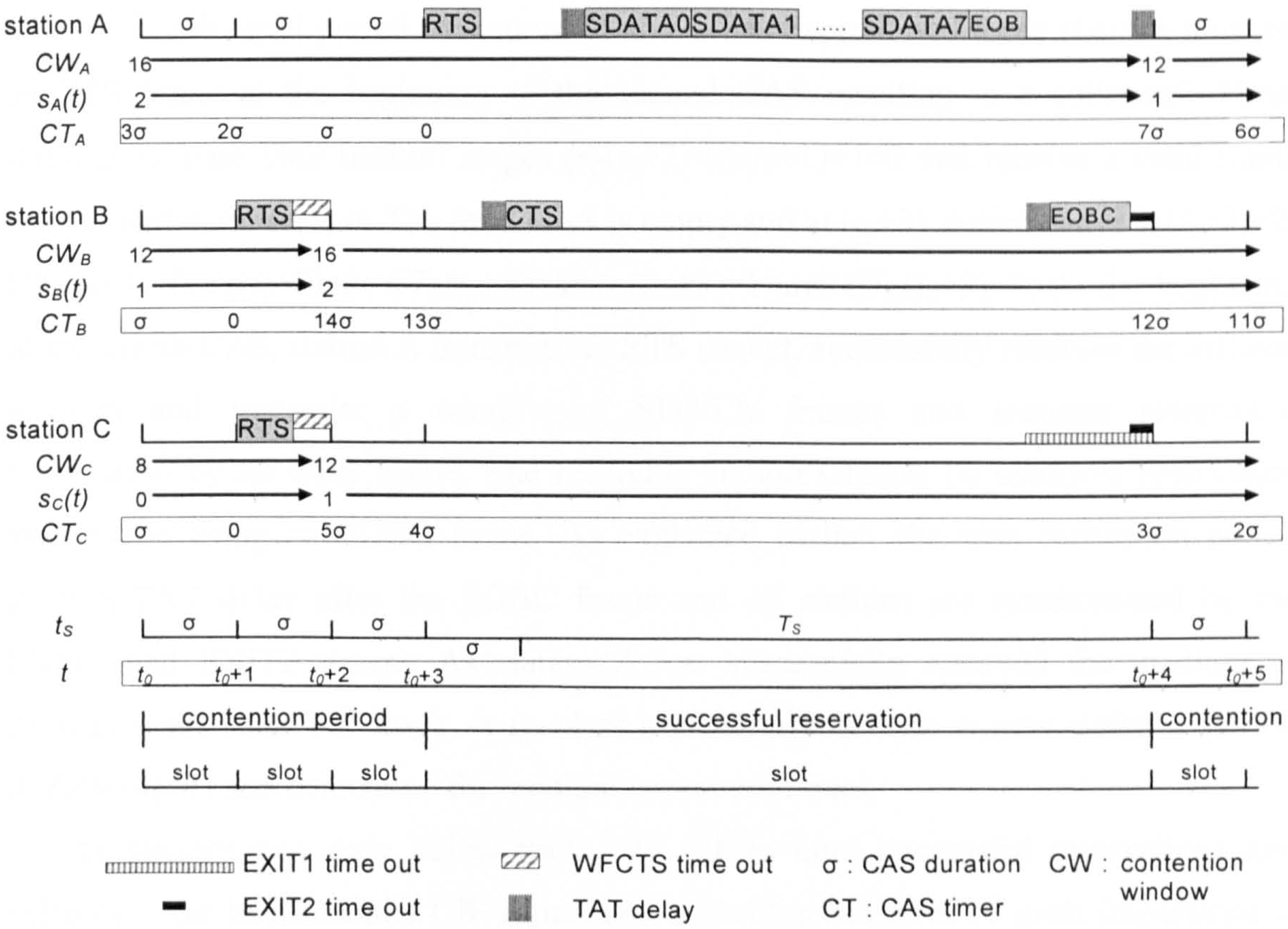


Figure 6.6 Operation of Collision Avoidance procedures

access by selecting a new CAS slot and re-attempting a reservation. To synchronize the colliding stations with the remaining stations, the WFCTS timer should expire at the end of the current time slot. The time required for transmitting control packets and packet elements as well as the implemented timers and time delay values are summarized in table 6.3 for C=4 Mbit/s.

Figure 6.6 illustrates in detail the station behavior for a LAN with three contending stations employing the Reserved transfer mode with Sequenced data. It is assumed that at the beginning of the contention ($t=t_0$), station A is at backoff stage 2 ($s_A(t_0)=2$) since its current CW value is 16 ($CW_A=16=8+4*2$). Station A also defers transmission for three CAS before transmitting an RTS packet, thus, its CT has a value of 3σ ($CT_A=3\sigma$), where σ is the CAS duration. Similarly, we assume that station B has $CW_B=12$, $s_B(t_0)=1$ and $CT_B(t_0)=1$ and station C has $CW_C=8$, $s_C(t_0)=0$ and $CT_C\ (t_0)=1$. As the first CAS is empty, all stations defer transmission, do not change backoff stage ($s_X(t_0+1)=s_X(t_0)$, $X=\{A,B,C\}$), decrement their CT value ($CT_X(t_0+1)=CT_X(t_0)-1$, $X=\{A,B,C\}$) and the slot duration is σ. As a result, at the beginning of the second CAS, $CT_A=2\sigma$, $CT_B=0$ and

CT_C=0. The deferral period of stations B and C has expired and these stations transmit an RTS frame at the beginning of the second CAS resulting in a collision[3]. These stations increase their backoff stages ($s_B(t_0+2)=s_B(t_0+1)+1=2$ and receive a valid frame and the slot duration is σ. The third CAS is empty and at (t_0+3) stations have $s_A(t_0+3)=2$, $CT_A(t_0+3)=0$, $s_B(t_0+3)=2$, $CT_B(t_0+3)=13$, $s_C(t_0+3)=1$ and $CT_C(t_0+3)=4$. At the beginning of the fourth CAS, station A transmits an RTS packet, successfully reserves the infrared medium and transmits a window of SDATA frames and requests reservation termination by an EOB frame. The receiving station (station B) confirms reservation termination using an EOBC frame. As explained earlier, the next contention period starts a TAT delay after the EOBC frame and all stations are synchronized by the EXIT1 and EXIT2 timers. As station A has successfully reserved the medium, it decreases its back off stage ($s_A(t_0+4)=1$), randomly selects a new deferral period ($CT_A(t_0+4)=7$) and contention for medium access continues.

As stations can only adjust their *CW* values after successful reservations and collisions, the implemented *CW* adjustment algorithm becomes of great importance if maximum utilization is to be achieved. Small *CW* size values result in a very high collision probability and, therefore, to low performance due to the increased number of collisions. When large *CW* size values are implemented, the increased number of empty CAS will also result in low medium utilization because a long CAS duration is defined. A station can only estimate the suitable *CW* value it should implement based on the experienced successful reservations and collisions. AIr specifications define that the AIrLM layer selects the *CW* value to be used in every reservation attempt and pass it down to the MAC layer. The AIrLM layer does not provide rules for *CW* size adjustment but suggests guidelines by utilizing a linear algorithm for incrementing and decrementing *CW* after a collision and a successful reservation attempt, respectively. This *CW* size adjustment can be achieved since the transmitting station always 'remembers' the *CW* value used in the previous reservation attempt. If this attempt was successful, *CW* is decreased by 4 (see station A in figure 6.6); if it resulted in a collision, *CW* is increased by 4 (stations B and C in figure 6.6). A minimum *CW* value of 8 (lower limit) and a maximum *CW* value of 256 (upper limit) are also defined.

[3] As no CTS packets are generated, stations B and C realize the collision by the expiration of their WFCTS timers at the start of the next CAS; this actually synchronizes stations B and C with station A.

6.5 Analytical model

Our mathematical analysis consists of three parts. The first part presents the assumptions and the parameters that we utilize in our analysis. The second part considers the behavior of a single station to compute the conditional probability p and the stationary probability τ that a station transmits a RTS packet in a randomly chosen CAS for a network of n stations. Note that both probabilities are independent of the reserved access scheme employed by the stations. Finally, in the last part, by examining the events that can occur in a randomly chosen CAS, we derive the average delay performance of packets being transmitted by the AIr protocol; simple equations calculate packet delay as a function of probabilities p and τ. The key assumption used in our model is that an RTS transmission always collides with probability p regardless of the CW value used to select the deferral period for the reservation attempt.

A. Analysis assumptions and parameter definitions

This work concentrates on the packet delay performance for a fixed number of stations under saturation conditions. In saturation conditions, every station has immediately a burst of packets ready for transmission, after the completion of each successful burst transmission. In other words, the transmission queue for every station is always non-empty. All burst of packets are considered "consecutive"; each one needs to wait for a random backoff time before being transmitted. All stations always employ the Reserved transfer mode with Sequenced data although the analytical model can be easily altered to evaluate performance of the remaining reserved transfer modes supported by the AIr MAC (figure 6.4(d)). After a successful reservation attempt, a station transmits packets per burst (*ppb*) of fixed payload size of l bits at a fixed data rate of C Mbit/s.

We also assume ideal channel conditions; a non-colliding packet is always received error free to all network stations. Current analysis also assumes that reservation control packets (RTS, CTS, EOB and EOBC) are always transmitted error free. This is a realistic assumption because, since control packets CTS, EOB and EOBC contain only an RH portion which is transmitted using the maximum repetition rate RR=16 to minimize transmission errors. RTS control packet has also an MBR field consisting of only 48 bits, which is transmitted using variable RR. This MBR length is extremely small for the expected link quality and the assumption that the RTS packets are always transmitted error free also holds true. Moreover, since we do not consider channel bit errors, a RR increase resulting in higher symbol capture probability at the receiver is not considered. The MB of all data packets are always transmitted in the same RR and the RH portion is transmitted in the protocol suggested RR value of 16. We also assume that the one-way propagation delay is very small and can be safely neglected due to the fact that the considered indoor links operate at very short distances. Moreover, our analysis also assumes that there are no hidden stations. Thus, all stations will always receive the RTS and the CTS packets of a successful reservation. Therefore, there is no fairness problem as all stations have an equal chance to reserve the infrared medium.

B. Calculation of the RTS transmission probability

As explained earlier, stations operate in saturation conditions; a station has immediately a window of frames ready for transmission after it successfully captures the channel and transmits a window of frames. Time scale t is also slotted and t takes only integer values; t represents the beginning of a CAS and t+1 represents the beginning of the next CAS. Stations increment t at the beginning of a CAS. We will use the term 'slot' to denote an increase in this discrete time scale (can be also defined as the period at the end of which the backoff counter changes value). If the CAS is empty (the medium is idle) or contains a collision, an increase in the discrete time scale t (slot) corresponds to a CAS duration (σ). However, if the CAS contains a successful reservation, a slot corresponds to the total time required for a transmission of w frames during the successful reservation (see figure 6.5).

The backoff counter for every station depends on the collisions and on the successful reservation attempts experienced by the station in the past. The CSMA/CA protocol procedure specifies that before transmitting each station selects a random value for its backoff counter in the range [0, W-1]. If the reservation attempt failed (the RTS packet collided), then the AIr protocol employs the linear backoff i.e. the next backoff value will be selected in the range [0, (W+4)-1] and so forth. We define for convenience $W=CW_{min}$. Let m be the 'maximum backoff stage' defined as $CW_{max}=W+4m$. Since a station may be in stage $i \in [0,m]$, we adopt the following notation:

$$W_i = W + 4i, \quad i \in (0,m) \tag{6.1}$$

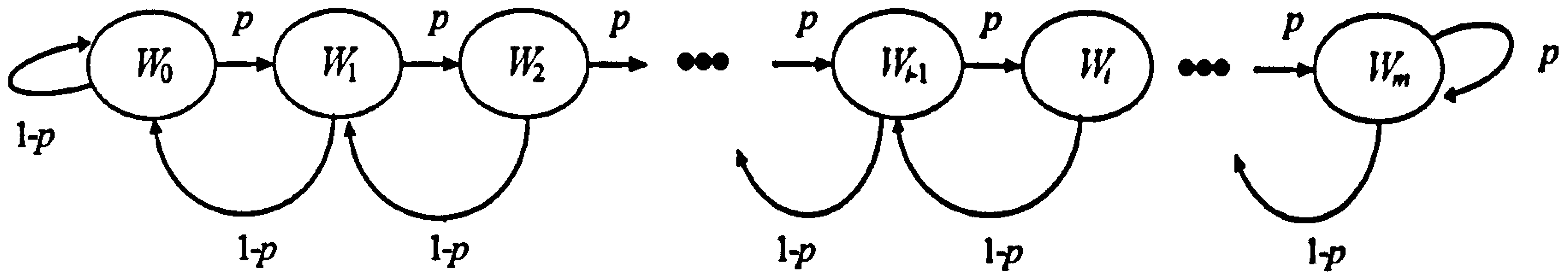


Figure 6.7 Markov chain model for backoff CW

where i is defined as the backoff stage that identifies the number of retransmissions suffered by a RTS packet According to the definition, we have $W=W_0=CW_{min}$ and $W_m=CW_{max}$. AIr standard specifies that CW_{min}=8, CW_{max}=256 and m=62.

We use $s(t)$ to denote the stochastic process representing the backoff stage $i \in (0, m)$ of the station at time t:

$$s(t) = i, \quad i \in [0, m] \tag{6.2}$$

As $s(t)=i$, the station's current CW value is W_i. If the backoff counter of a station has a value equal to k, where k is randomly selected in the range $(0, W_i\text{-}1)$, then the station will defer k slots before transmitting an RTS. We will use this fact to calculate the probability of a station starting transmission in a particular slot. By assuming that the probability of collision p is constant, regardless of the current backoff state Wi, the 2-dimensional Markov chain can be collapsed into a 1-dimensional model drawn in figure 6.7.

The only non-null, one-step transition probabilities of the 1-dimensional Markov chain shown in figure 6.7 are (using the short notation $P(s(t+1)=i_1 \mid s(t)=i_0)$):

$$\begin{cases} P\big(s(k+1)=i+1 \big| s(k)=i\big)=p & i=0,\dots,m-1 \\ P\big(s(k+1)=i-1 \big| s(k)=i\big)=1-p & i=1,\dots,m \\ P\big(s(k+1)=0 \big| s(k)=0\big)=1-p & i=0 \\ P\big(s(k+1)=m \big| s(k)=m\big)=p & i=m \end{cases} \tag{6.3}$$

The first and second equations in 6.3 represent the CW increase and decrease after a packet collision and a successful transmission, respectively. The third equation accounts for the fact that if the current backoff stage is 0, the CW value is not further decreased even after a successful packet transmission. Finally, the fourth equation considers that the CW is not increased after a collision if the maximum backoff stage m is reached.

If we let q_i to be the state probability of the 1-dimensional Markov chain being in state W_i, we have that:

$$q_i = \frac{p}{1-p} q_{i-1} \qquad i \in [1,m] \tag{6.4}$$

Furthermore, considering that: $q_0 = (1-p)q_0 + (1-p)q_1$, q_1 is given by:

$$q_1 = \frac{p}{1-p} q_0 \tag{6.5}$$

Since $q_1 = p \cdot q_0 + (1-p)q_2$, q_2 is given by:

$$q_2 = \left(\frac{p}{1-p}\right)^2 q_0 \tag{6.6}$$

Consequently, it can be easily shown that:

$$q_i = \left(\frac{p}{1-p}\right)^i q_0 = a^i q_0 \qquad i \in [0,m] \tag{6.7}$$

where for convenience we set: $a = \frac{p}{1-p}$ .By means of equation (6.7) and by applying the normalization factor, we will compute the value of q_0 :

$$1 = \sum_{i=0}^{m} q_i = \sum_{i=0}^{m} a^i \cdot q_0 = \frac{1-a^{m+1}}{1-a} \cdot q_0 = \frac{1-\left(\frac{p}{1-p}\right)^{m+1}}{1-\frac{p}{1-p}} \cdot q_0 = \frac{\frac{(1-p)^{m+1}-p^{m+1}}{(1-p)^{m+1}}}{\frac{1-p-p}{1-p}} \cdot q_0 \tag{6.8}$$

Finally, q_0 is calculated as:

$$q_0 = \frac{(1-2p)(1-p)^m}{(1-p)^{m+1}-p^{m+1}} \quad \text{or} \quad q_0 = \frac{1-a}{1-a^{m+1}} \tag{6.9}$$

By considering the previously calculated value of q_0, equation (6.7) becomes:

$$q_i = a^i \frac{1-a}{1-a^{m+1}} = \left(\frac{p}{1-p}\right)^i \frac{(1-2p)(1-p)^m}{(1-p)^{m+1}-p^{m+1}} \qquad i \in [0,m] \tag{6.10}$$

The average number of slots E[Y] spent in each state between transitions (after a collision or a successful transmission) is given by:

$$E[Y] = 1 + \sum_{i=0}^{m} \frac{W_i - 1}{2} \cdot q_i = 1 + \frac{1}{2}\left(\sum_{i=0}^{m} (W_i - 1) \cdot q_i\right)$$

$$=1+\frac{1}{2}\left(\sum_{i=0}^{m}W_i\cdot q_i-\sum_{i=0}^{m}q_i\right)$$

$$=1+\frac{1}{2}\frac{1-a}{1-a^{m+1}}\left(\sum_{i=0}^{m}W_i\cdot a^i-\sum_{i=0}^{m}a^i\right)$$

$$=1+\frac{1-a}{2\left(1-a^{m+1}\right)}\left(\sum_{i=0}^{m}(W+4i)\cdot a^i-\sum_{i=0}^{m}a^i\right)$$

$$=1+\frac{1-a}{2\left(1-a^{m+1}\right)}\left(\sum_{i=0}^{m}Wa^i+\sum_{i=0}^{m}4i\cdot a^i-\sum_{i=0}^{m}a^i\right)$$

$$=1+\frac{1-a}{2\left(1-a^{m+1}\right)}\left(W\frac{1-a^{m+1}}{1-a}+4\frac{a\left(1-(m+1)a^m+ma^{m+1}\right)}{(a-1)^2}-\frac{1-a^{m+1}}{1-a}\right)$$

$$=1+\frac{1}{2\left(1-a^{m+1}\right)}\left(\frac{(W-1)\left(1-a^{m+1}\right)(1-a)+4a\left(1-(m+1)a^m+ma^{m+1}\right)}{(1-a)}\right)$$

$$=1+\frac{(W-1)}{2}+\frac{1}{2\left(1-a^{m+1}\right)}\left(\frac{4a\left(1-(m+1)a^m+ma^{m+1}\right)}{(1-a)}\right)$$

$$=\frac{W+1}{2}+\frac{4a\left(1-(m+1)a^m+ma^{m+1}\right)}{2\left(1-a^{m+1}\right)(1-a)}$$

$$=\frac{W+1}{2}+\frac{4\frac{p}{1-p}\left(1-(m+1)\left(\frac{p}{1-p}\right)^m+m\left(\frac{p}{1-p}\right)^{m+1}\right)}{2\left(1-\left(\frac{p}{1-p}\right)^{m+1}\right)\left(1-\frac{p}{1-p}\right)} \tag{6.11}$$

After some algebra, $E[Y]$ is given by:

$$E[Y]=\frac{W+1}{2}+4p\left(\frac{\left((1-p)^{m+1}+(2m+1)p^{m+1}-(m+1)p^m\right)}{2\left((1-p)^{m+1}-p^{m+1}\right)(1-2p)}\right) \tag{6.12}$$

Therefore, the probability τ that a station transmits a packet in a randomly chosen slot time is equal to:

$$\tau = \frac{1}{E[Y]} = \frac{2}{(W+1)+4p\left(\frac{\left((1-p)^{m+1}+(2m+1)p^{m+1}-(m+1)p^{m}\right)}{\left((1-p)^{m+1}-p^{m+1}\right)(1-2p)}\right)} \tag{6.13}$$

Note that the above expression for the probability τ is consistent with the one found in [125] but not with the one in [6] for the IEEE 802.11 BEB algorithm or the one in chapter 4 for the DIDD backoff mechanism. From equation (6.13), we observe that the transmission probability τ depends on the collision the collision probability p, which is still unknown, and it will be derived next.

If we assume that all stations see the system at steady state and transmit with probability τ, the collision probability p is given by:

$$p = 1-(1-\tau)^{n-1} \tag{6.14}$$

Equations (6.13) and (6.14) represent a non-linear system with two unknowns p and τ. This system can be solved by utilizing numerical methods and evaluating p and τ for a certain W and m combination. Note that $p \in (0,1)$ and $\tau \in (0,1)$. As it has been shown in [20] throughout a detailed proof, this non-linear system has a unique solution.

Figure 6.8 shows how the collision and RTS transmission probabilities, p and τ respectively, are affected from the network size and the various backoff parameters. This figure plots both probabilities as a function of the number of stations n and for different CW and m values. When no CW size adjustment is enforced after a successful reservation or collision (m=0), the figure illustrates that the collision probability is highly dependent on the number of stations. A large network size results in a higher collision probability and, thus, in an increased number of collisions. Conversely, when m=0 the probability of an RTS transmission is practically not affected by network size. If CW size is increased or decreased after a collision or a successful reservation respectively, results show that the collision probability increases as network size increases for n<20. Moreover, the RTS transmission probability is decreased for n values less than 25; for larger network size scenarios τ attains roughly the same values having a slight decreasing trend.

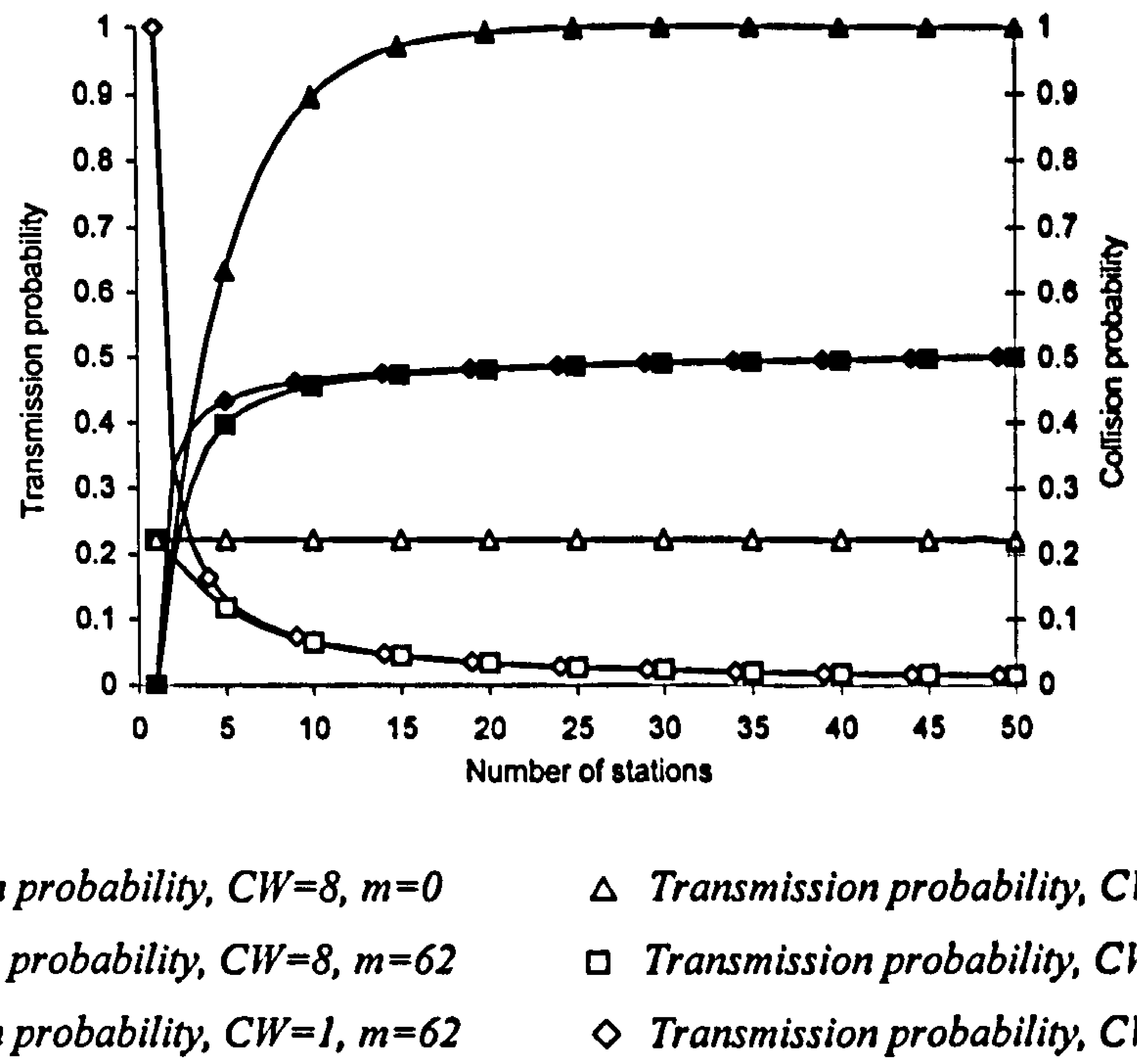


Figure 6.8 RTS packet collision and transmission probabilities, l=16 Kbits, ppb=1

C. Packet delay analysis

This section presents a neat method to calculate the average packet delay $E[D]$ for a successfully transmitted packet, defined to be the time interval from the time a packet is at the head of its MAC queue ready for transmission, until the instant of time the packet terminates a successful delivery.

Actually, by knowing the AIr saturation throughput S, we can calculate the average packet delay in a much more elegant way via the Little's Result as:

$$E[D] = \frac{E[N]}{S/l} \frac{1}{ppb} \tag{6.15}$$

where the numerator $E[N]$ represents the per-slot average number of competing stations which finally will successfully deliver a packet and the denominator S/l represents the departure rate of successful packets (i.e., the throughput measured in packets/seconds). It should be pointed out that (17) holds only for unbounded retries (no retry limits). In such case, clearly $E[N]=n$. Moreover, saturation throughput is defined as the time portion during which the infrared medium successfully transfers payload data and can be evaluated by dividing the time utilized for transmitting payload data by the average slot duration $E[slot]$[4]:

$$S = \frac{1}{RR} \frac{P_{tr} P_s \, ppb \, l}{E[slot]} \tag{6.16}$$

After some algebra, equation (6.15) becomes:

$$E[D] = \frac{E[slot]}{\tau(1-p)ppb} = \frac{E[slot]}{\tau(1-\tau)^{n-1} ppb} \tag{6.17}$$

From equation (6.17) we observe that the average packet delay depends on the average length of a slot time $E[slot]$ which is still unknown and will be calculated next. In order to compute $E[slot]$, we analyse all possible events that can occur in a randomly chosen slot time. Based on the calculated RTS collision probability p and transmission probability τ, we can now analyse all possible events that can occur in a randomly chosen slot. Let P_{tr} be the probability that at least one reservation attempt occurs (at least one station transmits an RTS packet) in the considered slot. For a LAN of n contenting stations, each transmitting with probability τ, P_{tr} is given by:

$$P_{tr} = 1-(1-\tau)^n \tag{6.18}$$

Let P_s be the conditional probability that an occurring RTS transmission is successful (P(success/trans)). The probability P_s can be evaluated as the probability that, in the considered slot time, only one station transmits an RTS attempting to reserve the infrared medium and the remaining (n-1) stations remain silent provided that at least one transmission occurs in the channel:

$$P_s = \frac{P_{success}}{P_{tr}} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n} \tag{6.19}$$

Therefore, a successful reservation attempt in a randomly selected slot time occurs with probability $P_{tr}P_s$ and the time utilized for transmitting payload information is $ppb \cdot t$, where ppb is the window size and t is defined as the time required for transmitting payload information data in an SDATA packet. The value of t is given by:

$$t = \frac{l}{C} \tag{6.20}$$

---

[4] The detailed derivation of saturation throughput can be found in [124][125] and [126].

where l is the packet payload data length and C is the data rate.

The average slot duration can now be evaluated by considering that a random slot time is empty with probability $1-P_{tr}$ and contains a collision with probability $P_{tr}(1-P_s)$. Thus, $E[slot]$ is equal to:

$$E[slot] = (1-P_{tr})\sigma + P_{tr}P_sT_s + P_{tr}(1-P_s)T_c \tag{6.21}$$

where σ is the duration of an empty slot time, T_s and T_C are the time durations the medium is sensed busy due to a successful reservation and a collision involving two or more simultaneous RTS packet transmissions respectively. A collision always lasts exactly one CAS duration and, therefore:

$$T_c = \sigma \tag{6.22}$$

Considering equation (6.22), the average slot duration $E[slot]$ can be easily reduced to:

$$E[slot] = P_{tr}P_sT_s + \sigma - P_{tr}P_s\sigma \tag{6.23}$$

For AIr networks employing the Reserved transfer mode with sequenced data (SDATA packet) (figure 6.4(d)), the duration of T_s is equal to:

$$T_S^{SDATA} = D_{over} + ppb(t + F_S + p_1) \tag{6.24}$$

where D_{over} is the reservation overhead that includes the transmission time of the RTS, CTS, EOB and EOBC packets and the TAT delays that follow these packets, F_S is the transmission time of the SDATA packet overhead (preamble, robust header, CRC etc.) and p_l is the preparation time of an SDATA packet (practically equal to zero). Assuming that the RTS MBR field is always transmitted using RR=1, D_{over} is given by:

$$D_{over} = T_{RTS} + TAT + T_{CTS} + TAT + T_{EOB} + TAT + T_{EOBC} + TAT \tag{6.25}$$

The value of F_S can be found as:

$$F_S = TT_{RH} + \frac{RR\, l_S'}{C} \tag{6.26}$$

where l_S' is the length of the MBR overhead of an SDATA packet (l_S' =80 bits) and TT_{RH} is the transmission time of a packet with no MBR field.

D. Time allocation tasks

Our analytical model also allows measurement of the time portion utilized on all component tasks affecting AIr delay performance. Such an evaluation reveals the impact of physical and link layer parameters to utilization. It is valuable for link designers in achieving high utilization at a reasonable cost and for link implementers in selecting suitable parameter values in order to maximize performance. Considering that a randomly selected slot is empty with probability $1-P_{tr}$ and that the CAS duration is σ, the time portion utilized in empty CAS because no station transmits is given by:

$$U_{empty} = \frac{(1-P_{tr})\sigma}{P_{tr}P_sT_s + \sigma - P_{tr}P_s\sigma} \tag{6.27}$$

A randomly selected slot contains a collision with probability $P_{tr}(1-P_s)$ and the time portion utilized on collisions when two or more stations are simultaneously trying to reserve the infrared channel is

$$U_{empty} = \frac{(1-P_{tr})\sigma}{P_{tr}P_sT_s + \sigma - P_{tr}P_s\sigma} \tag{6.28}$$

The time portion utilized on transmitting data packet overheads, the control packets (RTS/CTS/EOB/EOBC) and the associated TAT delays during a successful reservation period is given by:

$$U_{over} = \frac{P_{tr}P_s\left(T_s - \frac{l\,ppb}{C}\right)}{P_{tr}P_sT_s + \sigma - P_{tr}P_s\sigma} \tag{6.29}$$

As all component tasks that affect AIr delay performance are considered, the following equation always holds true:

$$U_{empty} + U_{coll} + U_{over} + U = 1 \tag{6.30}$$

where U is the channel utilization (throughput efficiency) found if we divide equation (6.16) by the data rate C. The correctness of equation (6.30) can be easily verified from equations (6.27)-(6.29).

6.6 Model validation

The previously presented analytical model is validated by comparing analytical with simulation results obtained using the AIr simulator introduced in [125]. This simulator was developed using the OPNET Modeler modeling and simulation software package and closely follows all timer values and packet element transmission times defined by AIr specifications. The OPNET AIr simulator emulates the real operation of a wireless station as closely as possible, by implementing the collision avoidance procedures and by closely following all parameters such as turnaround times, propagation delays and packet transmission times defined by AIr specifications. After the essential modification of the simulator in order to employ saturation conditions and to calculate AIr packet delay performance, we have run simulations on network sizes varying from 1 to 50 stations, in steps of 5. In all simulation runs, we assumed ideal channel conditions; an error free medium is assumed and no hidden stations are considered. Furthermore, the reserved transfer mode with sequenced data was employed; stations transmit *ppb* SDATA packets in every successful reservation attempt. Each SDATA packet is always carrying 16 Kbits of payload data (the maximum allowable size) at the 4 Mbit/s data rate.

Figure 6.9 provides packet delay performance results versus the number of stations and studies the accuracy of the developed mathematical analysis. The parameters used in both the analytical model and the simulation runs follow the parameters summarized in table 6.3. The performance comparison shows that analytical results (lines) coincide with simulation results (symbols) for different W, m and *ppb* values. Note that simulation results are acquired with a 95% confidence interval lower than 0.002. Our packet delay analysis gives results in high agreement with OPNET simulations and, therefore, it predicts very accurately AIr packet delay performance. Furthermore, an interesting observation is that packet delay significantly depends on the implemented backoff parameters such as W, m and *ppb*. The presented performance results are a strong indication of the great importance for the proper selection of the backoff and protocol parameters in order to reduce packet delay and, thus, maximize performance.

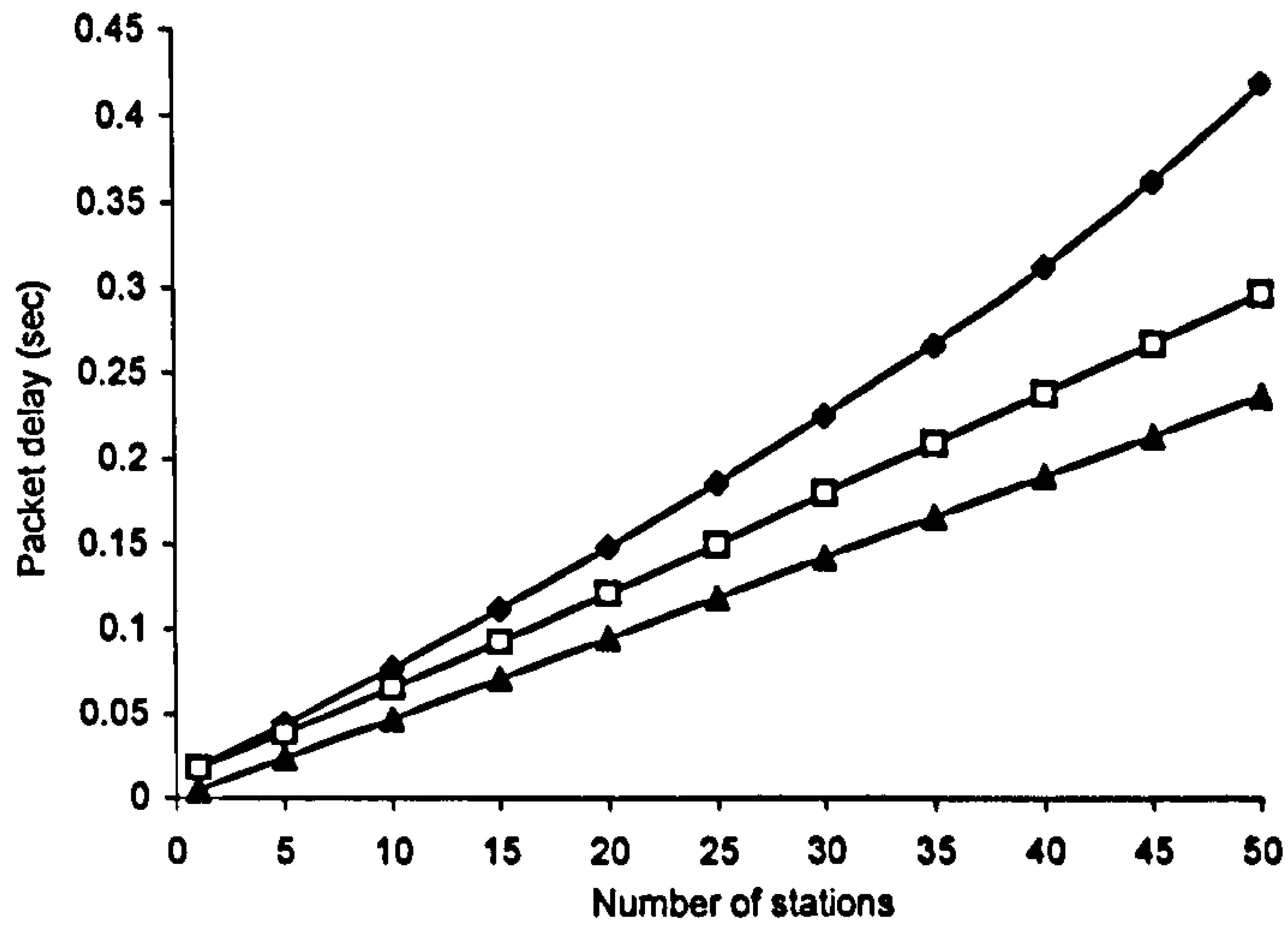


Figure 6.9 Packet delay: analysis (lines) versus simulation (symbols), l=16 Kbits, C=4 Mbit/s

6.7 Performance evaluation

This section presents a performance evaluation of the AIr protocol by employing the analytical model developed in section 6.5. This evaluation assumes that the infrared channel is error free and that no Repetition Rate coding (*RR*=1) is implemented. This section studies the impact of the backoff and system parameters on AIr MAC protocol performance by employing the previously derived mathematical analysis.

6.7.1 The effect of Contention Window (*CW*) size and packets burst size (*ppb*)

Figure 6.10 explores the dependency of performance on the *CW* size. Packet delay results are plotted versus number of stations when no *CW* size adjustment is imposed (*m*=0) after a successful reservation or collision. The figure shows that packet delay is not practically affected when a large *CW* size is implemented (*CW*=32 or 64) for any network scenario. Conversely, when a lower *CW* size is being used, packet delay is highly dependent on the network size *n*. When *n* increases, the increased number of collisions results in high packet delay values and, therefore, in significant performance

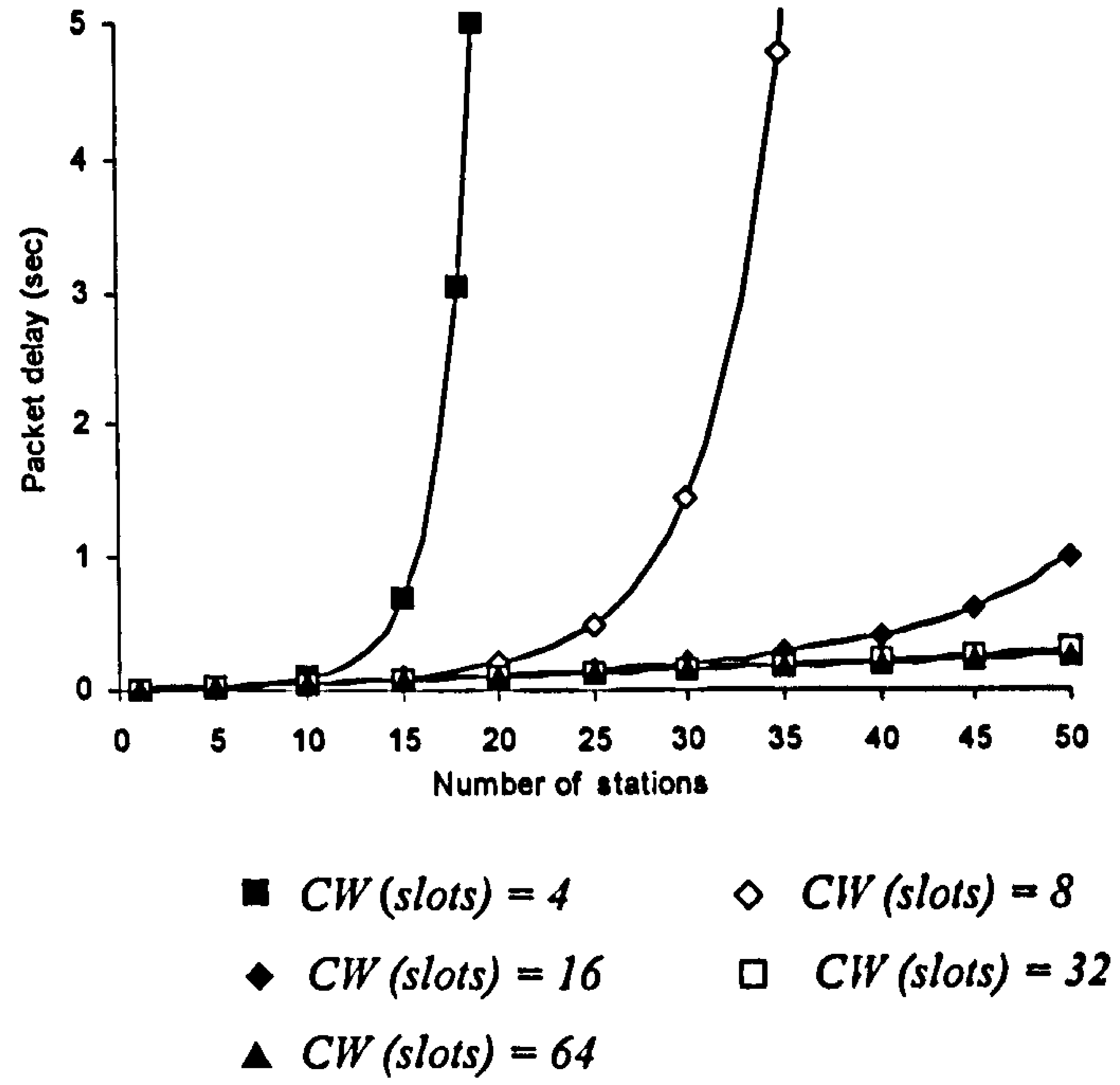


Figure 6.10 Packet delay versus n for fixed CW size, l=16 Kbits, ppb=4

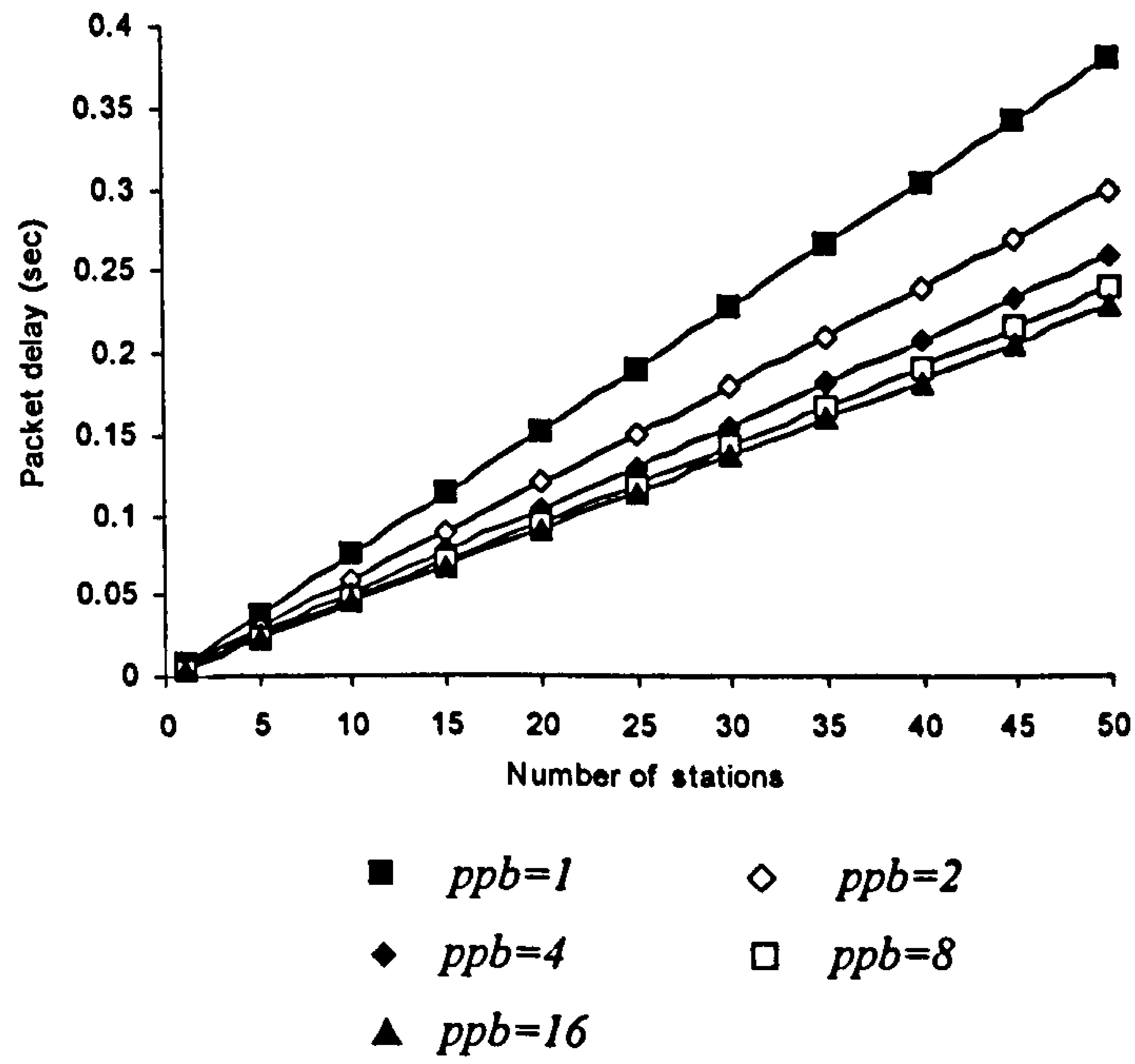


Figure 6.11 Packet delay versus n, for various ppb values, l=16 Kbits, CW=8, m=62

degradation, especially for small *CW* size values. Thus, for a given network size and when no *CW* size adjustment is implemented, an appropriate *CW* size value should be chosen in order to obtain minimum packet delay and as a result maximum performance.

The effect of the number of packets per burst (*ppb*) on performance is examined in figure 6.11 by plotting packet delay versus network size for different *ppb* values (*ppb*=1,2,4,8 and 16). Results show that performance is significantly improved by

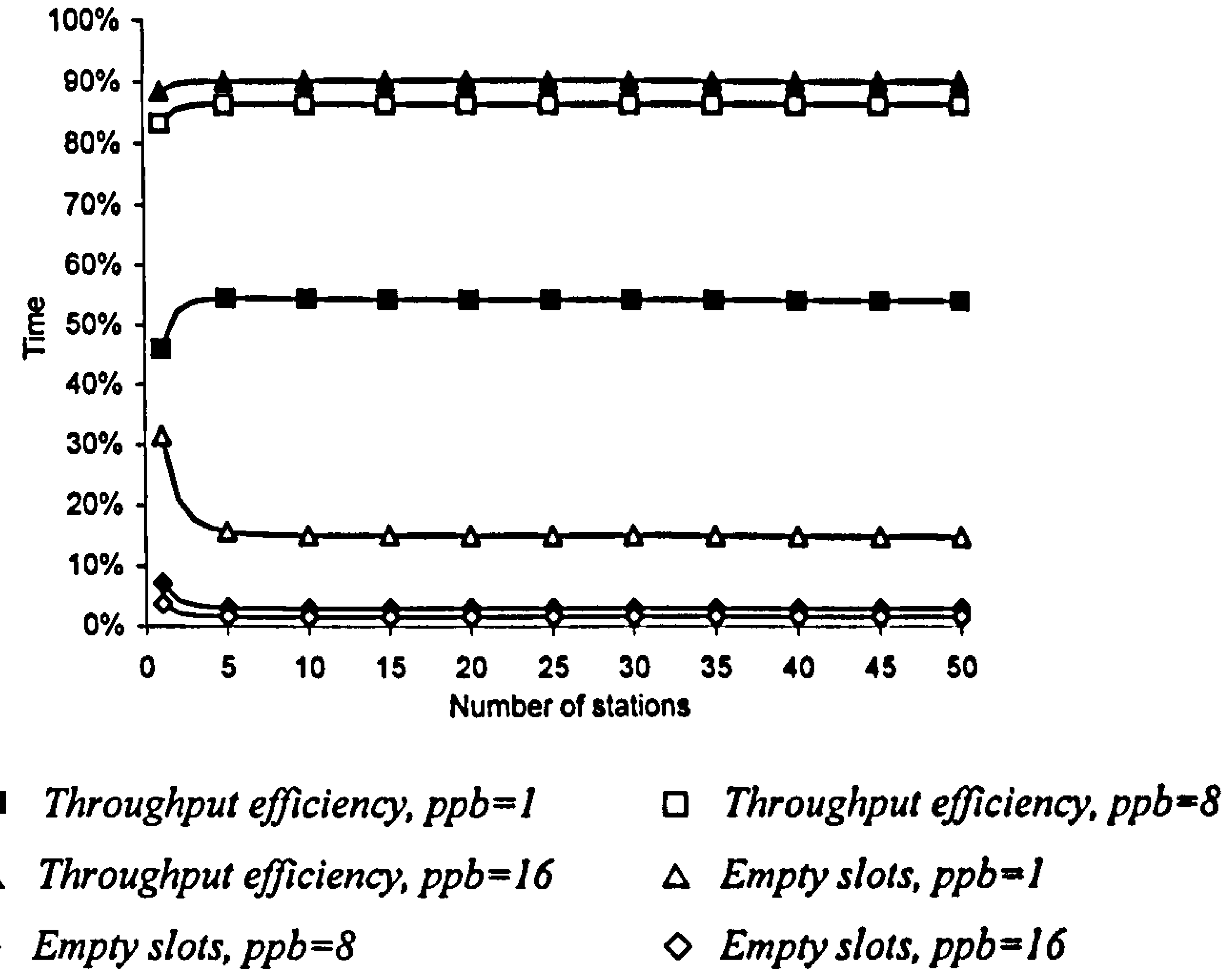


Figure 6.12 Time allocation of various AIr tasks versus n, l=16 Kbits, CW =8, m=62

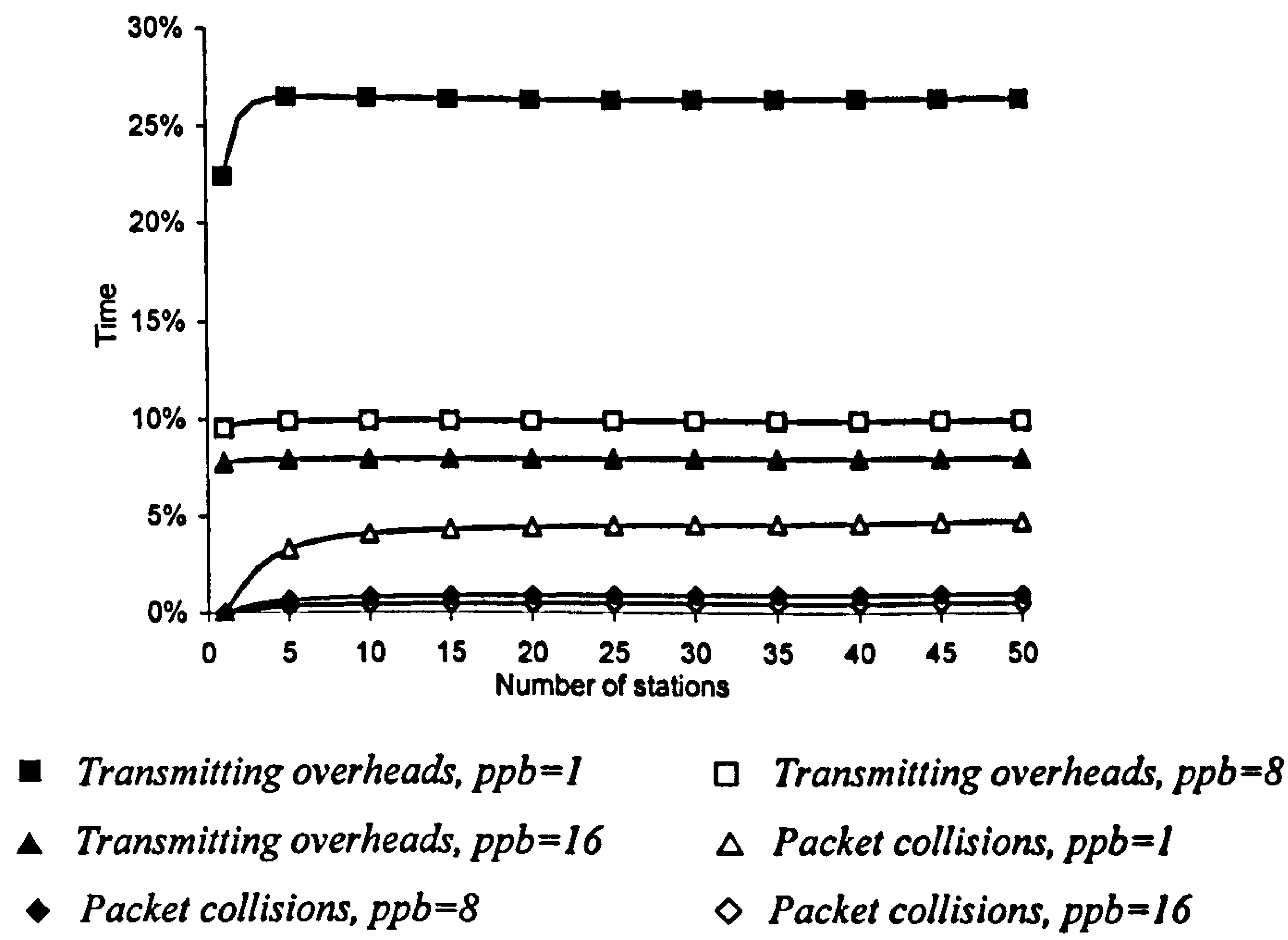


Figure 6.13 Time allocation of various AIr tasks versus n, l=16 Kbits, CW =8, m=62

putting multiple packets, and not only one, into each burst transmission. The situation is justified by noting that for each packet transmission a separate set of overhead information and delays (reservation time, inter-frame spaces, backoff time and acknowledgements) is needed. With packet bursting, instead of several sets of overhead for each packet, only one set of overhead information will be used. In this way, the packet delay can be reduced and the performance is significantly improved. Another useful observation is that the performance is not considerably enhanced when *ppb*=16 compared to the case of *ppb*=8.

Figures 6.12 and 6.13 clearly show all the factors affecting packet delay versus network size for two different *ppb* values (*ppb*=1, 16). Figure 6.12 plots throughput efficiency and the time portion utilized on empty slots. The figure depicts that when the number of packets per burst is either equal to 8 or 16 instead of *ppb*=1, throughput efficiency of the communication is increased. Note that when *ppb*=1, only 55% of the time is devoted in useful transmission in contrast with the case of *ppb*=8 or 16 when the equivalent amount of time is 82-88%. Furthermore, high *ppb* values significantly decrease the amount of time consumed on empty slots. Figure 6.13 plots the time portion utilized in packet collisions and in transmitting overheads during a successful transmission i.e. SDATA packet headers and control packets such as RTS, CTS, EOB and EOBC. The figure shows that when *ppb* increases, the amount of time consumed on inadequate tasks like packet collisions or transmitting overheads is significantly reduced. Thus, high *ppb* values appear to be a necessity in improving performance.

6.7.2 CW adjustment mechanism

The effectiveness of the proposed *CW* size adjustment mechanism and CW_{max} value is explored in figures 6.14 and 6.15. These figures plots packet delay versus *CW* size and maximum backoff stage m respectively, for 5 different network sizes (*n*=1, 5, 10, 20 and 50). Both figures show that the network size affects considerably performance; in large networks packet delay attains higher values than smaller networks due to the increased number of collisions no matter the *CW* size or the maximum backoff stage. Figure 6.14 depicts that the choice of *CW* size does not practically affect packet delay when a *CW* size adjustment mechanism exists (*m*=62), especially for high values of n. This conclusion is significantly different to the expressed conclusion in figure 6.10 that an appropriate *CW* size value is essential for maximum performance.

Figure 6.15 examines the appropriateness of the CW_{max} value selected in the AIr standard. In the case of large network sizes, the choice of m plays a key role in reducing packet delay; small m values result in a significant high packet delay and, therefore, impair performance. Moreover, performance results show that the dependence of the packet delay from the maximum backoff stage *m* for small networks is marginal. However, the figure illustrates that packet delay and, therefore, performance is not

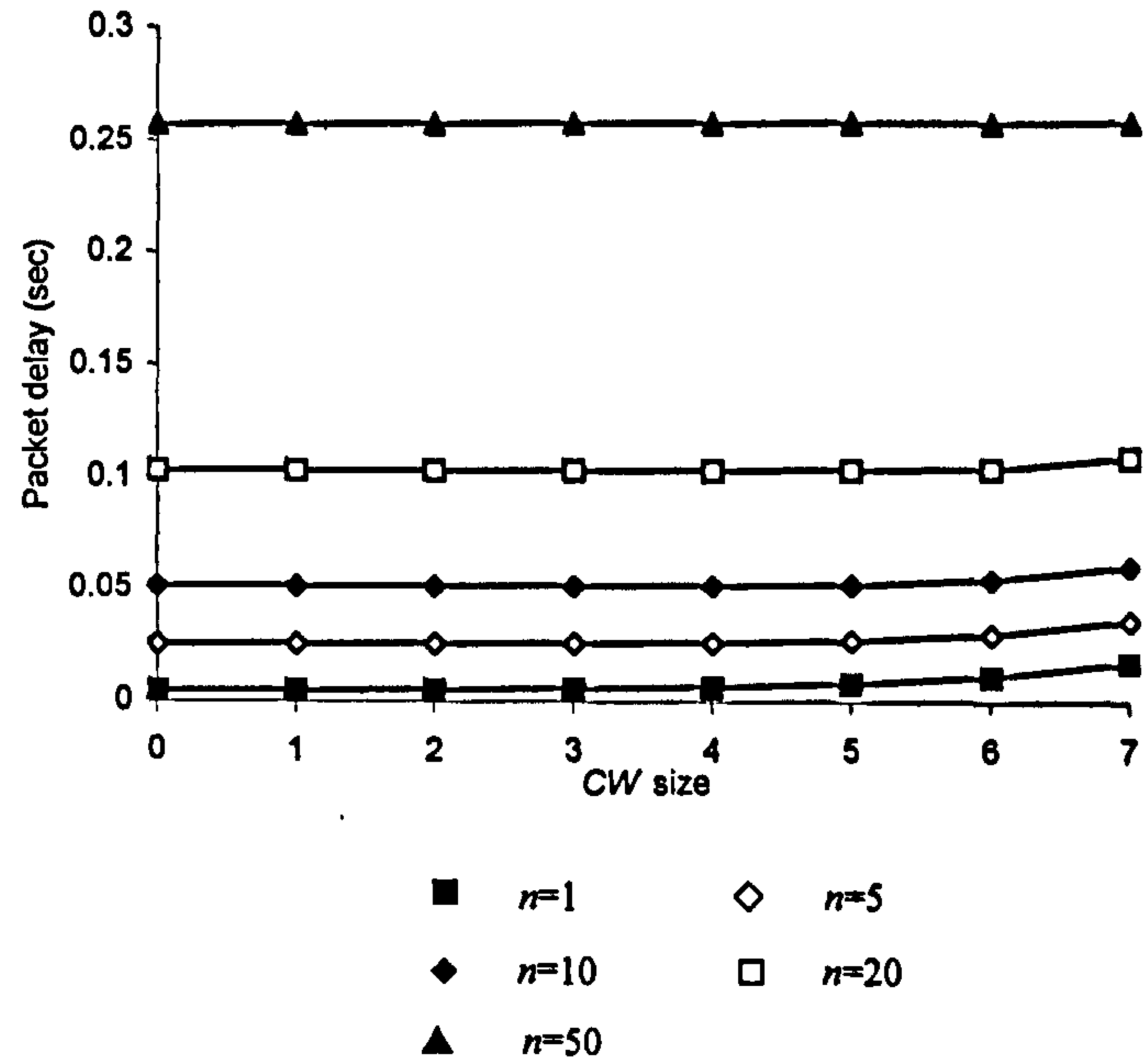


Figure 6.14 Packet delay versus CW size, for various n values, l=16 Kbits, CW =8, m=62, ppb=4

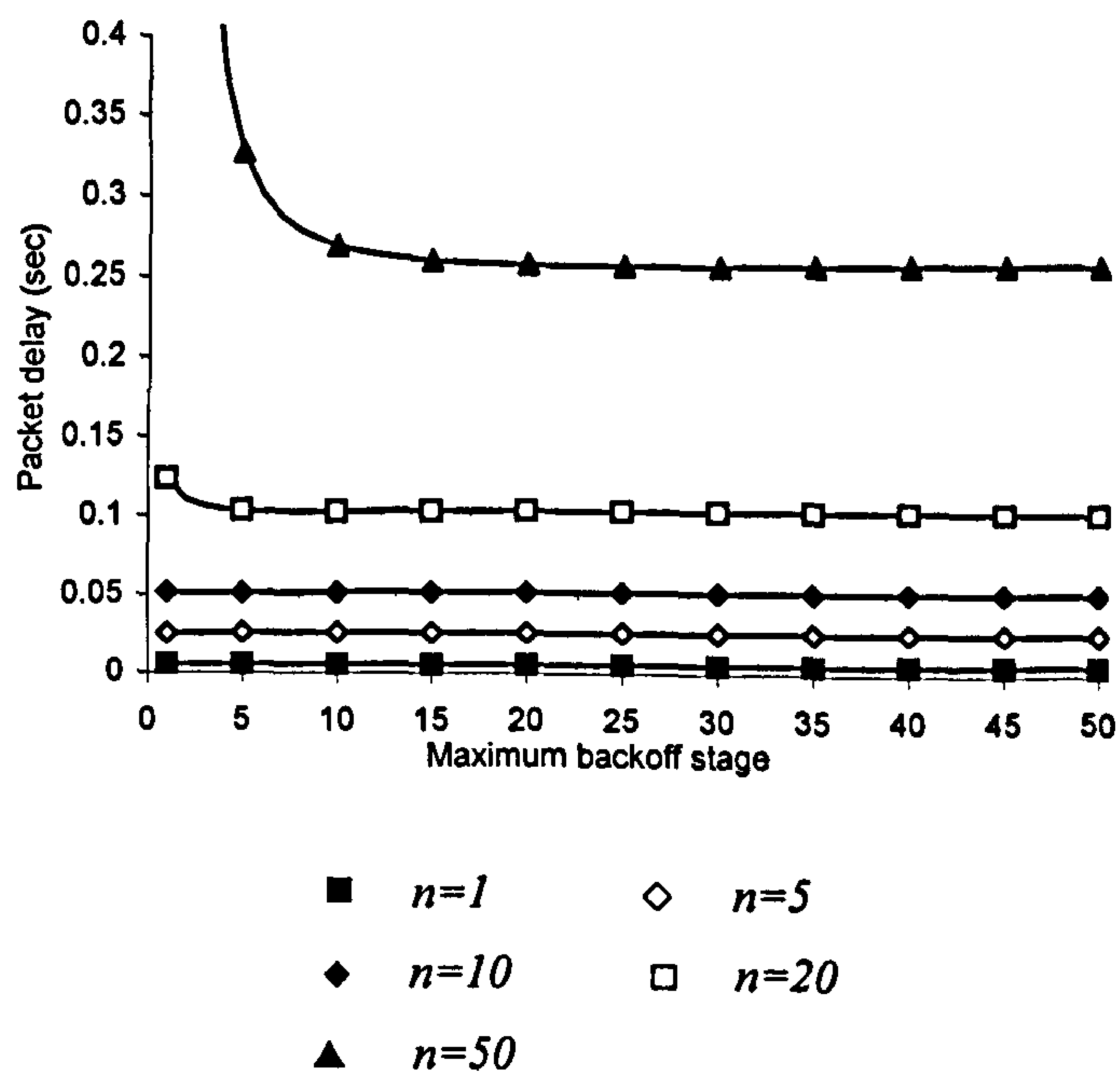


Figure 6.15 Packet delay versus m, for various n values, l=16 Kbits, CW =8, ppb=4

practically affected when the maximum backoff stage m is greater than 20. This means that a CW_{max} value of 64 instead of the proposed value of 256 is sufficient enough for maximum performance. This result can be explained as follows. Since AIr utilizes a linear adjustment of the *CW* size and the *CW* is increased by 4 (a relatively small value) after a packet collision. For this reason, large CW_{max} values are rarely used and only after a large number of consecutive packet collisions. As a result, CW_{max} can be safely lowered even for large network sizes.

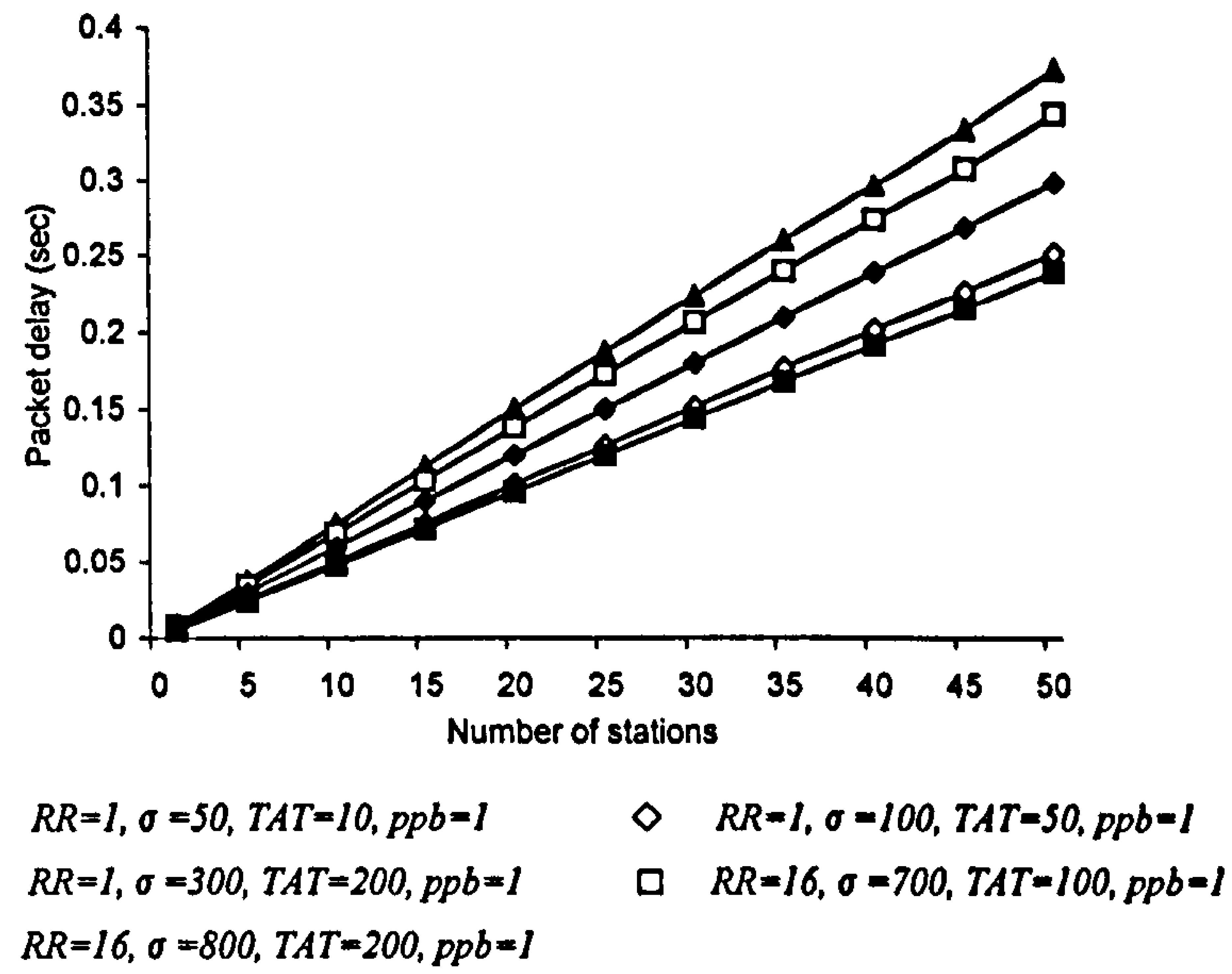


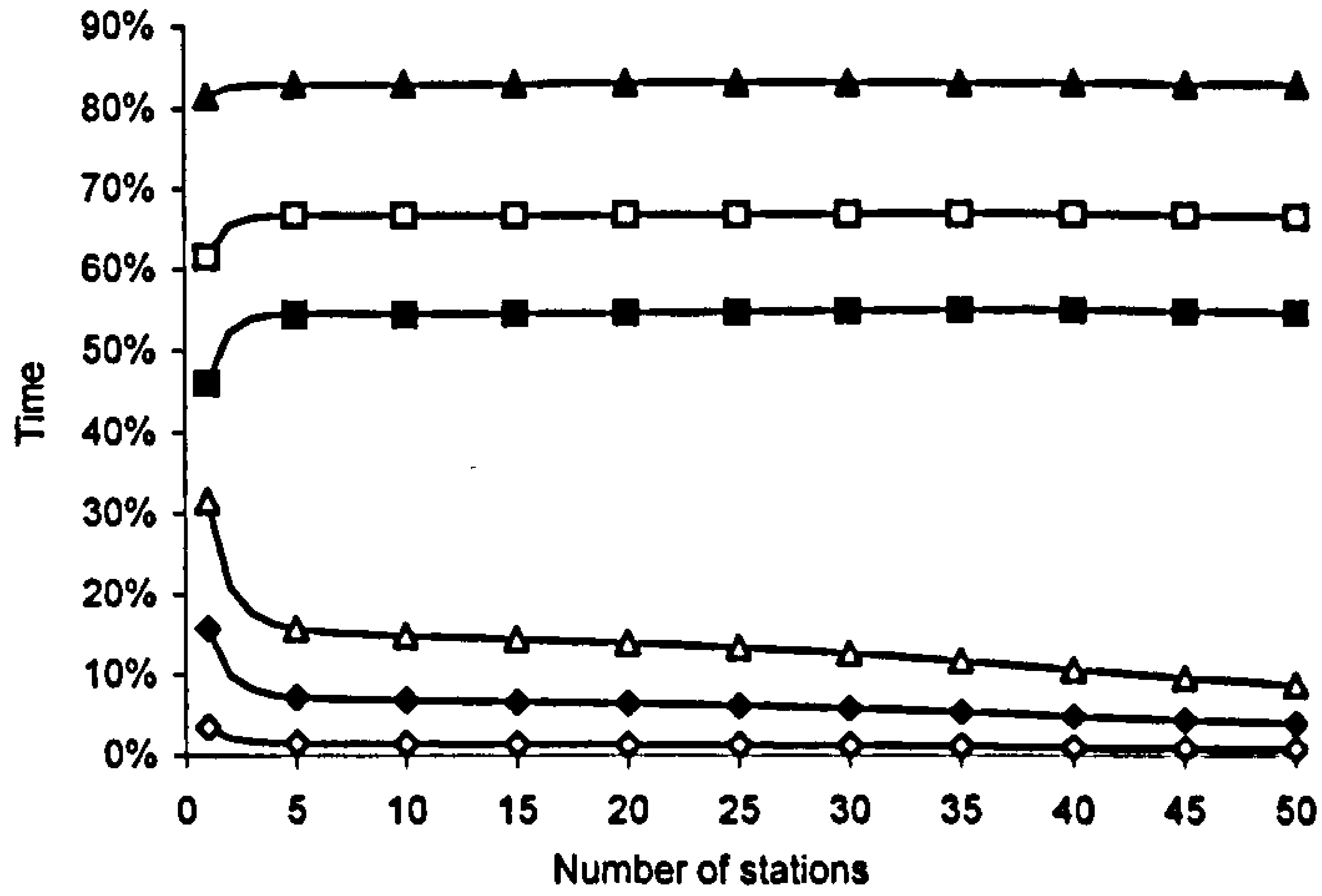
Figure 6.16 Packet delay versus n, for various physical layer parameters, l=16 Kbits, CW =8, m=20

Parameter	*RR=16*	*RR=16*	*RR=1*	*RR=1*	*RR=1*
TAT	200	100	200	50	10
$T_{RTS}+TAT+T_{PA}+T_{SYNC}$	548	448	428	278	238
CAS duration (σ)	800	700	300	100	50
F_s	252	252	132	132	132
D	1740	1340	1260	660	500

Table 6.4 AIr physical and link layer parameters for improved performance

6.7.3 The effect of high RH RR value, TAT delay and packet payload size

Figure 6.16 investigates the dependency of performance on various physical layer parameters (like RH RR and TAT delay values) by plotting packet delay against network size. The figure reports packet delay for the case of transmitting the RH field of all packets using *RR*=1 (instead of *RR*=16) for scenarios in which only one data packet of 16 Kbits (*ppb*=1) is transmitted as well as for lower minimum turnaround time (*TAT*) and CAS slot size (σ) values. Note that the implemented CAS slot time has to obey the restriction that $\sigma > T_{RTS}$+TAT+TT_{PA}+TT_{SYNC} in order to ensure that a station not hearing the RTS control packet will hear the beginning of the CTS control packet during the CAS time duration and defer transmission. The set of values utilized to derive figure 6.16 are displayed in table 6.4. As we can observe packet delay is considerably



■ *Throughput efficiency, RH in RR=16, σ =800, TAT=200*

□ *Throughput efficiency, RH in RR=1, σ =300, TAT=200*

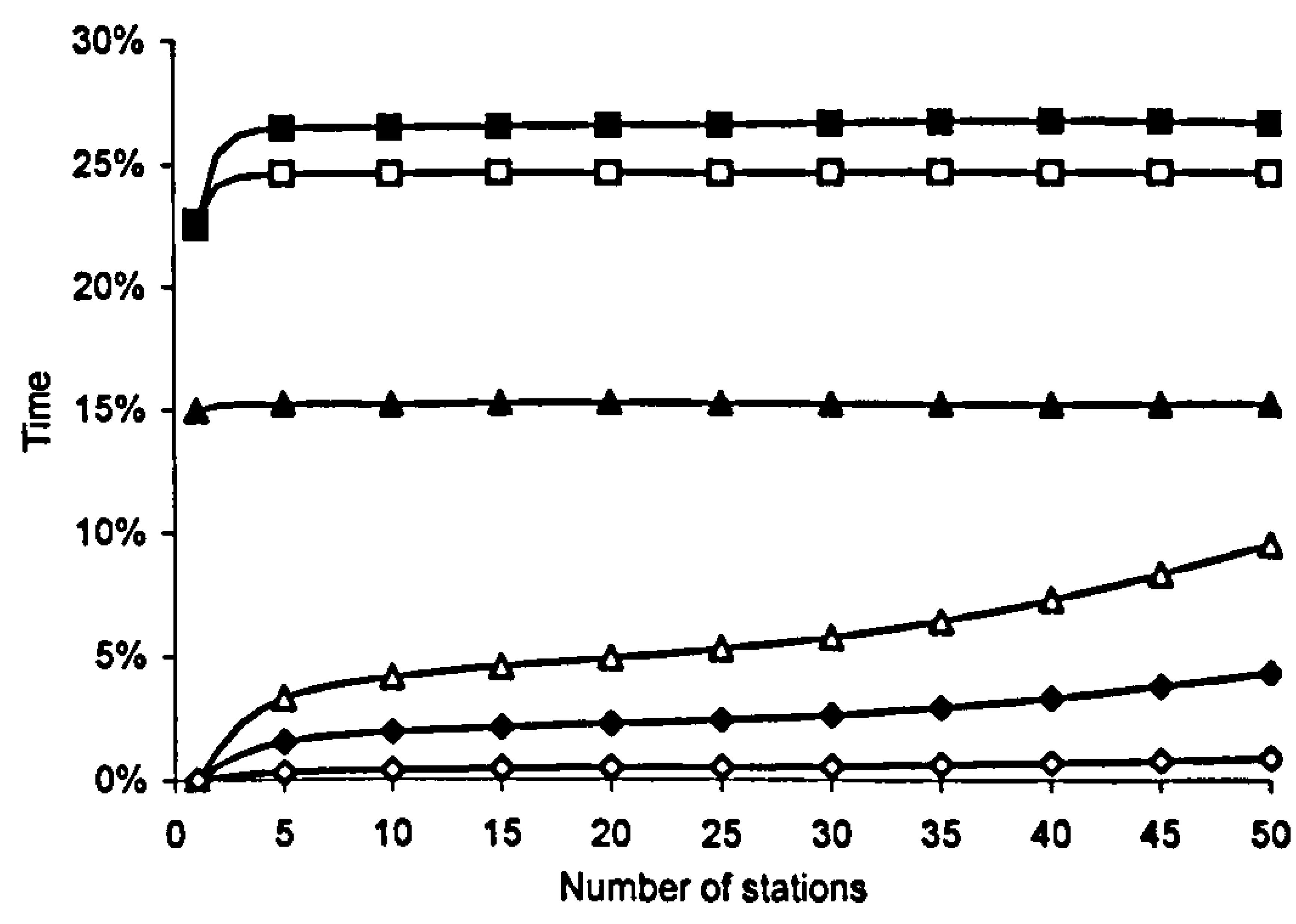
▲ *Throughput efficiency, RH in RR=1, σ =50, TAT=10*

△ *Empty slots, RH in RR=16, σ =800, TAT=200*

◆ *Empty slots, RH in RR=1, σ =300, TAT=200*

◇ *Empty slots, RH in RR=1, RH in RR=1, σ =50, TAT=10*

Figure 6.17 Time allocation of various AIr tasks versus n, l=16 Kbits, W=8, m=20, ppb=1



■ *Transmitting overheads, RH in RR=16, σ =800, TAT=200*

□ *Transmitting overheads, RH in RR=1, σ =300, TAT=200*

▲ *Transmitting overheads, RH in RR=1, σ =50, TAT=10*

△ *Packet collisions, RH in RR=16, σ =800, TAT=200*

◆ *Packet collisions, RH in RR=1, σ =300, TAT=200*

◇ *Packet collisions, RH in RR=1, σ =50, TAT=10*

Figure 6.18 Time allocation of various AIr tasks versus n, l=16 Kbits, W=8, m=20, ppb=1

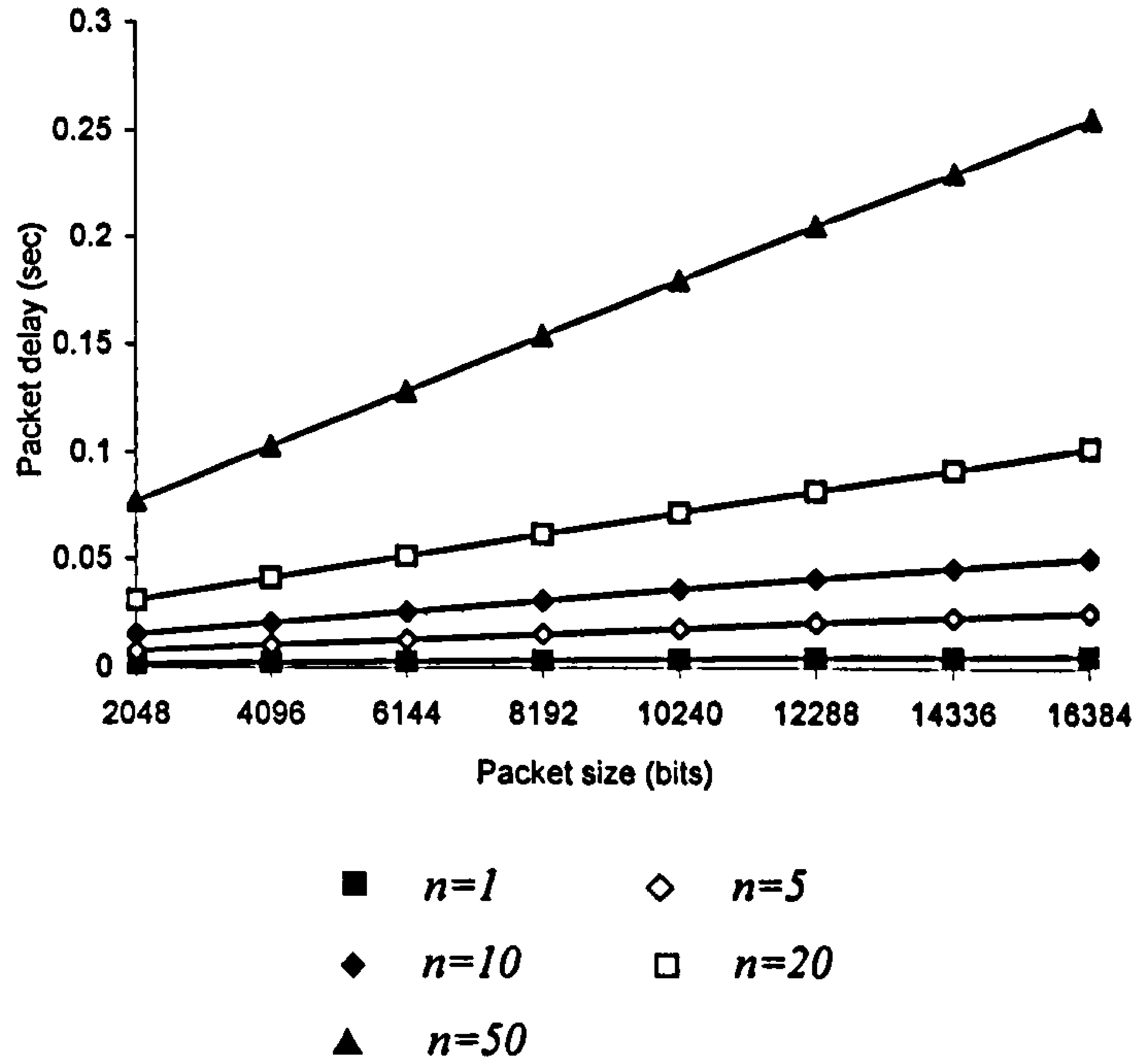


Figure 6.19 Packet delay versus l, for various n values, W=8, m=20, ppb=4

decreased if the RH field of all packets is transmitted using RR=1. Moreover, performance is also improved by appropriately reducing the minimum turnaround time and the CAS slot size. Therefore, the lowest packet delay values are achieved for RR=1 and for the much smaller than the proposed TAT and σ values (TAT=10, σ =50).

In figures 6.17 and 6.18 we clearly demonstrate why performance is considerably improved when the RH field of all packets is transmitted using RR=1 and when smaller values TAT and σ values are implemented. These figures plot throughput efficiency and the time portion utilized on empty slots (figure 6.17) and the time portion utilized in packet collisions and in transmitting overheads during a successful transmission (figure 15). Figure 14 shows that performance is significantly enhanced when RH is transmitted in RR=1 and for lower TAT and σ values since more time is utilized for useful transmission and less time is consumed in empty slots. Figure 6.18 also confirms the previous conclusion since much less time is utilized in transmitting overheads or in packet collisions.

Finally, in figure 6.19 packet delay is plotted against packet size in bits in order to study the effect of the packet payload size on performance. As it is expected, packet delay increases when packet size increases, especially in large network scenarios. However, it is understandable that throughput efficiency is improved when large size packets are transmitted since in this way the negative effect on performance of the packet overhead is minimized. For this reason, there is a trade-off between throughput efficiency and packet delay performance as the network size increases.

CHAPTER 7

Conclusions and Suggestions for Future Research

7.1 Conclusions

This work has focused on the efficient link layer design of WLAN connectivity utilizing the IEEE 802.11 protocol as well as multipoint infrared links based on IrDA AIr proposals. Several implementation issues have been discussed and two protocol stack proposals developed by the IEEE and Infrared Data Association (IrDA) have been considered, namely the widely used IEEE 802.11 protocol for mainly radio LANs and the AIr proposal for multipoint infrared links. The following protocol stack layers have been studied:

- IEEE 802.11 Medium Access Control (802.11 MAC)
- Advanced Infrared Medium Access Control layer (AIr MAC)

The main contributions of this thesis are:

a) The derivation of accurate mathematical equations which allow calculation of the performance of the IEEE 802.11 protocol by considering a variety of different performance metrics.

b) The proposal, analysis and modeling of certain easy-to-implement modifications of the legacy IEEE 802.11 protocol stack which significantly improve the operation and performance of IEEE 802.11 wireless links.

c) The detailed analysis and modeling of the AIr protocol performance, the study of link layer and protocol parameters that maximize performance for AIr.

d) The mathematical modeling and performance analysis of both the IEEE 802.11 and AIr protocols is derived by employing three simple and intuitive different modeling approaches.

7.1.1 Conclusions for the IEEE 802.11 protocol

a) The developed OPNET simulation results validate the accuracy of the derived packet delay analysis and show that the model, which takes into account packet retry limits, is more realistic and predicts very accurately DCF performance in terms of throughput and packet delay. In fact, the assumption of infinite retransmissions leads to an overestimated packet delay performance since it includes the long time delay of unlimited retransmissions of packets that should have been discarded as specified in the 802.11 standard. Moreover, simulation results are compared with results derived from the well known Bianchi's Markov chain model. It has been revealed that this model that considers consecutive packet transmissions without activating the backoff procedure overestimates packet delay.

b) According to the results of the analytical approach, the network size, the initial Contention Window (*CW*) size, the maximum *CW* size and the data rate all affect the performance of both access mechanisms considerably. In large network scenarios, performance drops off significantly due to the increased number of packet collisions. High values for the initial *CW* size improve the performance producing lower packet drop probability and higher throughput but at the same time increase packet drop time and in certain cases packet delay. Moreover, the increase on the maximum *CW* size enhances performance since the number of packet collisions is significantly decreased. Furthermore, increasing the data rate of the transmitted packets results in a considerable degradation of packet delay and packet drop time. Conversely, the increase on data rate does not affect at all the packet drop probability, increases throughput but results in decrease of throughput efficiency.

c) Performance is found to be significantly affected depending on the packet retry limits; a small retry limit value brings about a considerably lower packet delay and packet drop time but at the expense of lower throughput and also more packets are being dropped. Although, a high retry limit value enhances throughput efficiency and results in a lower packet dropping probability, at the same time it considerably increases packet delay as well as packet drop time. Thus, the value of 6 for the packet retry limit as defined in the IEEE 802.11 standard is proved to be a good trade-off.

d) Performance results demonstrate the deficiency of the RTS/CTS scheme for high data rates in both IEEE 802.11b and 802.11a protocols, unlike common expectation due to the exchange of the RTS and CTS reservation packets at a much lower control rate. It has been found that only very large packet size values render the RTS/CTS beneficial compared to the basic access scheme. Moreover, the RTS threshold that optimizes the use of the RTS/CTS scheme is highly dependent on the data rate; when the data rate increases, the threshold values increase significantly. Finally, a small retry limit coupled with large network size and small contention windows, decrease the RTS threshold values due to increased packet collision probability.

e) The proposed Double Increment Double Decrement (DIDD) backoff scheme (apart from the throughput improvement) achieves no packet drops. Under DIDD, every packet is being retransmitted until its successful transmission but with a decreased collision probability compared to the legacy DCF. The small price we pay for this performance improvement is that DIDD attains higher packet delay values comparing to the legacy DCF since it includes the time delay of packets that would have been discarded using the legacy DCF.

f) The idea of transmitting more than one data packets after winning DCF contention can be easily implemented using the fragmentation mechanism of the legacy DCF. The increased throughput and lower packet delay values due to the reduction of contention periods and RTS/CTS exchanges come at the cost of short-time unfairness. However, fairness improves in both cases of packet bursting and of the legacy DCF for long-term scale ensuring that all contending stations experience on average the same number of collisions.

g) Although, throughput increases with increasing packet length in an ideal channel (*BER*=0), in an error-prone environment there exist a trade-off between reducing the effect of the overhead by adopting a larger packet size and the need to reduce the effect of packet error rates by using smaller packet length. In fact, there is a packet size that maximizes throughput performance in a heavily error-prone channel. The optimal packet length does not depend on the number of contending stations but significantly depends on the *BER*. Furthermore, when burst transmission errors take place, performance significantly depends on the time spent in the *GOOD* and *BAD* states of the Gilbert-Elliot model.

7.1.2 Conclusions for the AIr standard

a) The proposed AIr MAC Collision Avoidance (CA) procedures perform effectively. The procedures include the Contention Window adjustment algorithm to minimize collisions (when two or more stations simultaneously transmit) and the large Collision Avoidance Slot (CAS) duration that avoids collisions from hidden stations. If the CAS duration is larger than necessary for small networks where collisions are less probable, the extra CAS slots result in an excessive time overhead which deteriorates performance. Thus, AIr in order to deal with this problem utilizes a dynamic CAS window backoff process with a linear adjustment of the Contention Window (*CW*).

b) A fixed CAS value (no *CW* size adjustment is imposed after a successful reservation or collision) could be used for a low-cost reservation system. In such a case, packet delay is significantly affected particularly for small network sizes and a suitable *CW* size value should be chosen in order to obtain minimum packet delay. It has been shown that a value of 32 slots provides a low delay performance for large networks (above 30 devices) and a value of 8 slots to achieve low delay performance for small network sizes (less than 10 stations).

c) When there are no hidden stations and when a *CW* size adjustment mechanism is in place, the proposed lower and upper limits for the *CW* size significantly affect packet delay and, therefore, impair performance. The proposed lower limit for the *CW* size results in performance degradation for a small network size whereas it does not practically affect packet delay for large network sizes. Therefore, the lower limit of 8 slots should be lowered to 1. The new lower limit results in significant performance improvement for one or a few transmitting stations. Moreover, for any network scenario, performance does not depend significantly on the maximum backoff stage for values greater than 20. Thus, the proposed upper limit of 256 for the *CW* size could be lowered to 88, even for large network sizes. This value actually corresponds to 20 backoff stages (88=8+20*4), when every stage utilizes the contention window size of the previous stage increased by 4.

d) A contending station that successfully reserves the infrared medium by using the RTS/CTS packet exchange transmits a burst of data packets. It has seen shown that performance is significantly improved with putting multiple packets into each burst transmission since the amount of time consumed on detrimental tasks like packet collisions, empty slots or transmitting overheads is significantly reduced. Thus, high burst size values appear to be a necessity in improving performance, however, the burst size may be limited by the requirements of upper layer protocols.

e) It has been also shown that an important parameter that affects packet delay is the payload size of the transmitted data packets. As it is expected, packet delay increases when packet size increases due to the increased transmission time delay. However, it is understandable that throughput efficiency is improved when large size packets are transmitted since this way the negative effect on performance of the packet overhead is minimized. Therefore, performance results point out that there is a trade-off between throughput efficiency and packet delay performance, especially for large network sizes.

f) AIr protocol utilizes a sufficiently large Collision Avoidance Slot (CAS) time to cover the time for an RTS transmission and the beginning of a CTS reply. This prevents overlapping of reservation attempts and also addresses the 'hidden station' problem. However, in certain situations this large CAS time as well as the high *RR*=16 used in transmitting the RH field can cause an excessive overhead that considerably affects performance. Finally, for indoor environments in which small amounts of data (one packet per burst) are transmitted in every successful reservation attempt, both the utilized RR for all packets as well as the minimum turnaround time and the CAS size should be lowered in order to reduce packet delay and, thus, enhance performance.

7.2 Suggestions for future research

a) The explosive growth of multimedia applications in the recent years raised the need for Quality of Service (QoS) support such as guaranteed delay, jitter and bandwidth for these applications. A new standard, the IEEE 802.11e, is developed to offer certain QoS support and multimedia support to the existing 802.11/.11b/.11a WLAN standards. The development of an analytical model that accurately evaluates the IEEE 802.11e protocol performance is an open challenge. An analysis as such could examine the suitability of IEEE 802.11e in providing QoS for real-time multimedia applications for a number of different network scenarios. The proposed study could also examine the adaptation of the protocol parameters to the traffic load and the optimization of the tradeoff between efficiency, priority and fairness among stations.

b) The packet delay analysis of the IEEE 802.11 and AIr protocols could be extended in order to calculate the important properties of the constituent curves of the delay distribution curve. Performance results will probably indicate that the time delays of packets are not close to their average value; most packets have very low time delays and a small number of packets experience very high delays. A mathematical model that calculates the average packet delay per stage and the probability that a stage is utilized for a successful packet transmission is an open issue. Such an analysis will reveal an additional insight view of the internal mechanisms of the DCF affecting packet delay performance.

c) Since the focus in this work was on the IEEE 802.11 and AIr MAC layers, we have made the implicit assumption that hidden stations are not present in the considered environment. However, as the presence of hidden stations is probable in wireless communications, and an extensive analysis of the performance degradation due to hidden stations is definitely of high priority for future research in the area of both IEEE 802.11 and AIr protocols.

d) Previous research on the area of wireless communications has shown that the implementation of optimum window and packet size values significantly improves the performance of wireless point-to-point infrared links, especially at high BER. The IrDA AIr protocol introduces significant turn around delays arising from the collision avoidance procedures and the RTS/CTS/EOB/EOBC packet exchange. The derivation of optimum burst size and packet length values that maximize AIr protocol performance can be considered as interesting area of future research. A mathematical analysis could recommend suitable window and packet size values that the transmitter should implement in order to cope with transmission errors in wireless communications.

REFERENCES

[1] Aad I. and Castelluccia C., "Differentiation Mechanisms for IEEE 802.11", in Proceedings of the Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2001), Alaska, USA, pp. 209 – 218, 2001.

[2] Bantz D. F. and Bauchot F. J., "Wireless LAN design alternatives", IEEE Network, vol. 8, no. 2, pp. 43-53, Apr. 1994.

[3] Barry J. R., "Wireless Infrared Communications", Kluwer Academic Publishers, 1994.

[4] Bharghavan V., Demers A., Shenker S. and Zhang L., "MACAW: A media access protocol for wireless LANs", in Proceedings of the Conference on Communications architectures, protocols and applications, SIGCOMM'94, pp. 212-225, Oct. 1994.

[5] Bianchi G., Fratta L. and Oliveri M., "Performance Evaluation and Enhancement of the CSMA/CA MAC Protocol for 802.11 Wireless LANs", in Proceedings of the 7th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 1996), Taipei, vol. 2, pp. 392-396, 1996.

[6] Bianchi G., "Performance Analysis of the IEEE 802.11 Distributed Coordination Function", IEEE Journal on Selected Areas in Communications, vol.18, no.3, pp. 535-547, 2000.

[7] Bianchi G. and Tinnirello I., "Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network", in Proceedings of the 22nd Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003), vol. 2 , pp.844 – 852, 2003.

[8] Bisdikian C., Bhagwat P. and Golmie N., "Wireless Personal Area Networks", IEEE Network Magazine, vol. 15, no. 5, pp. 10-11, Sept/Oct. 2001.

[9] Boucouvalas A. C. "IEC 825-1 Eye Safety Classification of some consumer electronic products", in Proceedings of IEE Colloquium on Free-Space Communication Links, IEE Savoy Place, London, no. 13, pp.1-6, Feb. 1996.

[10] Boucouvalas A. C. and Vitsas V., "100 Mb/s IrDA Protocol Performance Evaluation", in Proceedings of IASTED International Conference on Wireless and Optical Communications (WOC 2001), Banff, Canada, pp.49-57, June 2001.

[11] Boucouvalas A. C. and Vitsas V., "Optimum Window and Frame Size for IrDA Links", IEE Electronic Letters, vol. 37, no. 3, pp194-196, Feb. 2001.

[12] Bruno R., Conti M. and Gregori E., "IEEE 802.11 optimal performance: RTS/CTS mechanism vs. basic access", in Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2002), vol.4, pp. 1747 -1751, Sep 2002.

[13] Calì F., Conti M. and Gregori E., "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit", IEEE/ACM Transactions on Networking, no. 8, pp. 785–799, 2000.

[14] Carvalho M. M. and Garcia-Luna-Aceves J. J., "Delay Analysis of IEEE 802.11 in Single-Hop Networks", in Proceedings of 11th IEEE International Conference on Network Protocols (ICNP 2003), Atlanta, USA, pp. 146 – 155, 2003.

[15] Chatschik, B., "An Overview of Bluetooth Wireless Technology", IEEE Communications Magazine, vol. 39, issue 12, pp. 86-94, Dec. 2001.

[16] Chow F. and Kahn J. M., "Effect of non-reciprocity on infrared wireless Local Area Networks", in Proceedings of the IEEE Global Telecommunications Conference (Globecom 1999), Rio de Janeiro, Brazil, vol. 1a, pp. 330-338, Dec. 5-9, 1999.

[17] Chatzimisios P. and Boucouvalas A. C., "Packet Delay Analysis of the Advanced Infrared (AIr) CSMA/CA MAC Protocol in Optical Wireless LANs", Special Issue in Indoor Optical Wireless Communication Systems and Networks, Wiley International Journal of Communication Systems, accepted (to appear), 2004.

[18] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "Effectiveness of the RTS/CTS Handshake in IEEE 802.11a Wireless LANs", IEE Electronic Letters, vol. 40, no. 14, pp. 915-916, 8th July 2004.

[19] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "IEEE 802.11 Packet Delay - A Finite Retry Limit Analysis", in Proceedings of IEEE Global

Telecommunications Conference (Globecom), San Francisco, USA, vol. 2, pp. 950-954, Dec. 2003.

[20] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "IEEE 802.11 Wireless LANs: Performance Analysis and Protocol Refinement", EURASIP Journal on Wireless Communications and Networking, accepted (to appear), 2004.

[21] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "IEEE 802.11 WLANs: Performance Analysis in Presence of Bit Errors", in Proceedings of the International Conference on Communication Systems, Networks and Digital Processing (CSNDSP 2004), Newcastle, UK, pp. 355-358, July 2004.

[22] Chatzimisios P. and Boucouvalas A. C., "Improving Performance Through Optimization of the RTS/CTS Mechanism in IEEE 802.11 Wireless LANs", in Proceedings of the International Conference on Communication Systems, Networks and Digital Processing (CSNDSP 2004), Newcastle, UK, pp. 363-368, July 2004.

[23] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "Influence of BER on the Performance of IEEE 802.11 DCF", IEE Electronic Letters, vol. 39, no. 23, pp. 1687-1689, 13th Nov. 2003.

[24] Chatzimisios P. and Boucouvalas A. C., "IrLAP IrDA Protocol Throughput Dependence on Processor Speed", in Proceedings of the International Symposium on Communication Systems, Networks and Digital Processing (CSNDSP 2002), Stafford, UK, pp. 272-275, July 2002.

[25] Chatzimisios P. and Boucouvalas A. C., " Optimisation of IrLAP IrDA Protocol Including Processor Speed ", in Proceedings of the 3rd Annual Convergence of Telecommunications, Networking and Broadcasting Conference (PGNET 2002), pp. 298-303, Liverpool, UK, June 2002.

[26] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "Optimisation of RTS/CTS handshake in IEEE 802.11 Wireless LANs for maximum performance", in Proceedings of the IEEE Global Telecommunications Conference (Globecom 2004), Dallas Texas, USA, pp. 270-275, 29 Nov.- 3 Dec. 2004.

[27] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "Packet Delay Analysis of the IEEE 802.11 MAC Protocol", IEE Electronic Letters, vol. 39, no. 18, pp. 1358-1359, 4th Sept. 2003.

[28] Chatzimisios P., Boucouvalas A. C. and Vitsas V., " Performance Analysis of IEEE 802.11 DCF in the Presence of Transmission Errors ", in Proceedings of the IEEE International Conference on Communications (ICC 2004), Paris, France, vol. 7, pp. 3854 – 3858, June 2004.

[29] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "Performance analysis of the IEEE 802.11 MAC protocol for Wireless LANs", Wiley International Journal of Communication Systems, accepted (to appear), 2004.

[30] Chatzimisios P., Boucouvalas A. C. and Vitsas V., "Throughput and Delay analysis of IEEE 802.11 protocol", in Proceedings of IEEE International Workshop on Networked Appliances (IWNA), Liverpool, UK, pp. 168-174, October 2002.

[31] Chhaya H. S. and Gupta S., "Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol", Wireless Networks, pp.217-234, 1997.

[32] Comer D. E., "Computer Networks and Internets with internet applications", Prentice Hall, 2001.

[33] Crow B. P., Widjaja I., Kim J. G. and Sakai P. T, "IEEE 802.11 wireless local area networks", IEEE Communications Magazine, vol. 35, no. 9, pp. 116-126, 1997.

[34] Crow B. P., Widjaja I., Kim J. G. and Sakai P. T, "Investigation of the IEEE 802.11 medium access control (MAC) sublayer functions", in Proceedings of the Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 1997), vol. 1, pp. 126-133, 1997.

[35] Dietrich P. and Rao R. R., "Second order bounds on block error probabilities in stationary, time varying channels", in Proceedings of Conference on Information and Systems, Maryland , USA, Mar.1996.

[36] Duchamp D. and Reynolds N. F., "Measured Performance of a Wireless LAN", in Proceedings of IEEE Conference on Local Computer Networks (LCN 1992), Minneapolis, USA, pp. 494-499, Sep. 1992.

[37] Dun A. and Langston T., "Technical Summary of the SWAP Specification", HomeRF, March 1998.

[38] Dutta-Ray A., "Netwoks for homes", IEEE Spectrum, vol. 36, no. 12, pp. 26-33, Dec. 1999.

[39] Elliot E. O., "Estimates on error rates for codes on burst-noise channels", Bell Syst. Tech. J. 42, pp. 1977–1997, 1963.

[40] Fang Z., Bensaou B. and Wang Y., "Performance Evaluation of a Fair Backoff Algorithm for IEEE 802.11 DFWMAC", in Proceedings of ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) 2002, Lausanne, Switzerland, pp. 48 - 57 , June 2002.

[41] Garcia-Frias J. and Crespo P. M., "Hidden Markov models for burst error characterization in indoor radio channels", IEEE Transactions on Vehicular Technology, vol. 46, pp. 1006–1020, Nov. 1997.

[42] Fritchman B. D., "A binary channel characterization using partitioned Markov chains", IEEE Transactions on Information Theory, vol. 13, pp. 221–227, Apr. 1967.

[43] Gfeller F. and Hirt W., "A robust wireless infrared system with channel reciprocity", IEEE Communications Magazine, vol. 36, no. 12, pp. 100-106, 1998.

[44] Gilbert E. N., "Capacity of a Burst-Noise Channel", Bell Systems Technical Journal, Vol. 39, pp. 1253-1266, Sept. 1960.

[45] Chandra A., Gummalla V. and Limb J. O., "Wireless medium access control protocols", IEEE Communications Surveys, vol. 3, no. 2 Second Quarter 2000.

[46] Haas Z. J., Deng J., Liang B., Papadimitratos P., and Sajama S., "Wireless Ad Hoc Networks", Encyclopedia of Telecommunications, John Proakis (editor), John Wiley, 2002.

[47] Jianhua H., Zuoyin T., Zongkai Y., Wenqing C. and Chun C. "Performance evaluation of distributed access scheme in error-prone channel", in Proceedings of IEEE Conference on Computers, Communications, Control and Power Engineering (TENCON), vol. 2, pp. 1142-1145, 2002.

[48] Heatley D. J. T. and Neild I., "Optical wireless-the promise and the reality", IEE Colloquium on Optical Wireless Communications, Savoy Place, London, pp. 1/1-1/6, 1999.

[49] Higginbottom G. N., "Performance Evaluation of Communication Networks", Artech House, 1998.

[50] HiperLAN- High Performance Radio Local Area Network Draft ver. 1.1, ETSI, 1995.

[51] Ho T. and Chen K., "Performance Analysis of IEEE 802.11 CSMA/CA Medium Access Control Protocol", in Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 1996), Taipei, Taiwan, pp. 407-411, 1996.

[52] HomeRF, HomeRF Working Group, HomeRF, http://www.homerf.org

[53] Huang P., Chatzimisios P. and Boucouvalas A. C., "A Study of Link Propagation Delay and Processing Speed for IrDA Links at Gbit/s Data Rate", in Proceedings of the IrDA International Optical Wireless Communications Symposium, Warwick, UK, pp. 21-30, Sep. 2003.

[54] Huang P. and Boucouvalas A. C., "OBEX Performance Evaluation and parameter Optimization for High Speed IrDA Links", in Proceedings of IEEE International Conference on Communications (ICC 2004), Paris, France, vol.7, pp. 3849 – 3853, June 2004.

[55] IEEE TG3, IEEE 802.15.3 (TG3) Criteria Definitions, IEEE P802.15-00/110r13, Sept. 2000.

[56] IEEE 802.11e draft/D4.0, Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), November 2002.

[57] IEEE Std 802.11h-2003 (Amendment to IEEE Std 802.11, 1999 Edn. (Reaff 2003)) , Pages 1 – 59, 2003.

[58] IEEE Std 802.11f-2003, pp. 1 – 67, 2003.

[59] IEEE Std 802.11j-2004, pp. 1 – 40, 2004.

[60] IEEE Std 802.11g-2003 (Amendment to IEEE Std 802.11, 1999 Edn. (Reaff 2003) as amended by IEEE Stds 802.11a-1999, 802.11b-1999, 802.11b-1999/Cor 1-2001, and 802.11d-2001), pp. 1 – 67, 2003.

[61] IrDA: "Technical Overview of an Advanced IrDA Link Access Protocol and Physical Layer – Request for Comments – version 0.1", (IrDA, 1997).

[62] IrDA: "Advanced Infrared Draft Physical Layer Specification (AIr-PHY) – version 0.4", (IrDA, 1998).

[63] IrDA: "Advanced Infrared Physical Layer Specification (AIrPHY) – version 1.0", (IrDA, 1998).

[64] IrDA: "Serial Infrared Link Access Protocol (IrLAP) – version 1.1", (IrDA, 1996).

[65] IrDA: "Advanced Infrared Medium Access Control (AIrMAC) Draft Protocol Specification - version 1.0", (IrDA, 1999).

[66] IrDA: "Advanced Infrared Link Manager (AIrLM) Draft Specification – version 0.3", (IrDA, 1999).

[67] IrDA: "Advanced Infrared Logical Link Control (AIrLC) Specification, version 0.1", (IrDA, 1999).

[68] Jain R., Chiu W. and Hawe W. R., "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," DEC Tech. Rep., 1984.

[69] Johnsson M., "HIPERLAN/2 - The broadband radio transmission technology operating in the 5 GHz frequency band", ver 1.0, HiperLAN/2 Global Forum, 1999.

[70] Jungnickel V., Forck A., Haustein T., Krüger U., Pohl V. and von Helmolt C., "Wireless infrared communication beyond 100 Mbit/s: System design, transmission experiments and potential", in Proceedings of World Multiconference on Systemics, Cybernetics and Informatics (SCI 2001), Orlando, Florida, vol. IV, pp. 413-420, July 2001.

[71] Kahn J. M. and Barry J. R., "Wireless infrared communications", in Proceedings of the IEEE, vol. 85, pp. 265-298, Feb. 1997.

[72] Kahn J. M., Barry J. R., Audeh M. D., Carruthers J. B., Krause W. J. and Marsh G. W., "Non-directed infrared links for high-capacity wireless LANs", IEEE Personal Communications, vol. 1, no. 2, pp. 12-25, Second Quarter 1994.

[73] Karande S. K., Khayam, S. A., Krappel M. and Hayder R., "Analysis and Modeling of Errors at the 802.11b Link-Layer", in Proceedings of IEEE International Conference on Multimedia and Expo (ICME), vol. 1, 673-676, July 2003.

[74] Karaoguz J., "High-rate wireless personal area networks", IEEE Communications Magazine, vol. 39, no. 12, pp. 96-102, Dec. 2001.

[75] Karn P., "MACA – A new channel access method for packet radio", in Proceedings of the 9th Amateur Radio Computer Networking Conference (ARRL/CRRL), pp. 134-140, Sep. 1990.

[76] Khayam, S. A., Karande S. K., Hayder R. and Dmitri L., "Performance Analysis and Modeling of Errors and Losses over 802.11b LANs for High-Bitrate Real-Time Multimedia", Signal Processing: Image Communication - Special Issue on Recent Advances in Wireless Video, vol.18, no.7, pp. 575-595, 2003.

[77] Khirat S. and Kadhi A., "Wireless Personal Area Networks and HomeRF", DESS Reseaux de Radio communications avec des mobiles, Feb. 2002.

[78] Khurana S., Kahol A., Gupta S. K. S and Srimani P. K., "Performance evaluation of distributed co-ordination function for IEEE 802.11 wireless LAN protocol in presence of mobile and hidden terminals", in Proceedings of International Symposium on Analysis and Simulation of Computer and Telecommunication Systems, pp. 40 – 47, Oct. 1999.

[79] Kim Y. Y. and Li S.-qi, "Modeling multipath fading channel dynamics for packet data performance analysis", in Proceedings of the Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 1998), pp. 1292–1300, 1998.

[80] Kim Y. Y. and Li S.-qi, "Capturing important statistics of a fading/shadowing channel for network performance analysis", IEEE Journal Selected Areas in Communications, vol. 17, pp. 888–901, May 1999.

[81] Kleinrock L. and Tobagi F., "Packet switching in radio channels, part II – the hidden terminal problem in carrier sense multiple access and the busy tone solution", IEEE Transactions in Communications, vol. 23, no.12, pp. 1417-1433, Dec. 1975.

[82] Koksal C. E., Kassab H. and Balakrishnan H., "An analysis of short-term fairness in wireless media access protocols", in Proceedings of ACM SIGMETRICS, California, USA, pp.118-119, 2000.

[83] Kristensen F. and Sandgren M., "A system simulation of wireless local area networks operating in the 5 GHz band", Master's thesis, Lunds Institute of Technology, Sweden, 2001.

[84] LaMaire R. O., Krishna A. and Bhagwat P., "Wireless LANs and mobile networking: standards and future directions", IEEE Communications Magazine, vol. 34, no. 8, pp. 86-94, Aug. 1996.

[85] Lettieri P. and Srivastava M. B., "Adaptive Frame Length Control for Improving Wireless Network Link Throughput, Range and Energy Efficiency", in Proceedings of the Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 1998), vol. 2, pp. 564 – 571, 1998.

[86] Ludwig R., Konrad A., Joseph A. and Katz R., "Optimizing the End-to-End Performance of Reliable Flows over Wireless Links", Wireless Networks Journal (Special Issue: Selected papers from MobiCom '99), vol. 8, pp. 289-299, ACM and Baltzer Science, 2002.

[87] Mallick M., "Mobile and Wireless Design Essentials", John Wiley & Sons, 2003.

[88] McCullough R. H., "The binary regenerative channel", Bell Syst. Tech. J., vol. 47, pp. 1713–1735, Oct. 1968.

[89] Natkaniec M. and Pach A. R., "An Analysis of the Influence of the RTS_Threshold Parameter on the IEEE 802.11 Network Performance", in Proceedings of the Wireless Communications and Networking Conference (WCNC 2000), Chicago, U.S.A., vol. 2, pp. 819 – 823, Sep. 2000.

[90] Natkaniec M. and Pach A. R., "An Analysis of Modified Backoff Mechanism in IEEE 802.11 Networks", in Proceedings of the Polish – German Teletraffic Symposium (PGTS 2000), Dresden, Germany, pp. 89-96, Sep. 2000.

[91] Natkaniec M. and Pach A. R., "PUMA – A New Channel Access Protocol for Wireless LANs", in Proceedings of the Wireless Personal Multimedia Communications (WPMC 2002), Honolulu, Hawaii, U.S.A., vol. 3, pp. 1351 – 1355, October 2002.

[92] Negus K. J., Stephens A. P. and Lansford J., "HomeRF: wireless networking for the connected home", IEEE Personal Communications Magazine, vol. 7, no. 1, pp. 20-27, Feb. 2000.

[93] Nguyen G. T., Katz R. H., Noble B. and Satyanarayanan M., "A Trace-Based Approach for Modeling Wireless Channel Behavior", in Proceedings of the Winter Simulation Conference, Coronado, CA, pp. 597–604, Dec. 1996.

[94] Ozugur T., Naghshineh M., Kermani P. and Copeland J. A., "On the performance of ARQ protocols in infrared networks", International Journal of Communication Systems, vol.13, pp. 617–638, 2000.

[95] Ozugur T., Copeland J. A., Naghshineh M. and Kermani P., "Next-generation indoor infrared LANs: issues and approaches", IEEE Personal Communications, vol. 6, no. 6, pp. 6-19, Dec. 1999.

[96] Ozugur T., Naghshineh M., Kermani P., Olsen C. M., Rezvani B. and Copeland J. A., "Performance evaluation of L-PPM links using repetition rate coding", in Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 1998), Boston, USA, pp. 698-702, Sept. 1998.

[97] Ozugur T., Naghshineh M.; Kermani P. and Copeland J. A., "Fair media access for wireless LANs", in Proceedings of IEEE Global Telecommunications Conference (Globecom 1999), vol. 1b, pp. 570-579, 1999.

[98] Ozugur T., Naghshineh M., Kermani P., Olsen C. M., Rezvani B. and Copeland J. A., "Balanced media access methods for wireless networks", in Proceedings of IEEE/ACM MobiCom'98, pp. 21-32, Oct. 1998.

[99] Ozugur T., Naghshineh M., Kermani P., Olsen C. M., Rezvani B. and Copeland J. A., "ARQ protocol for infrared wireless LANs: packet-level ACK or no-packet-level ACK?", in Proceedings of the IEEE International Conference on Universal Personal Communications, (ICUPC '98), vol. 2, pp. 1235-1239, 1998.

[100] Paulson L. D., "Exploring the wireless LANscape", Computer, vol. 33, no. 10, pp. 12-16, Oct. 2000.

[101] Papadimitratos P. and Haas Z. J., "Secure Data Transmission in Mobile Ad Hoc Networks", in Proceedings of the Second ACM Workshop on Wireless Security (WiSe'03), in conjunction with ACM MobiCom 2003, Atlanta, USA, pp. 41-50, Sep. 2003.

[102] Qiao D. and Choi S., "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs", IEEE Transactions on Mobile Computing, vol. 1, issue 4, pp. 278 – 292, Oct.-Dec. 2002.

[103] Ramamurthi B., Goodman D. J. and Saleh A., "Perfect Capture for Local Radio Communications", IEEE JSAC, vol. 5, no. 5, pp. 806–13, June 1987.

[104] Rappaport T. S., "Wireless Communications: Principles and Practice", Prentice Hall, 2002.

[105] Raptis P., Vitsas V., Paparrizos K., Chatzimisios P. and Boucouvalas A. C., "Packet Delay Distribution of the IEEE 802.11 Distributed Coordination Function", accepted (to appear) in Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2005), Taormina, Italy, June 2005.

[106] Ritter M. B., Gfeller F., Hirt W., Rogers D. and Gowda S., "Circuit and system challenges in IR wireless communication", in Proceedings of the IEEE International Solid-State Circuits Conference (42nd ISSCC), Digest of Technical Papers, pp. 398-399, 1996.

[107] Sadeghi B., Kanodia V., Sabharwal A. and Knightly E., "Opportunistic Media Access for Multirate Ad Hoc Networks", in Proceedings of the ACM MOBICOM, Georgia, USA, pp. 24-35, 2002.

[108] Sakidu M. N. O., "Optical and Wireless Communications – Next Generation Networks", CRC press, 2002.

[109] Sheu S.-T., Chen T., Chen J. and Ye F., "The impact of RTS Threshold on IEEE 802.11 MAC protocol", in Proceedings of the 9th IEEE International Conference on Parallel and Distributed Systems (ICPADS 2002), pp. 267-272, Dec. 2002.

[110] Sheu S.-T., Chen T., Chen J. and Ye F., "An Improved Data Flushing MAC Protocol for IEEE 802.11 Wireless Ad Hoc Network", in Proceedings of the IEEE 55th Vehicular Technology Conference (VTC Fall 2002), vol. 4, pp. 2435-2439, 2002.

[111] Stajano F. and Anderson R., "The Resurrecting Duckling: Security Issues for Ad Hoc Wireless Networks", in Proceedings of 7th International Workshop on Security Protocols, LNCS, Springer-Verlag, pp. 172 – 194, 1999.

[112] Stallings W., "Wireless Communications and Networks", Prentice Hall, 2002.

[113] Sakidu M. N. O., "Optical and Wireless Communications – Next Generation Networks", CRC press, 2002.

[114] Zuoyin T., Zongkai Y., Jianhua H. and Yanwei L., "Impact of bit errors on the performance of DCF for wireless LAN", in Proceedings of IEEE Conference on Communications, Circuits and Systems and West Sino Expositions, vol. 1, pp. 529-533, 2002.

[115] Tay Y. C. and Chua K. C., "A capacity analysis for the IEEE 802.11 MAC protocol", Wireless Networks, Volume 7, Issue 2 , pp. 159-171, March 2001.

[116] Turin W. and Nobelen R. van, "Hidden Markov modeling of flat fading channels", IEEE Journal Selected Areas in Communications, vol. 16, pp. 1809–1817, Dec. 1998.

[117] Hadzi-Velkov Z. and Spasenovski B., "Saturation throughput - delay analysis of IEEE 802.11 DCF in fading channel", in Proceedings of the IEEE International Conference on Communications (ICC 2003), pp. 121 – 126, 2003.

[118] Vishnevsky V. M and Lyakhov A. I., "IEEE 802.11 Wireless LAN: Saturation Throughput Analysis with Seizing Effect Consideration", Cluster Computing, vol. 5, pp.133–144, 2002.

[119] Vitsas V., "Link layer protocol performance of indoor infrared wireless communications", PhD thesis, Bournemouth University, UK, 2002.

[120] Vitsas V. and Boucouvalas A. C., "Automatic repeat request schemes for infrared wireless communications", IEE Electronics Letters, vol. 38, no. 5, pp. 254-246, Feb 2002.

[121] Vitsas V. and Boucouvalas A. C., "Effectiveness of packet level acknowledgement in infrared wireless LANs", in Proceedings of the IEEE 55th Vehicular Technology Conference, (VTC Spring 2002), Birmingham, AL, vol. 4, pp. 1814-1818, May 2002.

[122] V. Vitsas and A.C. Boucouvalas, 'Effectiveness of Selective Reject (SREJ) Automatic Repeat Request (ARQ) scheme with RR-coding in Infrared Wireless LANs', in Proceedings of the 3rd International Symposium on Communication Systems, Networks and Digital Signal Processing, Stafford, UK, pp.495-498, July 2002.

[123] V. Vitsas and A.C. Boucouvalas, "IrDA IrLAP protocol performance and optimum link layer parameters for maximum throughput", in Proceedings of the

IEEE Global Telecommunications Conference (Globecom 2002), Taipei, Taiwan, vol. 3, pp. 2270-2275, Nov. 2002.

[124] Vitsas V. and Boucouvalas A. C., "Performance Analysis of the Collision Avoidance Procedure of the Advanced Infrared (AIr) CSMA/CA Protocol for Wireless LANs", in Proceedings of the IEEE 55th Vehicular Technology Conference, (VTC Spring 2002), vol.3, pp. 1502-1506, 2002.

[125] Vitsas V. and Boucouvalas A. C., "Performance Analysis of the AIr-MAC Optical Wireless Protocol", in Proceedings of International Conference on System Engineering, Communications and Information Technologies, (ICSECIT 2001), Punta Arenas, Chile, 2001.

[126] Vitsas V. and Boucouvalas A. C., "Performance Evaluation of IrDA Advanced Infrared AIr -MAC Protocols", in Proceedings of the 5th Multi-Conference on Systemics, Cybernetics, and Informatics (SCI 2001), Orlando, USA, pp. 347-352, 2001.

[127] V. Vitsas and A.C. Boucouvalas, "Simultaneous optimisation of window and frame size for maximum throughput IrDA links", IEE Electronics Letters, vol. 37, no. 16, pp.1042-1043, 2nd August 2001.

[128] V. Vitsas, "Throughput analysis of linear backoff scheme for wireless LANs", IEE Electronics Letters, vol. 39, no 1, pp. 99-100, 9th January 2003.

[129] Vitsas V., Chatzimisios P., Boucouvalas A.C., Raptis P., Paparrizos K. and Kleftouris D., "Enhancing performance of the IEEE 802.11 Distributed Coordination Function via packet bursting", invited paper in Proceedings of the IEEE Global Telecommunications Conference (Globecom 2004), Dallas Texas, USA, pp. 245-252, 29 Nov.- 3 Dec., 2004.

[130] Wang H. S. and Moayeri N., "Finite state Markov channel—A useful model for radio communication channels", IEEE Transactions in Vehicular Technology, vol. 44, pp. 163–171, Feb. 1995.

[131] Weinmiller J., Woesner H., Ebert J-P. and Wolisz A., "Analysing the RTS/CTS Mechanism in the DFWMAC Media Access Protocol for Wireless LANs", in Proceedings of Personal Wireless Communications (PWC 1995), Czech Republic, pp. 117-130, 1995.

[132] Weinmiller J., Woesner H., Ebert J-P. and Wolisz A., "Analyzing and Tuning the Distributed Coordination Function in the IEEE 802.11 DFWMAC Draft Standard", in Proceedings of MASCOT, California, pp. 200-206, 1996.

[133] Willig A., Kubisch M., Hoene C., and Wolisz A., "Measurements of a Wireless Link in an Industrial Environment using an IEEE 802.11-Compliant Physical Layer", IEEE Transactions on Industrial Electronics, vol. 49, issue 6, pp. 1265 – 1282, Dec. 2002.

[134] Willig A., Kubisch M. and Wolisz A., "Measurements and stochastic modeling of a wireless link in an industrial environment", Technical report, Telecommunication Networks Group, Tech. Univ. Berlin, Mar. 2001.

[135] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, ISO/IEC 8802-11:1999(E), 1999.

[136] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY): High-Speed Physical Layer Extension in the 2.4 GHz Band, IEEE 802.11b WG, 1999.

[137] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY): High-Speed Physical Layer Extension in the 5 GHz Band, IEEE 802.11a WG, 1999.

[138] Wisely D., "A 1 Gbit/s optical wireless tracked architecture for ATM delivery", IEE Colloquium on Optical Free Space Communication, London, pp. 14/1-14/7, Feb. 96.

[139] Wu H., Peng Y., Long K., Cheng S. and Ma J., "Performance of Reliable Transport Protocol over IEEE 802.11 Wireless LAN: Analysis And Enhancement", in Proceedings of the Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), vol.2, pp. 599-607, 2002.

[140] Wu H., Long K., Cheng S. and Ma J., "IEEE 802.11 Distributed Coordination Function (DCF): Analysis and Enhancement", in Proceedings of International Conference on Communications (ICC 2002), vol. 1, pp. 605 – 609, April 2002.

[141] Xiao Y., "A Simple and Effective Priority Scheme for IEEE 802.11", IEEE Communications Letters, vol. 7, no. 2, pp. 70-72, 2003.

[142] Xiao Y. and Rosdahl J., "Throughput and Delay Limits of IEEE 802.11," IEEE Communications Letters, Vol. 6, No. 8, Aug. 2002, pp. 355-357.

[143] Xu S. and Saadawi T., "Revealing the problems with 802.11 medium access control protocol in multi-hop wireless ad hoc networks", Computer Networks, vol 38, no. 4, pp. 531-548, 2002.

[144] Xu K., Gerla M. and Bae S., "How Effective is the IEEE 802.11 RTS/CTS Handshake in Ad Hoc Networks?", in Proceedings of the IEEE Global Telecommunications Conference (Globecom 2002), Taipei, vol. 1, pp. 72-76, 2002.

[145] Xu K., Gerla M. and Bae S., "Effectiveness of RTS/CTS Handshake in IEEE 802.11 Based Ad Hoc Networks", Ad Hoc Networks Journal, vol. 1, no.1, pp. 107-123, July 2003.

[146] Yeo J. and Agrawala A., "Packet Error Model for the IEEE 802.11 MAC Protocol", in Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2003), Beijing, China, vol. 2, pp. 1722 – 1726, Sep. 2003.

[147] Zhou L. and Haas Z. J., "Securing Ad Hoc Networks," IEEE Network Magazine, vol,13, no.6, pp. 24 – 30, Nov./Dec. 1999.

[148] Ziouva E. and Antonakopoulos T., "CSMA/CA performance under high traffic conditions: throughput and delay analysis", Computer Communications, vol. 25, pp. 313-321, 2002.

[149] Zorzi M., Rao R. R. and Milstein L. B., "A Markov model for block errors on fading channels", in Proceedings of Personal, Indoor and Mobile Radio Communications (PIMRC 96), vol. 3, pp. 1074 – 1078, Oct 1996.

[150] Zorzi M., Rao R. R. and Milstein L. B., "On the accuracy of a first order Markov model for data transmission on fading channels", in Proceedings of the 4th International Conference on Universal Personal Communication, Tokyo, Japan, pp. 211 – 215, Nov. 1995.

APPENDIX A IEEE 802.11 BACKGROUND

A.1 IEEE 802.11 standard architecture

IEEE 802.11 standard defines a hierarchical network architecture that enables WLAN equipment to be configured in a variety of ways. The basic and smallest building block of a WLAN is a Basic Service Set[1] (BSS), which is simply a group of stations executing the same MAC protocol and competing for access to the same shared wireless medium. An BSS may be isolated or it may connect to a backbone distribution system (DS) through an access point (AP) that functions like a bridge. The DS can be a switch, a wired network or a wireless network.

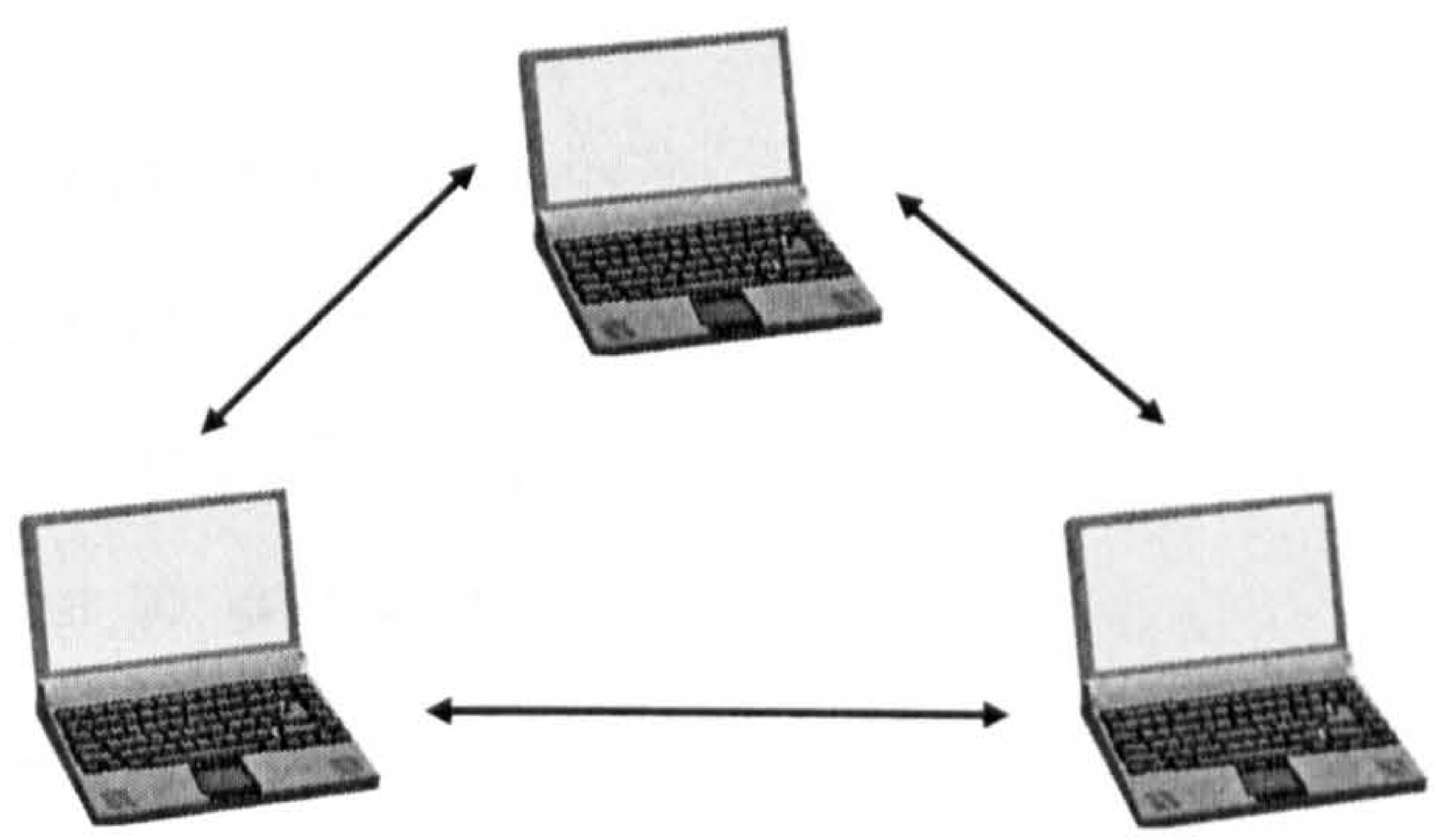


Figure A.1 An example of an IBSS

The simplest configuration is called Independent Basic Service Set (IBSS). An IBSS is a self-contained network of a minimum of two mobile stations, where direct communication between stations is supported, without the use of any kind of centralized coordination (coordination of the channel is distributed among the stations). Figure A.1 illustrates how three stations can form an IBSS and communicate directly with each other. These networks do not involve pre-planning, they are usually composed of a small number of stations set up for a specific purpose and for a short period of time. A typical example would be some people with laptops create short-lived network to support a single meeting or to collaborate on a presentation at a conference. Due to their short duration, small size and focused purpose, IBSSs are sometimes referred to as ad hoc BSSs or ad hoc networks.

---

[1] A service set is a logical grouping of devices.

When a BSS includes one station acting as AP that has access to the wired network, the BSS is no longer independent and is called an infrastructure BSS but referred simply as BSS (never called an IBSS). In a BSS, all stations do not communicate directly with other stations but through the AP. Thus, if mobile stations in the BSS want to communicate with each other, they communicate with the AP and the AP forwards the data to the destination stations. This causes communication to consume twice the bandwidth that the same communication would consume if sent directly from one mobile station to another.

The interconnection of two or more BSSs via the DS is called an Extended Service Set (ESS) and extends the range of station's mobility. Typically, the distribution system is a wired backbone LAN but can be any communications network. An ESS appears as a single logical WLAN to the Logical Link Control (LLC) layer. A simple but comprehensive configuration is shown in figure A.2, in which each station (STA) belongs to a single BSS and is within range only of other stations within the same BSS. The AP is implemented as part of the station that provides access to DS by providing DS services in addition to acting as a station. In order to integrate the IEEE 802.11 architecture with a traditional wired LAN, a portal is utilized (usually a device such as a bridge or a router). In the real-world deployment generally it is possible for two BSSs to overlap geographically and a station to participate in more than one BSS.

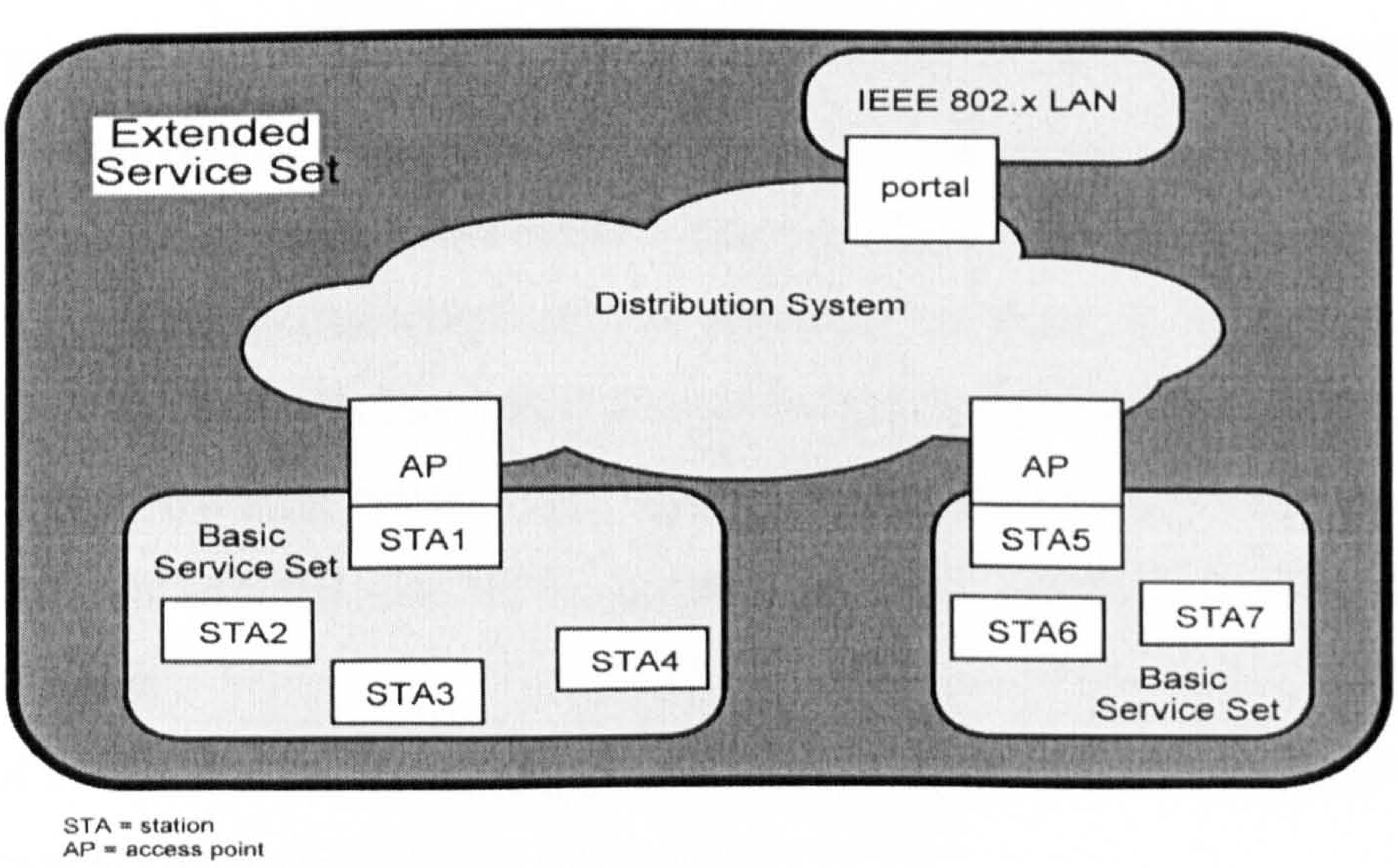


Figure A.2 An example of IEEE 802.11 Architecture

Figure A.3 shows the most common scenario; an IEEE 802.11 WLAN connected to a switched IEEE 802.3 Ethernet via a bridge. The upper part of the data link control layer, the logical link control (LLC), covers the differences of the medium access control layers needed for the different media. This allows existing network protocols to run over IEEE 802.11 without any special consideration. Applications should not notice any difference apart from a lower bandwidth. Therefore, the higher layers (application, TCP/IP, etc.) in a wireless station look the same as the wired station.

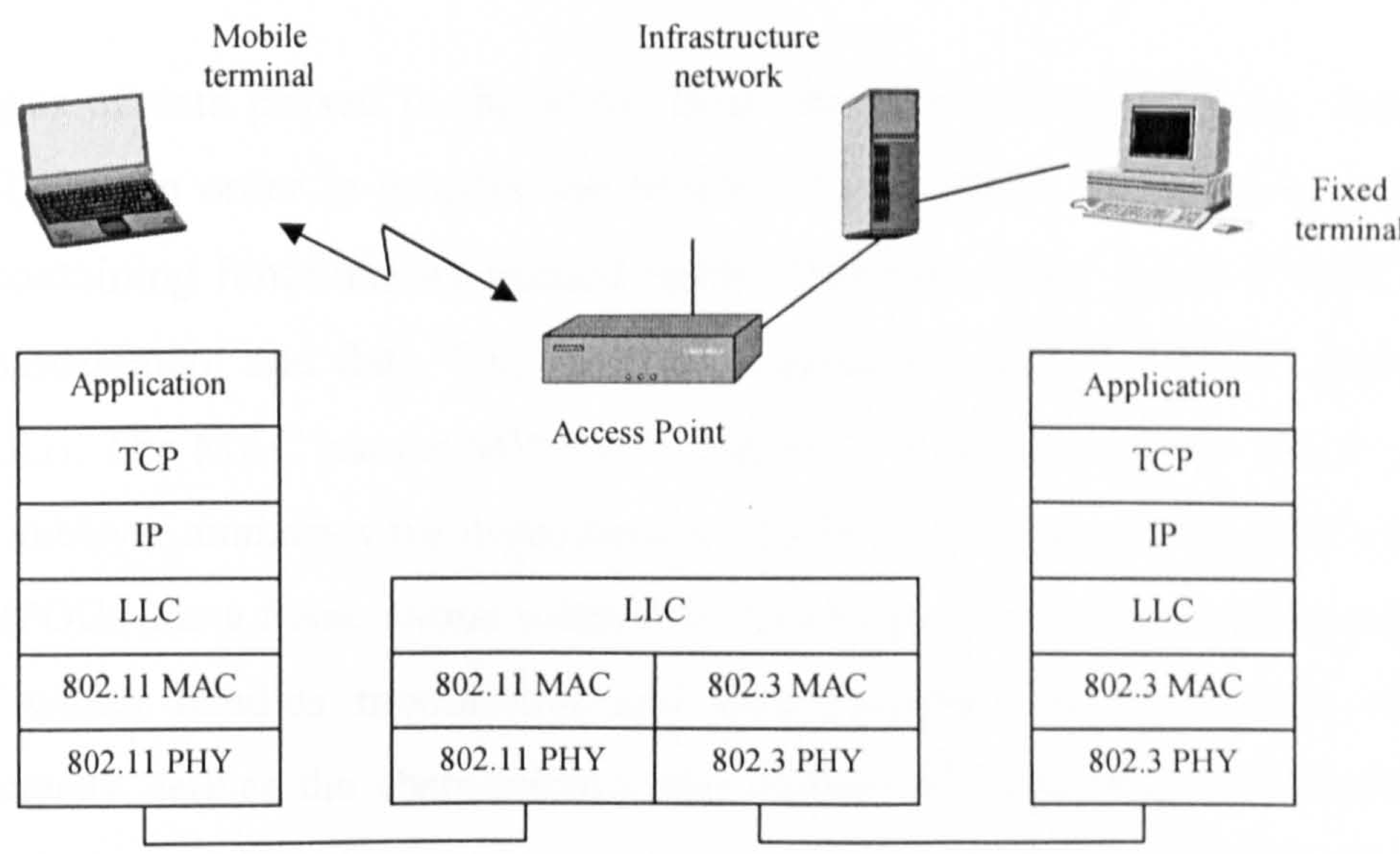


Figure A.3 An example of IEEE 802.11 Architecture

The IEEE 802.11 standards only cover the Medium Access Control (MAC) layer and the physical layer (PHY) like the other 802.x LANs do. The basic tasks of the MAC layer are medium access, fragmentation of user data, encryption. The IEEE 802.11 PHY essentially provides wireless transmission mechanisms for the MAC layer, in addition to supporting secondary functions such as assessing the state of the wireless medium and reporting it to MAC. By providing these transmission mechanisms independently of the MAC, IEEE Working Group (WP) has developed advances by adding the higher data rate 802.11b and 802.11a PHYs. In fact, the MAC layer for each of the 802.11 PHYs is the same.

Each of the 802.11 PHYs is subdivided in two sublayers (shown in figure A.4):

- Physical Layer Convergence Procedure (PLCP)
- Physical Medium Dependant (PMD)

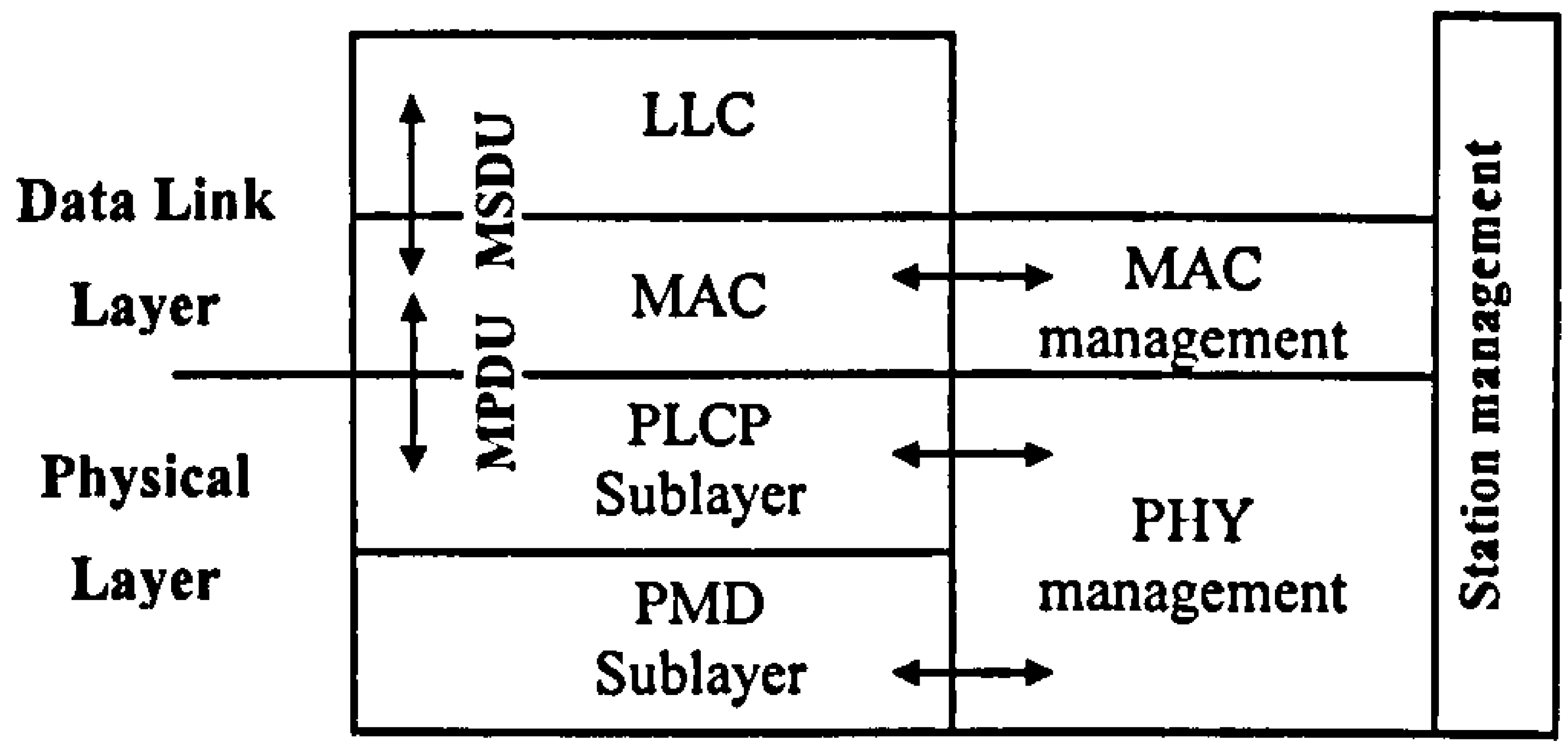


Figure A.4 PHY sublayers and protocol management

Packages of data passed to the MAC from the LLC are called MAC service data units (MSDUs). In order to transfer the MSDUs to the PHY, the MAC uses messages (frames) containing functionality related fields. There are three types of MAC frames: control, management and data. One of these messages is called a MAC protocol data unit (MPDU). The MAC passes MPDUs to the PHY layer through the PLCP sublayer. The PLCP sublayer minimizes the dependence of the MAC sublayer on the PMD sublayer by mapping MPDUs into a frame format suitable for transmission over the wireless medium by the PMD, which handles modulation and encoding/decoding of signals. The PMD sublayer actually defines the characteristics and method of transmitting and receiving data through a wireless medium between two or more stations. Moreover, the PLCP sublayer provides a carrier sense signal, called clear channel assessment (CCA). This needed for the MAC mechanisms controlling the medium access and indicates if the medium is currently idle.

Apart from the protocol sublayers, the 802.11 standards specify management layers (MAC and PHY management) and the station management. The MAC management controls authentication mechanisms, encryption and power management. The main tasks of the PHY management includes channel tuning, whereas station management interacts with both management layers and is responsible for higher layer functions.

A.2 IEEE 802.11 Services

IEEE 802.11 defines nine services that need to be provided by the WLAN. Table A.1 lists the various services to manage authentication, de-authentication, privacy and data transfer and indicates two ways of categorizing them. These services are divided into the Station Service (SS) and the Distribution System Service (DSS)[2]:

---

[2] More details can be found in [104] and [112].

Service	Provider	Used to support
Association	Distribution system	MSDU delivery
Dissassociation	Distribution system	MSDU delivery
Authentication	Station	LAN access and security
Deauthentication	Station	LAN access and security
Distribution	Distribution system	MSDU delivery
Integration	Distribution system	MSDU delivery
Data transfer	Station	MSDU delivery
Privacy	Station	LAN access and security
Reassocation	Distribution system	MSDU delivery

Table A.1 IEEE 802.11 Services

Association: The association service enables the establishment of wireless links between wireless clients and APs in infrastructure networks.

Disassociation: The service, which cancels the wireless links between wireless clients and APs in infrastructure networks.

Authentication: The authentication service is the process of proving client identity, which takes place prior to a wireless client associating with an AP. By default, IEEE 802.11 devices operate in an Open System, where essentially any wireless client can associate with an AP without checking credentials. True authentication is possible with the use of the 802.11 option known as *Wired Equivalent Privacy* or WEP, where a shared key is configured into the AP and its wireless clients. Only those devices with a valid shared key will be allowed to be associated to the AP.

De-authentication: The de-authentication function is performed by the base station. It is a process of denying client credentials, based on incorrect authentication settings, or applied IP or MAC filters.

Distribution: The distribution function is performed by DS and it is used in special cases in frame transmission between APs.

Integration: This is a function performed by the *portal*, where essentially the portal is design to provide logical integration between existing wired LANs and 802.11 LANs.

Privacy: By default, data is transferred in the clear allowing any 802.11-compliant device to potentially eavesdrop on similar PHY 802.11 traffic within range. The WEP option encrypts data before it is sent, using a 40-bit encryption algorithm known as RC4. The same shared key used in authentication is used to encrypt or decrypt the data, allowing only wireless clients with the exact shared key to correctly decode the data.

***Re-association*:** The re-association service occurs in addition to association when a wireless client moves from one BSS to another. Two adjoining BSSs form an ESS if they are defined by a common ESSID, providing a wireless client with the capability to roam from one area to another. Although reassociation is specified in 802.11, the mechanism that allows AP-to-AP coordination to handle roaming is not specified.

***Data transfer*:** The primary service of MAC layer is to provide frame exchange between MAC layers. Wireless clients use a Collision Sense Multiple Access with Collision Avoidance (CSMA/CA) algorithm as the media access scheme.

A.3 IEEE 802.11 framing in detail

There are three categories of frames in the IEEE 802.11 MAC:

- Control frames – These frames facilitate data frames during 802.11 data exchanges.
- Management frames – These frames facilitate WLAN connectivity, authentication and status.
- Data frames – These frames carry station data between the transmitter and the receiver.

All 802.11 MAC frames leverage the 802.11 general frame. The three frame types use specific portions of the general MAC frame for their specific purposes. Each frame consists of the following basic components:

a) A *MAC header*, which comprises frame control, duration, address, and sequence control information;
b) A variable length *frame body* that contains information specific to the frame *type*;
c) A *frame check sequence* (FCS), which contains a 32-bit Cyclic Redundancy Code (CRC).

The MAC protocol data units (MPDUs) or frames in the MAC sublayer are described as a sequence of fields in specific order. Each figure in this section depicts the fields/subfields as they appear in the MAC frame and in the order in which they are passed to the physical layer convergence protocol (PLCP), from left to right. In the following figures, all bits within fields are numbered, from 0 to k, where the length of the field is k+1 bits. The octet (byte) boundaries within a field can be obtained by taking the bit numbers of the field modulo 8. Octets within numeric fields that are longer than a single octet are depicted in increasing order of significance, from lowest numbered bit to highest numbered bit.

The MAC frame format comprises a set of fields that occur in a fixed order in all frames. The fields Address 2, Address 3, Sequence Control, Address 4 and Frame Body are only present in certain frames. Figure A.5 depicts the general MAC frame format.

Frame Control	Duration/ID	Address 1	Address 2	Address 3	Sequence Control	Address 4	Frame Body	FCS
2 bytes	2 bytes	6 bytes	6 bytes	6 bytes	2 bytes	6 bytes	0-2312 bytes	4 bytes

Figure A.5 MAC frame format

A.3.1 Frame fields

A.3.1.1 Frame control field

The Frame Control field consists of the following subfields: Protocol Version, Type, Subtype, To DS, From DS, More Fragments, Retry, Power Management, More Data, Wired Equivalent Privacy (WEP), and Order. The format of the Frame Control field is illustrated in figure A.6.

Protocol Version	Type	Subtype	To DS	From DS	More Frag	Retry	Pwr Mgt	More Data	WEP	Order
2 bits	2 bits	4 bits	1 bit	1 bit	1 bit	1 bit	1 bit	1 bit	1 bit	1 bit

Figure A.6 Frame Control Field

A.3.1.1.1 Protocol Version field

The 2-bit Protocol Version field contains the version of the standard. For the current standard, the value of the protocol version is 0. A device that receives a frame with a higher revision level than it supports will discard the frame without indication to the sending station or to LLC layer.

A.3.1.1.2 Type and Subtype fields

The Type field is 2 bits in length, and the Subtype field is 4 bits in length. Together the two fields identify the function of the frame (control, management and data). Each of the frame types have several defined subtypes, some of them shown in the table A.2.

A.3.1.1.3 To DS and From DS fields

Both fields are 1 bit in length. The To DS field is set to "1" in data type frames destined for the DS. The From DS field is set to "1" in data type frames exiting the DS.

A.3.1.1.4 More Fragments field

The More Fragments field is 1 bit in length and is set to "1" in all data or management type frames that have another fragment of the current MSDU or current MMPDU to follow. It is set to 0 in all other frames.

A.3.1.1.5 Retry field

The 1-bit Retry field is set to "1" in any data or management type frame that is a retransmission of an earlier frame. It is set to 0 in all other frames. A receiving station can utilize this indication to aid in the process of eliminating duplicate frames.

Type Value	Type description	Subtype value	Subtype description
00	Management	0000	Association request
00	Management	0001	Association response
00	Management	0110-0111	Reserved
00	Management	1000	Beacon
00	Management	1101-1111	Reserved
01	Control	0000-1001	Reserved
01	Control	1011	Request To Send (RTS)
01	Control	1100	Clear To Send (CTS)
01	Control	1101	Acknowledgement (ACK)
01	Control	1110	Contention Free (CF)-End
10	Data	0000	DATA
10	Data	0001	DATA + CF-ACK
10	Data	0010	DATA + CF-Poll
10	Data	0011	DATA+ CF- ACK + CF- Poll
10	Data	0100	Null Function (no data)
10	Data	0101	CF- ACK (no data)

Table A.2 Valid Type/Subtype combinations

A.3.1.1.6 Power Management field

The Power Management field is 1 bit in length and is used to indicate the power management mode of a STA. The value of this field indicates the mode in which the station will be after the successful completion of the frame exchange sequence. A value of 1 indicates that the STA will be in power-save mode. A value of 0 indicates that the STA will be in active mode.

A.3.1.1.7 More Data field

The 1-bit More Data field is used to indicate to a STA in power-save mode that more MSDUs, or MMPDUs are buffered for that STA at the AP. A value of "1" indicates that at least one additional buffered MSDU, or MMPDU, is present for the same STA. The More Data field may be set to 1 in directed data type frames transmitted by a contention-free (CF)-Pollable STA to the point coordinator (PC) in response to a CF-Poll to indicate that the STA has at least one additional buffered MSDU available for transmission in response to a subsequent CF-Poll. The More Data field is set to 0 in all other directed frames. All uses of the More Data field are in detail reported in [135] and are out of the scope of this thesis.

A.3.1.1.8 WEP field

The Wired Equivalent Privacy (WEP) field is 1 bit in length and is set to "1" if the Frame Body field contains information that has been processed by the WEP algorithm. The WEP field is set to 0 in all other frames.

A.3.1.1.9 Order field

The Order field is 1 bit in length and is set to "1" in any data type frame that contains an MSDU, or fragment thereof, which is being transferred using the StrictlyOrdered service class. This field is set to 0 in all other frames.

A.3.1.2 Duration/ID field

The Duration/ID field is 16 bits in length and is used differently depending on whether a power save station is accessing the medium, the medium is in a PCF mode or a DCF station is accessing the medium. The contents of the Duration/ID field are as follows:

a) In control type frames of subtype Power Save (PS)-Poll, the Duration/ID field carries the association identity (AID) of the station that transmitted the frame in the 14 least significant bits (LSB), with the 2 most significant bits (MSB) both set to 1. The value of the AID is in the range 1–2007.

b) In all other frames, the Duration/ID field contains a duration value that is used in the medium access control algorithm such that it contains the amount of time (in us) the current transmission will be occupying the medium. For frames transmitted during the contention-free period (CFP), the duration field is set to 32,768. Whenever the contents of the Duration/ID field are less than 32,768, the duration value is used to update the network allocation vector (NAV). More details on the Duration/ID field can be found in [135].

A.3.1.3 Address fields

An 802.11 frame may contain up to four 48-bit address fields. The four types of addresses are the BSS identifier[3] (BSSID), DA, SA, RA, and TA, indicating basic service set identifier (BSSID), Destination Address, Source Address, Receiver Address, and Transmitter Address, respectively. Certain frames may not contain some of the address fields. Certain address field usage is specified by the relative position of the address field (1–4) within the MAC header, independent of the type of address present in that field. The general rule is that Address 1 is used for the receiver, Address 2 for the transmitter and Address 3 for filtering by the receiver.

A.3.1.4 Sequence Control field

The Sequence Control field is 16 bits in length and consists of two subfields, the Fragment Number (4 bits) and the Sequence Number (12 bits). Figure A.7 illustrates the format of the Sequence Control field.

Fragment Number	Sequence Number
4 bits	12 bits

Figure A.7 Sequence Control field

[3] Each BSS is assigned a BSSID that distinguishes it from other BSSs throughout the network.

A.3.1.4.1 Fragment Number field

The Fragment Number field is a 4-bit field indicating the number of each fragment of an MSDU or MMPDU. The fragment number is set to 0 in the first or only fragment of an MSDU or MMPDU and is incremented by one for each successive subsequent fragment of that MSDU or MMPDU. The fragment number remains constant in all retransmissions of the fragment.

A.3.1.4.2 Sequence Number field

The Sequence Number field is a 12-bit field indicating the sequence number of an MSDU or MMPDU and is used to detect duplicate frames. Each MSDU or MMPDU transmitted by a STA is assigned a sequence number. Sequence numbers are assigned from a single modulo 4096 counter, starting at 0 and incrementing by 1 for each MSDU or MMPDU. Each fragment of an MSDU or MMPDU contains the assigned sequence number. The sequence number remains constant in all retransmissions of an MSDU, MMPDU, or fragment thereof.

A.3.1.5 Frame Body field

The Frame Body is a variable length field that contains information specific to individual frame types and subtypes. The minimum frame body is 0 octets.

A.3.1.6 FCS field

The 32-bit Frame Check Sequence (FCS) field contains a 32-bit Cyclic Redundancy Check (CRC). The FCS is calculated over all the fields of the MAC header and the Frame Body field.

A.3.2 Format of individual frame types

A.3.2.1 Control frames

Control frames assist in the reliable delivery of data frames. The sub-fields within the Frame Control field of control frames are set as illustrated in figure A.8.

Protocol Version	Type	Subtype	To DS	From DS	More Frag	Retry	Pwr Mgt	More Data	WEP	Order
Protocol Version	Type	Subtype	0	0	0	0	Pwr Mgt	0	0	0
2 bits	2 bits	4 bits	1 bit	1 bit	1 bit	1 bit	1 bit	1 bit	1 bit	1 bit

Figure A.8 Frame Control field sub-field values within control frames

A.3.2.1.1 Request To Send (RTS) frame format

The frame format for the RTS frame is as defined in figure A.9.

MAC header

Frame Control	Duration	RA	TA	FCS
2 bytes	2 bytes	6 bytes	6 bytes	4 bytes

Figure A.9 RTS frame format

The RA of the RTS frame is the address of the station, on the WM, that is the intended immediate recipient of the pending directed data or management frame or RTS. The TA is the address of the station transmitting the RTS frame. The duration value is the time in microseconds (ms) required to transmit the pending data or management frame, plus one CTS frame, plus one ACK frame, plus three SIFS intervals.

A.3.2.1.2 Clear To Send (CTS) frame format

The frame format for the CTS frame is as defined in figure A.10.

MAC header

Frame Control	Duration	RA	FCS
2 bytes	2 bytes	6 bytes	4 bytes

Figure A.10 CTS frame format

The RA of the CTS frame is copied from the TA field of the immediately previous RTS frame to which the CTS is a response. The duration value is the value obtained from the Duration field of the immediately previous RTS frame, minus the time, in ms, required to transmit the CTS frame and its SIFS interval.

A.3.2.1.3 Acknowledgment (ACK) frame format

Figure A.11 defines the frame format for the ACK frame.

MAC header

Frame Control	Duration	RA	FCS
2 bytes	2 bytes	6 bytes	4 bytes

Figure A.11 ACK frame format

The RA of the ACK frame is copied from the Address 2 field of the immediately previous directed data, management, or PS-Poll control frame. If the More Fragment bit was set to 0 in the Frame Control field of the immediately previous directed data or management frame, the duration value is set to 0. If the More Fragment bit was set to 1 in the Frame Control field of the immediately previous directed data or management frame, the duration value is the value obtained from the Duration field of the immediately previous data or management frame, minus the time, in ms, required to transmit the ACK frame and its SIFS interval.

A.3.2.2 Management frames

Management frames are used to manage communication between stations and APs. Functions include management of associations (request, response, reassociation and authentication). The 802.11 management frames consist of the following:

- Beacon
- Probe request
- Probe response
- Authentication
- Deauthentication
- Association request
- Association response
- Reassociation request
- Reassociation response
- Disassociation
- Announcement Traffic Indication

Management frames are not being considered in this thesis and more details are given in [135].

A.3.2.3 Data frames

The 802.11 standard stipulates the following eight data frames:

- DATA
- DATA+CF-ACK
- DATA+CF-Poll
- DATA+CF-ACK+CF-Poll
- Null Function
- CF-ACK
- CF-Poll
- CF-ACK+CF-Poll

The first four subtypes define data frames that carry upper-level data from the source station to the destination station. The DATA is used in both a contention period as well as in a contention-free period. The DATA+CF-ACK data frame is sent during a contention-free (CP) period. In addition to carrying data, this frame acknowledges previously received data. The DATA+CF-Poll data frame is used by a point coordinator (PC) to deliver data to a mobile station and to request that the mobile station send a data frame that it may have buffered. The DATA+CF-ACK+CF-Poll data frame combines the functions of the DATA+CF-ACK+ and DATA+CF-Poll into a single frame.

The remaining four subtypes of data frames do not in fact carry any data. The Null Function data frame is used to carry the power management bit in the frame control field to the AP indicating that the station is changing to a low-power operating state. The remaining three frames (CF-ACK, CF-Poll, CF-ACK+CF-Poll) have the same functionality as the corresponding data frame subtypes in the preceding list (DATA+CF-ACK, DATA+CF-Poll, DATA+CF-ACK+CF-Poll) but without the data.

In fact, DATA is the simplest data frame and is shown in detail in figure A.5. The content of the Address fields of the data frame is dependent upon the values of the To DS and From DS bits. Table A.3 summarizes the use of the address fields in data frames.

To DS	From DS	Address 1 (receiver)	Address 2 (transmitter)	Address 3	Address 4
0	0	DA	SA	BSSID	N/A
0	1	DA	BSSID	SA	N/A
1	0	BSSID	SA	DA	N/A
1	1	RA	TA	DA	SA

Table A.3 Interpretation of the address fields in data frames

Where the content of a field is shown as not applicable (N/A), the field is omitted. Note that Address 1 always holds the receiver address of the intended destination or in the case of multicast frames, multiple destinations. In cases where the Address 1 field contains a group address, the BSSID also is validated to ensure that the broadcast or multicast originated in the same BSS. The Address 2 field always holds the address of the station that is transmitting the frame and is used to direct the acknowledgment if an acknowledgment is necessary. The DA (Destination Address) is the destination of the MSDU (or fragment thereof) in the frame body field. The SA (Source Address) is the address of the MAC entity that initiated the MSDU (or fragment thereof) in the frame body field. The RA (Receiver Address) is the address of the STA contained in the AP in the wireless distribution system that is the next immediate intended recipient of the frame. The TA (Transmitter Address) is the address of the station contained in the AP in the wireless distribution system that is transmitting the frame.

The frame body consists of the MSDU (or a fragment when fragmentation mechanism is utilized). The frame body is null (0 bytes in length) in data frames of Subtype Null function (no data), CF-ACK (no data), CF-Poll (no data), and CF-ACK +CF-Poll (no data). Within all data type frames sent during the CFP, the Duration field is set to the value 32,768. Within all data type frames sent during the contention period, the Duration field is set according to the following rules:

- If the Address 1 field contains a group address, the duration value is set to 0.
- If the More Fragments bit is set to 0 in the Frame Control field of a frame and the Address 1 field contains an individual address, the duration value is set to the time, in ms, required to transmit one ACK frame, plus one SIFS interval.
- If the More Fragments bit is set to 1 in the Frame Control field of a frame, and the Address 1 field contains an individual address, the duration value is set to the time, in ms, required to transmit the next fragment of this data frame, plus two ACK frames, plus three SIFS intervals.

A.4 IEEE 802.11 Physical Layer

The physical layer for IEEE 802.11 has been issued in four stages; the first part was issued in 1997 [135], two additional parts in 1999 [136][137] and the most recent in 2003. The first part, simply called IEEE 802.11, includes the MAC layer and three physical layer specifications, two in the 2.4 GHz ISM band and one in the infrared, all operating at 1 Mbit/s and 2 Mbit/s. IEEE 802.11a operates in the 5GHz band at data rates up to 54 Mbit/s. IEEE 802.11b operates in the 2.4 GHz band at 5.5 and 11 Mbit/s. IEEE 802.11g extends IEEE 802.11b to higher data rates up to 54 Mbit/s.

Three physical media were defined in the original 802.11 standard:

- Infrared at a wavelength between 850 and 950 nm, at data rates of 1Mbit/s and 2Mbit/s.
- Frequency Hopping Spread Spectrum operating in the 2.4 GHz ISM band, at data rates of 1Mbit/s and 2Mbit/s.
- Direct Sequence Spread Spectrum operating in the 2.4 GHz ISM band, at data rates of 1Mbit/s and 2Mbit/s.

A.4.1 Original IEEE 802.11

A.4.1.1 Infrared (IR) PHY

The IR implementation uses light in the 850 nm to 950 nm range for signalling. This range is similar to the spectral usage of common consumer devices like infrared remote controls, as well as other data communications equipment, such as IrDA devices. Unlike many other infrared devices, however, the IR PHY is omni-directional rather than point-to-point. That is, the receiver and transmitter do not have to be aimed at each other and do not need a clear line of sight. This permits the construction of a LAN system, whereas with an aimed system, it would be difficult to establish a LAN because of physical constrains. According to the standard [135], a pair of conformant infrared devices would be able to communicate in a typical environment at a range up to about 10 meters. The standard allows conformant devices to have more sensitive receivers and, thus, a range of up to 20 meters is possible. The Infrared links rely on both reflected infrared energy and line-of-sight for communication. The standard specifies the transmitter and receiver in such a way that a conformant design will operate well in most environments where there is no line-of-sight path between transmitter and receiver. However, in an environment that has few or no reflecting surfaces and where there is no line-of-sight, an Infrared system may suffer reduced range. Infrared radiation does not pass through walls. For this reason, the IR PHY operates only in indoor environments, usually in a single physical room, like a classroom or conference room. Different LANs using the IR PHY can operate in adjacent rooms separated only by a wall without interference and without the possibility of eavesdropping.

The IR PHY utilizes Pulse Position Modulation (PPM) to transmit data using IR radiation. The modulation scheme for the 1 Mbit/s data rate (basic access rate) is known as 16-PPM, where 16 symbols are used to transmit each group of 4 data bits. For the 2 Mbit/s data rate (enhanced access rate), each group of 2 data bits is mapped into one of four 4-bit sequences.

Figure A.12 shows the format for the PLCPDU including the PLCP Preamble, the PLCP Header, and the PSDU. The PLCP Preamble contains the following fields: Synchronization (SYNC) and Start Frame Delimiter (SFD). The PLCP Header contains the following fields: Data Rate (DR), DC Level Adjustment (DCLA), Length (LENGTH) and Cyclic Redundancy Check (CRC). Each of these fields is described in detail in [135].

← PLCP Preamble →		← PLCP Header →				
Sync	SFD	DR	DCLA	LENGTH	CRC	MAC Frame (PSDU)
57-73 bits	4 bits	3 bits	32 bits	16 bits	16 bits	variable

Figure A.12 PLCPDU frame format

The PLCP Preamble is always transmitted at 1 Mbit/s. The LENGTH and CRC fields as well as the PSDU are transmitted at one of two bit rates: 1 Mbit/s or 2 Mbit/s. Any conformant IR PHY shall be capable of receiving at 1 Mbit/s and 2 Mbit/s (transmission at 2 Mbit/s is optional).

A.4.1.2 Frequency-Hopping Spread Spectrum (FHSS) PHY

As the name suggests, Frequency-Hopping Spread Spectrum technology divides a radio signal spectrum into small segments and "hops" from one frequency to another many times per second as it transmits those segments. The transmitter and the receiver establish a synchronized hopping pattern that sets the sequence order in which they will use different subchannels. FHSS systems overcome interference from other users by using a narrow carrier-signal that changes frequency many times every second. Additional transmitter and receiver pairs can use different hopping patterns on the same set of subchannels at the same time. At any given point in time, each transmission is probably using a different subchannel, so there's no interference between signals. When a conflict does occur, the system resends the same packet until the receiver gets a clean copy and sends a confirmation back to the transmitting station.

As a result, FHSS has the advantage of allowing the coexistence of multiple networks in the same area by separating different networks using different hopping sequences from one channel to another. The original standard defines that the available spectrum is splitted into 79 non-overlapping hopping channels for North America and most of Europe, and 23 hopping channels for Japan. Each of the 79 channels is 1 MHz wide in the 2.4 GHz ISM band (across the 2.402 to 2.480 GHz frequency range). Because each frequency hop adds overhead to the data stream, FHSS transmissions are relatively slow.

The standard [135] specifies Gaussian shaped frequency shift keying (GFSK) as modulation for the FHSS PHY. For 1 Mbit/s a two-level GFSK is utilized (i.e. 1 bit is mapped to one frequency) and the bits zero and one are encoded as deviations from the current carrier frequency. For 2 Mbit/s a four-level GFSK is used (i.e. 2 bits are mapped to one frequency) in which four different deviations from the current frequency define the 2-bit combinations. While sending and receiving at 1 Mbit/s is mandatory for all devices, operation at 2 Mbit/s is optional. This facilitated the production of low-cost devices as well as more powerful devices for both transmission rates in the early days of IEEE 802.11. Figure A.13 shows the format of a frame used with the FHSS physical layer. The frame consists of two basic parts, the PLCP part (preamble and header) and the payload part. While the PLCP part is always transmitted at 1 Mbit/s, while the payload can utilize 1 or 2 Mbit/s.

PLCP Preamble		PLCP Header			
Sync	SFD	PLW	PSF	HEC	MAC Frame (PSDU)
80 bits	16 bits	12 bits	4 bits	16 bits	variable

Figure A.13 Format of a PHY frame using FHSS

The PLCP preamble consists of two separate subfields:

- **Sync:** The PLCP preamble starts with 80-bit sychronization (a 010101... bit pattern) used for synchronization of potential receivers and signal detection by the CCA.
- **Start frame delimiter (SFD):** This subfield is 16 bits long and consists of a specific bit string (0000 1100 1011 1101) to indicate the start of the frame and provide frame sychronization and timing.

The PLCP header consists of three separate subfields:

- **PLCP_PDU length word:** This first field of the PLCP header (12 bits long) specifies the size of the MAC frame payload (PSDU) in bytes including the 32 bit CRC at the end of the payload. PLW can range between 0 and 4,095.
- **PLCP signalling field (PSF):** This 4-bit field indicates the data rate of the payload following and takes the values 0000 and 0010 for the data rates of 1 Mbit/s and 2 Mbit/s, respectively.
- **Header error check (HEC):** To protect errors in the PLCP header, a 16-bit CRC is calculated and placed in this field.

Finally, the size of the payload field ranges from 0 to 4095 bytes.

A.4.1.3 Direct Sequence Spread Spectrum (DSSS) PHY

Spread spectrum was first developed by the military as a secure wireless technology. It modulates (changes) a radio signal pseudo-randomly so it is difficult to decode. This modulation provides some security, however, because the signal can be sent great distances, there is interception risk. To provide complete security, most spread spectrum products include encryption.

DSSS is an alternative spread spectrum method separating by code time domain and not by frequency. It works by taking a data stream of zeros and ones and modulating it with a second pattern, the *chipping sequence*. The sequence is also known as the Barker code, which is an 11-bit sequence (10110111000). The chipping or spreading code is used to generate a redundant bit pattern to be transmitted, and the resulting signal appears as wide band noise to the unintended receiver. The DSSS signaling technique divides the 2.4 GHz band into 14 channels of 22 MHz width, of which 11 adjacent channels overlap partially and the remaining three do not overlap. Data is sent across one of these 22 MHz channels without hopping to other channels. To reduce the number of re-transmissions and noise, chipping is used to convert each bit of user data into a series of redundant bit patterns called "chips". A DSSS transmitter transmits the chips to a receiver that reassembles them back into a data stream that is identical to the original. The inherent redundancy of each chip, combined with spreading the signal across the 22 MHz channel, provides the error checking (the receiver can usually identify noise) and correction functionality to recover the data.

As Figure A.14 shows, a DSSS transmission uses more bandwidth but less power than a conventional signal. The digital signal on the left is a conventional transmission, in which the power is concentrated within a tight bandwidth. The DSSS signal on the right uses the same amount of power, but it spreads that power across a wider band of radio frequencies. Obviously, the 22 MHz DSSS channel is a lot wider than the 1 MHz channels used in FHSS systems.

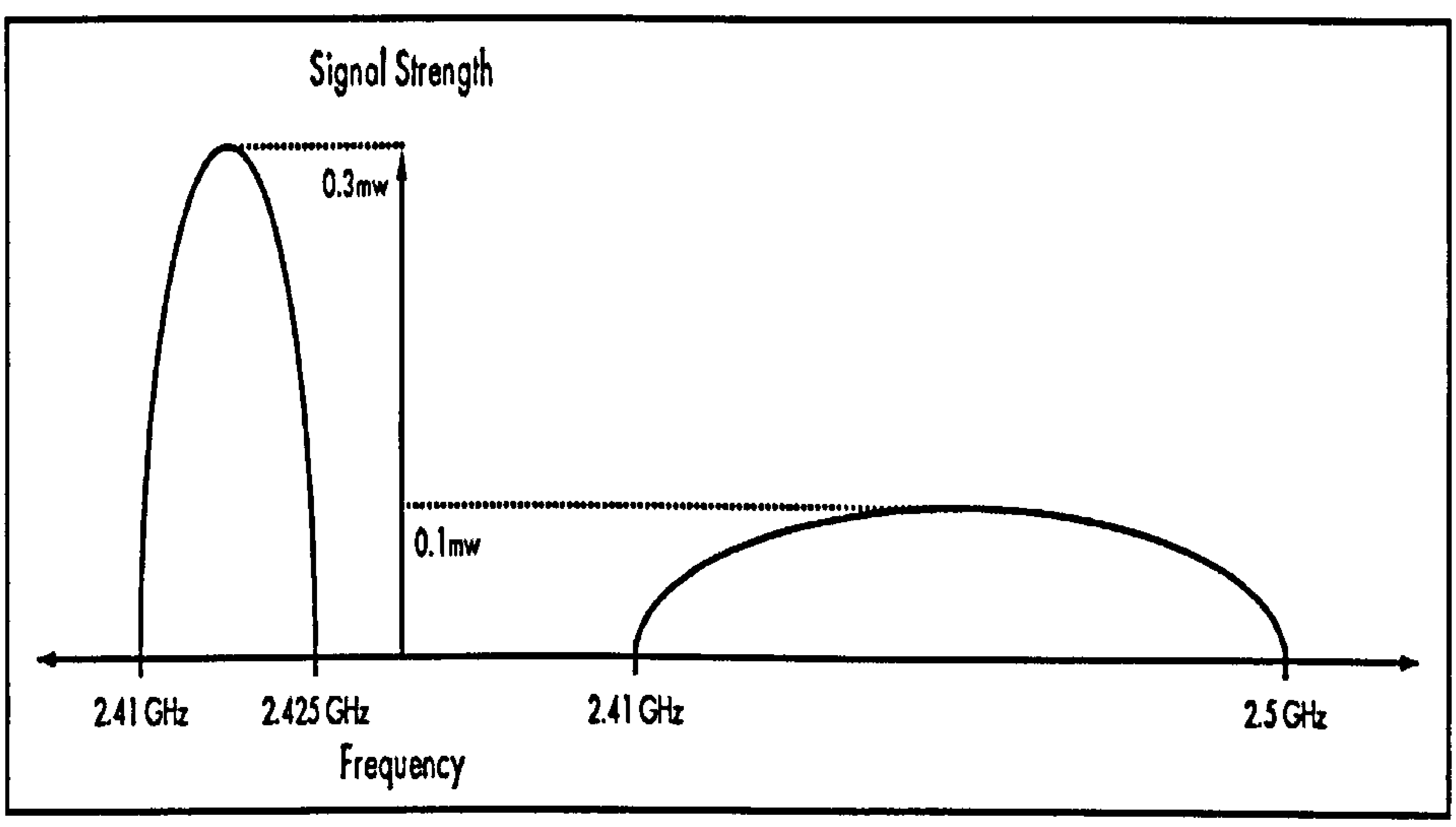


Figure A.14 Conventional and DSSS radio signals

IEEE 802.11 DSSS PHY offers both 1 and 2 Mbit/s data rates. The modulation is being achieved using either differential binary phase shift keying (DBPSK) for 1 Mbit/s transmission or differential quadrature phase shift keying (DQPSK) for 2 Mbit/s transmission. Figure A.15 shows the format of a frame utilizing the DSSS physical layer. The frame consists of two basic parts, the PLCP part (preamble and header) always transmitted at 1 Mbit/s and the payload part.

PLCP Preamble		PLCP Header				
Sync	SFD	SIGNAL	SERVICE	LENGT	CRC	MAC Frame (PSDU)
128 bits	16 bits	8 bits	8 bits	16 bits	16 bits	variable

Figure A.15 Format of a PHY frame using DSSS

The PLCP preamble consists of two separate subfields:

- **Sync:** The Sync subfield is 128 bits long and provides synchronization for the receiving station.
- **Start frame delimiter (SFD):** The SFD subfield is 16 bits long and consists of a specific bit string (1111 0011 1010 0000) to mark the start of every frame and provide frame timing.

The PLCP header is 48 bits long and comprises four separate subfields:

- **SIGNAL:** This 8-bit subfield specifies the modulation and data rate for the frame. In fact, it indicates the data rate of the payload following and takes the values 0000 and 0010 for the data rates of 1 Mbit/s and 2 Mbit/s, respectively.
- **SERVICE:** This 8-bit subfield is reserved for future use, meaning that it was left undefined (but reserved) at the time of the specification was written so that future changes to the standard could use this subfield.
- **LENGTH:** 16 bits are used for length indication of the payload in microseconds (useful for the PHY to correctly detect the end of the packet).
- **CRC:** This 16-bit subfield protects the PLCP header (SIGNAL, SERVICE and LENGTH subfields) with a CRC-16 frame check sequence (FCS).

Finally, the payload can be sent either at 1 or 2 Mbit/s. The size of the payload subfield is adjustable and ranges from 4 to 8191 bytes.

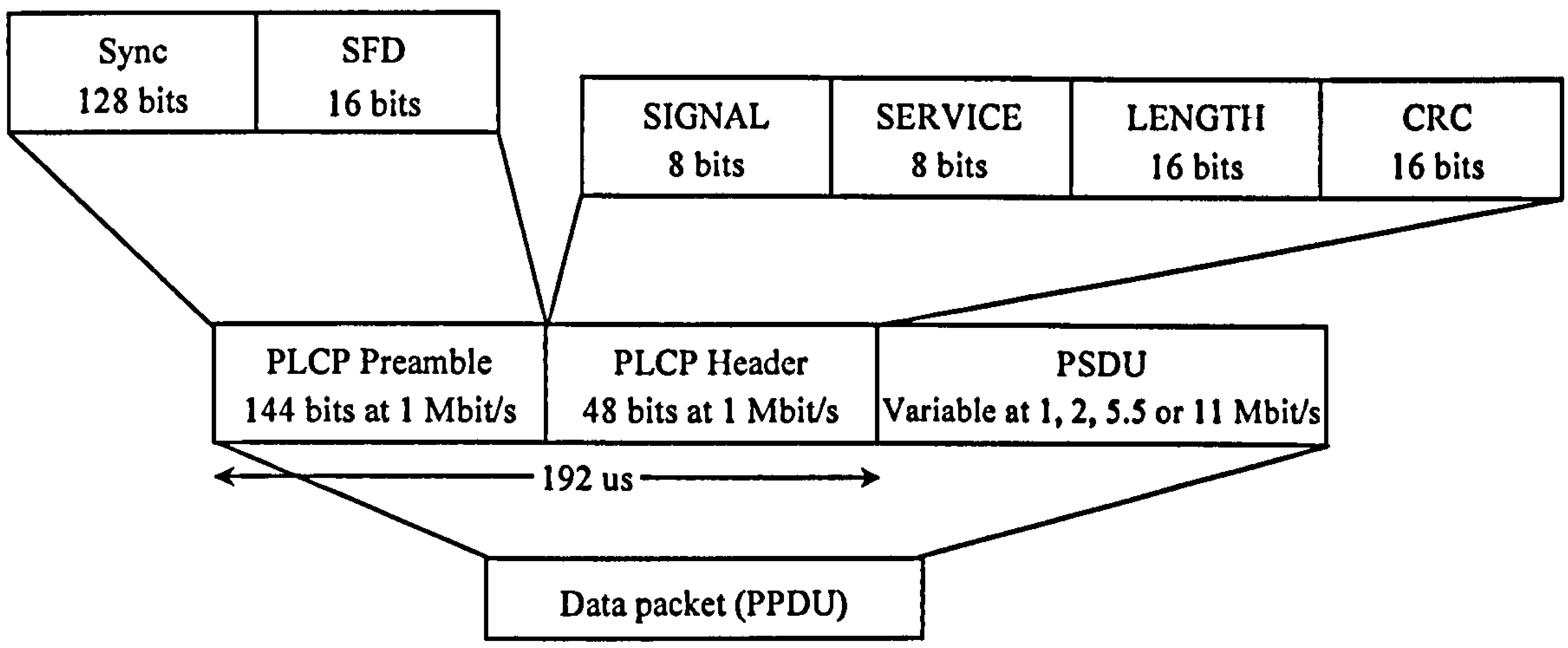
A.4.2 IEEE 802.11b

A.4.2.1 High-Rate DSSS (HR-DSSS) PHY

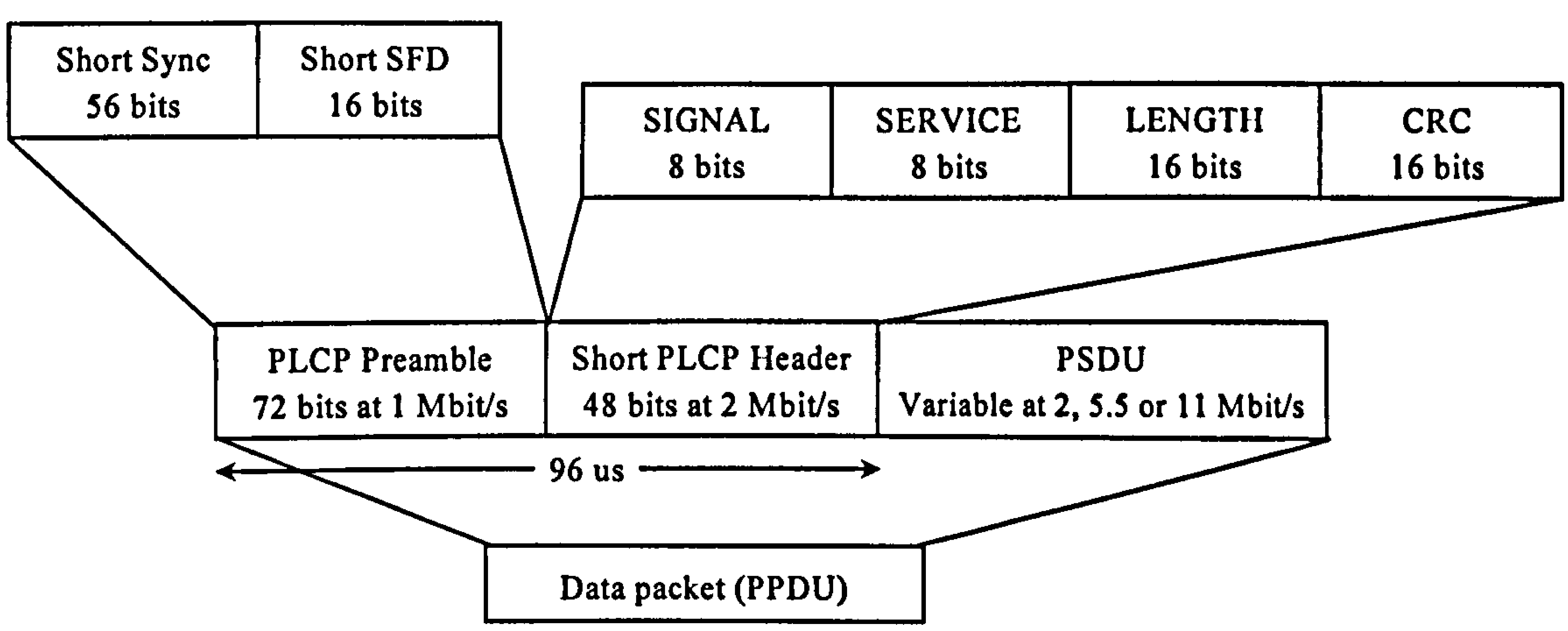
All previously mentioned coding techniques for 802.11 provide a speed of 1 to 2 Mbit/s. The only technique capable of providing higher data rates is DSSS. Extending the DSSS physical layer specification in 1999, the IEEE 802.11b [136] introduced High-Rate DSSS (HR-DSSS) as supplement to the original standard, providing data rates 1 or 2 Mbit/s and the higher data rates of 5.5 and 11 Mbit/s in the 2.4 GHz ISM frequency band[4]. HR-DSSS utilizes the same channelization scheme and chipping rate of 11 MHz as DSSS, thus providing the same 22 MHz bandwidth and 11 channels, 3 non-overlaping. To achieve a higher data rate in the same bandwidth at the same chipping rate, a new modulation technique, the Complementary Code Keying (CCK) is utilized. Rather than the two 11-bit Barker code, CCK uses a set of 64 eight-bit unique code words, thus up to 6 bits can be represented by any code word (instead of the 1 bit represented by a Barker symbol). The 5.5 Mbit/s rate uses CCK to encode 4 bits per carrier, while the 11 Mbit/s rate encodes 8 bits per carrier. Both data rates use the DQPSK modulation technique and encode 2 bits of information in the same space as BPSK encodes 1. The CCK modulation scheme is quite complex and is not explained in detail here.

The High Rate PHY uses the same PLCP preamble and header as the DSSS PHY, so both PHYs can co-exist in the same BSS. In addition to providing higher speed extensions to the DSSS system, there are a number of optional features that allow the performance to be improved. An optional mode that allows data throughput at the higher rates (2, 5.5, and 11 Mbit/s) to be significantly increased is the use of a Short PLCP preamble called HR/DSSS/short (in contrast with the Long Preamble in HR/DSSS). This Short Preamble mode can coexist with DSSS, HR/DSSS, or HR/DSSS/short under limited circumstances, such as on different channels or with appropriate CCA mechanisms. Figure A.16 illustrates the two packet formats of the PPDU used in the HR/DSSS and HR/DSSS/short PHY.

[4] This physical layer extension is backward compatible with legacy DSSS 802.11 systems.



(a) *HR/DSSS*



(b) *HR/DSSS/short*

Figure A.16 Format of a PHY frame used in 802.11b

Furthermore, in order to support very noisy environments as well as extended range, 802.11b wireless LANs are also capable of "dynamic rate shifting", which allows data rates to be automatically adjusted to compensate for the varying nature of the radio channel. This feature can be useful when there's a source of electrical noise near the receiver or when the transmitter and receiver are too far apart to support full-speed operation. Initially, the equipment tries to connect at the full 11 Mbit/s rate. If the devices move beyond the optimal range for 11 Mbit/s operation, or if considerable interference is encountered, then the 802.11b devices will "fall back" and transmit at the lower speed of 5.5 Mbit/s. If 5.5 Mbit/s is still too fast for the link to handle, it drops again, down to 2 Mbit/s, or even 1 Mbit/s. Likewise, if the device moves closer or if the interference disappears, then the connection will automatically increase to 11 Mbit/s. Rate shifting is a Physical Layer mechanism that's transparent to the user and the upper layers of the OSI protocol stack.

A.4.3 IEEE 802.11a

A.4.3.1 Orthogonal Frequency Division Multiplexing (OFDM) PHY

IEEE 802.11a [137] was developed in response to a demand for high data rate WLANs. Unlike IEEE 802.11b, IEEE 802.11a utilizes Orthogonal Frequency Division Multiplexing (OFDM) modulation to spread the transmitted signal over a wide bandwidth and was designed to operate in the 5 GHz UNII (Unlicensed National Information Infrastructure) frequency radio band[5]. Unlike ISM band, which offers about 83 MHz spectrum in the 2.4 GHz frequency for 802.11b devices, IEEE 802.11a utilizes almost four times that of the ISM band. In fact, at 5 GHz the UNII band offers 300 MHz of relatively free of interference spectrum. This is achieved by segmenting the UNII band into three different 100 MHz bands for operation in the USA, each with a different legal maximum power output; the lower band ranges from 5.15 –5.25 GHz (50 mW), the middle band ranges from 5.25-5.35 GHz (250 mW) and the upper band ranges from 5.725-5.825 GHz (1 W)[6]. Within this spectrum, the lower and middle bands accommodate 8 channels in a total bandwidth of 200 MHz (intended for in-building applications) and the upper band accommodates 4 channels in a 100 MHz bandwidth (for outdoor use, i.e building-to-building). The frequency channel center frequencies are spaced 20 MHz apart. The outermost channels of the lower and middle bands are centered 30 MHz from the outer edges. In the upper band the outermost channel centers are 20 MHz from the outer edges.

Depending on the modulation scheme employed, the 802.11a PHY layer can support data rates from 6 Mbit/s up to 54 Mbit/s. In fact, OFDM technique provides communication capabilities of 6, 9, 12, 18, 24, 36, 48, or 54 Mbit/s with the support of

---

[5] These two characteristics make 802.11a networks incompatible with 802.11b networks.

[6] The European Telecommunications Standards Institute (ETSI) defines different frequency bands for Europe: 5.15-5.35 GHz and 5.47-5.725 GHz. Japan allows operation in the frequency range 5.15-5.25 GHz. Up to now, only 100 MHz are available "worldwide" at 5.15-5.25 GHz.

transmitting and receiving at data rates of 6, 12, and 24 Mbit/s being mandatory.

The basic principle of OFDM technique is to divide a high-speed binary carrier generated by the sender into several lower-speed sub-carriers, which are then transmitted in parallel to each other (figure A.17). Each 20 MHz wide high-speed carrier is broken up into 52 OFDM sub-carriers, each approximately 300 KHz wide (shown in detail in figure A.17). OFDM uses 48 of these sub-carriers for actual data, while the remaining 4 pilot sub-carriers are used for error correction. Each of the sub-carriers is spaced 0.3125 MHz apart (for a 20 MHz-wide channel with 64 possible sub-carrier frequency slots) and a guard time[7] of 800 ns is added to each symbol making the total symbol duration 4 μs.

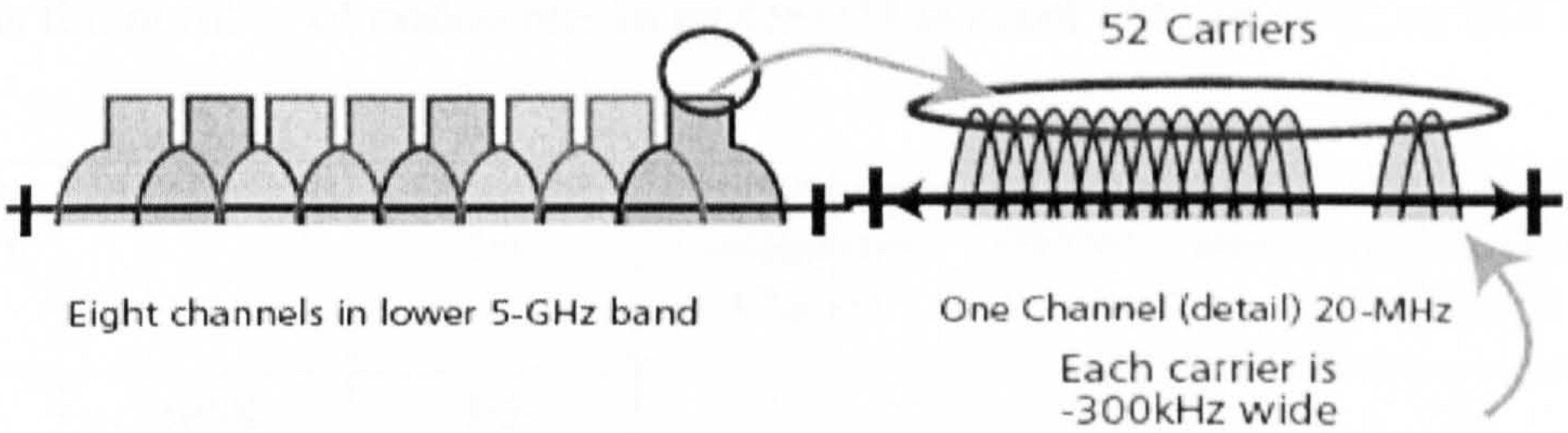


Figure A.17 The OFDM technique used in IEEE 802.11a

A key feature of the IEEE 802.11a PHY is to provide eight data rates (called PHY modes) with different modulation schemes and coding rates. Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), 16 Quadrature Amplitude Modulation (16-QAM) and 64 Quadrature Amplitude Modulation (64-QAM) are the supported modulation schemes. These modulation schemes are coupled with the various forward error correction convolutional encoding scheme (with a coding rate of 1/2, 2/3, or 3/4) to give a multitude of number of the data bits per symbol (N_{DBPS}) performance.

There are four rate tiers with the OFDM PHY: 6 and 9 Mbit/s, 12 and 18 Mbit/s, 24 and 36 Mbit/s, and 48 and 54 Mbit/s. Support is required for 6, 12, and 24 Mbit/s, which are lowest speeds in each of the first three tiers, and therefore the most robust in the presence of interference. The lowest tier uses BPSK to encode 1 bit per subchannel, or 48 bits per symbol. The convolution coding means that either half or one quarter of the bits are redundant bits used for error correction, so there are only 24 or 36 data bits per

---

[7] With so much information per transmission, it obviously becomes important to guard against data loss. Forward Error Correction (FEC) was added to 802.11a for this purpose (FEC does not exist in 802.11b).

symbol. The next tier uses QPSK to encode 2 bits per subchannel, for a total of 96 bits per symbol. After subtracting overhead from the convolution code, the receiver is left with 48 or 72 data bits. The third and fourth tiers use generalized forms of BPSK and QPSK known as QAM. 16-QAM encodes 4 bits using 16 symbols and 64-QAM encodes 6 bits using 64 symbols. The third tier uses 16-QAM along with the standard r=1/2 and r=3/4 convolution codes, achieving data rates of 24 and 36 Mbit/s, respectively. To achieve higher rates (48 and 54 Mbit/s with 64-QAM, however, the convolution codes use r=2/3 and r=3/4. Table A.4 shows how each supported data rate is mapped to the appropriate OFDM PHY parameters where N_{BPSC} is the coded bits per subchannel being a function of the modulation (BPSK, QPSK, 16-QAM, or 64-QAM) and N_{CBPS} is the number of coded bits in an OFDM symbol (48, 96, 192, or 288 bits).

Data rate (Mbit/s)	Modulation	Coding rate (r)	Coded bits per subcarrier (N_{BPSC})	Coded bits per OFDM symbol (N_{CBPS})	Data bits per OFDM symbol (N_{DBPS})
6	BPSK	1/2	1	48	24
9	BPSK	3/4	1	48	36
12	QPSK	1/2	2	96	48
18	QPSK	3/4	2	96	72
24	16 QAM	1/2	4	192	96
36	16 QAM	3/4	4	192	144
48	64 QAM	2/3	6	288	192
54	64 QAM	3/4	6	288	216

Table A.4 Rate-dependent parameters of 802.11a

Due to the nature of OFDM, the PDU on the physical layer of IEEE 802.11a looks quite different from 802.11b or the original 802.11 physical layers. The basic structure of an IEEE 802.11a PLCP Frame format is shown in figure A.18 and it includes a PLCP Preamble, a SIGNAL field, and a DATA field:

- **PLCP preamble**: This field is used to acquire the incoming OFDM signal and train and synchronize the demodulator. The PLCP preamble consists of 12 symbols; it begins with 10 short training symbols of 0.8 μs followed by two long training symbols of 4.0 μs each. The short symbols are used to train the receiver's automatic gain control (AGC) and to estimate a coarse estimate of the carrier frequency and the channel. The long symbols are used to fine-tune the frequency and the channel estimates. Thus, the training of an OFDM receiver is accomplished in 16 μs.

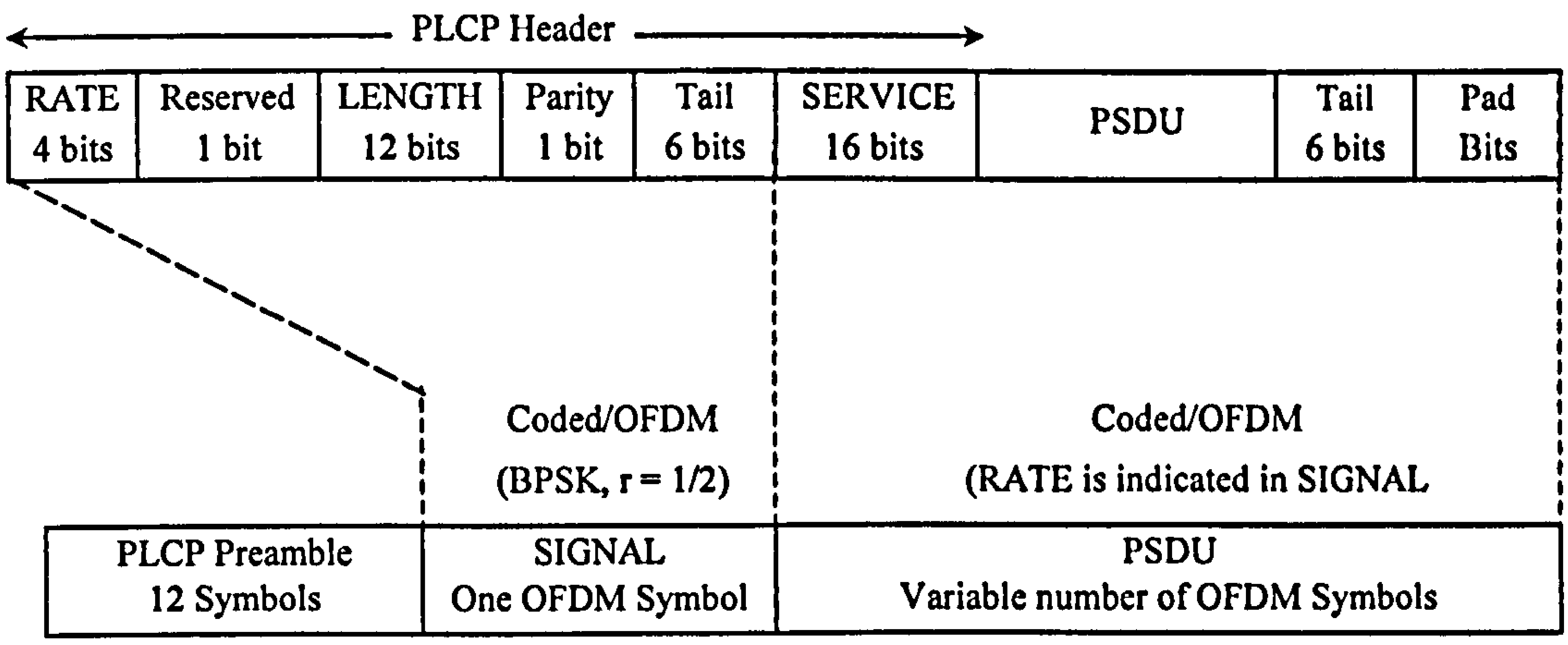


Figure A.18 PPDU frame format of the IEEE 802.11a OFDM PHY

- **SIGNAL**: This 24-bit field contains information about the rate and length of the PSDU and it is composed of a Rate subfield, a reserved bit, a Length subfield, a Parity bit, and a Tail, all combined to form one OFDM symbol. The SIGNAL field is always transmitted utilizing the most robust combination of BPSK modulation (at 6 Mbit/s) and a coding rate of r = 1/2. The 4-bit (R1-R4) Rate subfield is used to encode the rate. The mapping of these bits to the data rate is given in table A.5. The next bit is reserved for future use. The Length subfield (12 bits long) identifies the number of octets in the PSDU. A continuation is an even parity bit (for the previous 17 bits of the SIGNAL field) and finally the last subfield contains 6 tail bits. The six "zero" tail bits are used to return the convolutional codec to the "zero state" (making possible to decode the Rate and Length fields immediately following receipt of the Tail bits).

Rate (Mbit/s)	R1–R4
6	1101
9	1111
12	0101
18	0111
24	1001
36	1011
48	0001
54	0011

Table A.5 Rate subfield mapping

- **PSDU (DATA)**: The SERVICE subfield and the PSDU (with 6 "zero" tail bits and pad bits appended), denoted as DATA, are transmitted at the data rate described in the RATE subfield and may constitute multiple OFDM symbols (each data symbol is 4.0 μs long).

As figure A.18 shows, the PLCP Header is defined as the field used to generate the SIGNAL field plus the 16-bit SERVICE field. In fact, the PLCP Header, expect the SERVICE field, constitutes a single OFDM symbol. The pad bits are used to make the resulting bit string into a multiple of OFDM symbols. Both the RATE and LENGTH are required for decoding the DATA part of the packet. In addition, the Clear Channel Assessment (CCA) mechanism is used to predict the duration of the packet from the contents of the rate and length fields. The encoding process is quite complex (composed of many steps) and can be found in [137].

A.4.4 IEEE 802.11b and 802.11a pros and cons

When deploying a WLAN, there is a number of considerations to be considered when deciding on whether to deploy 2.4 GHz (802.11b) or 5 GHz (802.11a) solutions. The move to the 5 GHz band and OFDM modulation provides two important benefits of 802.11a over 802.11b. First, it increases the maximum speed per channel from 11 Mbit/s to 54 Mbit/s. This is a tremendous boost, especially useful for wireless multimedia, large file transfers, and fast Internet access. Second, the bandwidth available in the 5-GHz range is larger than available at 2.4 GHz, allowing for more simultaneous users without potential conflicts. Additionally, the 5-GHz band is not as congested at the 2.4-GHz band, resulting in less interference. 2.4GHz WLANs can experience interference from cordless phones, microwaves, and other WLANs.

These advantages come with some downsides. The higher operating frequency equates to a shorter range. This means that to maintain the high data rates, a larger number of 802.11a access points are required to cover the same area (increased cost), versus 802.11b. While 802.11b access points have a typical range of 100 meters, 802.11a access points are often limited to between 25 and 50 meters. In addition, OFDM requires more power than DSSS, leading to higher power consumption by 802.11a products. This is definitely a disadvantage for mobile devices that have limited battery power. Another downside is that 802.11a and 802.11b products are not compatible. With the large number of 802.11b products on the market, this has a negative effect on the adoption of 802.11a products.

Another very important issue in wireless communications is the frequency allocation. As the number of wireless devices dramatically increases, it is not difficult to see why there is a need for regulatory agencies like International Telecommunication Union (ITU) and the Federal Communications Commission (FCC) that are responsible for the development and operation of wireless systems. By international agreement, a window of the radio spectrum near 2.4 GHz is supposed to be reserved for unlicensed industrial, scientific, and medical (ISM) services, including spread-spectrum wireless data networks. However, the exact frequency allocations are slightly different from one part of the world to another; the authorities in different countries have assigned slightly different frequency bands. Thus, if implementing a spread spectrum system it is important to avoid overlapping between adjacent channels by investigating the local regulations. On the other hand, the use of 5GHz for WLANs is somewhat limited. For example, the U.S. allows operation of 5GHz WLANs, but other countries (e.g., China) do not.

A.5 Point Coordination Function (PCF)

Point Coordination Function (PCF) is used to implement time-critical services, like voice or video transmission and it is defined for use only in an infrastructure-based network. PCF is optional and it is based on polling the stations to access the channel and, therefore, it is contention-free. In PCF, a single AP controls access to the media and a point coordinator resides in the AP. If a BSS is set up with PCF enabled, time is sliced between the system being in PCF mode and in DCF mode. Details about PCF can be found in [135], nevertheless we briefly cite the main features of this function.

Figure A.19 shows the PCF operational mode; the AP starts the contention-free period (CFP) periodically by transmitting a beacon frame, which updates the NAVs of the stations with the maximum expected CFP time. After sending the beacon, the AP starts polling each station for data, and after a given time move on to the next station, according to a polling list. Polls and ACKs can be piggybacked to data frames so bandwidth is efficiently used. As with SIFS and DIFS, PCF gets priority over DCF by waiting the channel being idle for a Polling IFS (PIFS) before it grabs the channel. PIFS is shorter than DIFS, giving the AP absolute priority to transmit before any of the stations try to contend using DCF. When the CFP ends, DCF mode starts being used by stations randomly contending to access the medium each time they detect the medium idle for a period longer than DIFS, as described before.

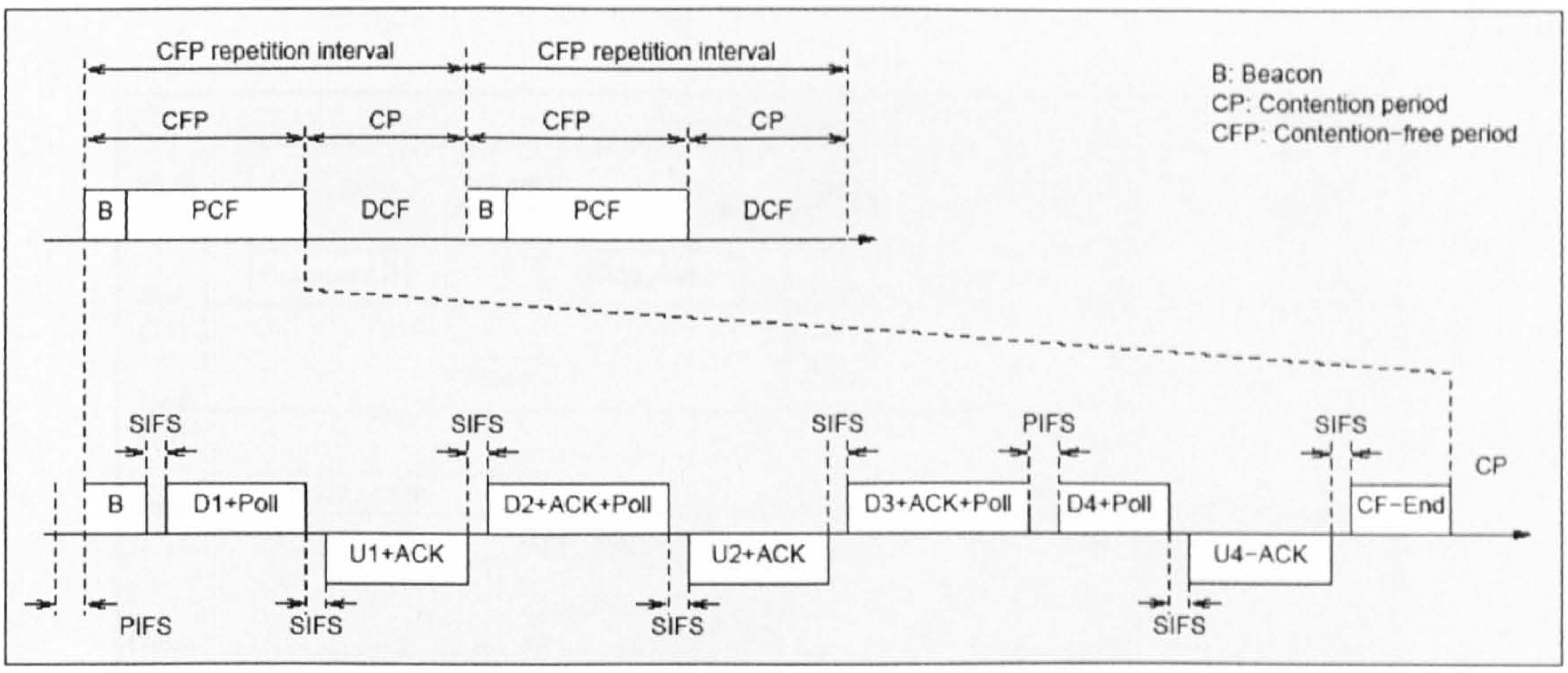


Figure A.19 Polling Coordination Function (PCF)

When using the PCF, only station-AP or AP-station transmissions are possible. Therefore, a communication between two stations in the same BSS has to go through the AP, wasting bandwidth. Another limitation of PCF is it is not particularly scalable due to the fact a single AP needs to have control of media access and must poll all stations, which can be ineffective in large networks. However, one of the advantages of this mechanism (apart from providing a guaranteed maximum latency) is that the AP provides a good power saving capability; the AP can store incoming packets in a buffer allowing the destination station to stay in sleep mode during relatively long periods, in order to save battery power.

A.6 MAC Packet Fragmentation

The MAC layer can then optionally break the original data packet into smaller fragments for sequential individual transmission; this is called fragmentation. Very large frames may reduce transmission reliability too, e.g. a transmission error in a large packet wastes more bandwidth and transmission time than an error in a shorter packet. An optimization parameter is used, the fragmentation threshold, above which packets are fragmented as shown in Fig. 3.6. Packet fragments are transmitted on the channel separated by SIFS, so no new packet can interrupt the current transmission. Each fragment is acknowledged separately, else, the fragment is retransmitted before any other fragments and keeping the sequence in order. If any fragment of a packet encounters any errors or collision, only the fragment needs to be retransmitted, not the entire packet, enhancing the performance of the medium.

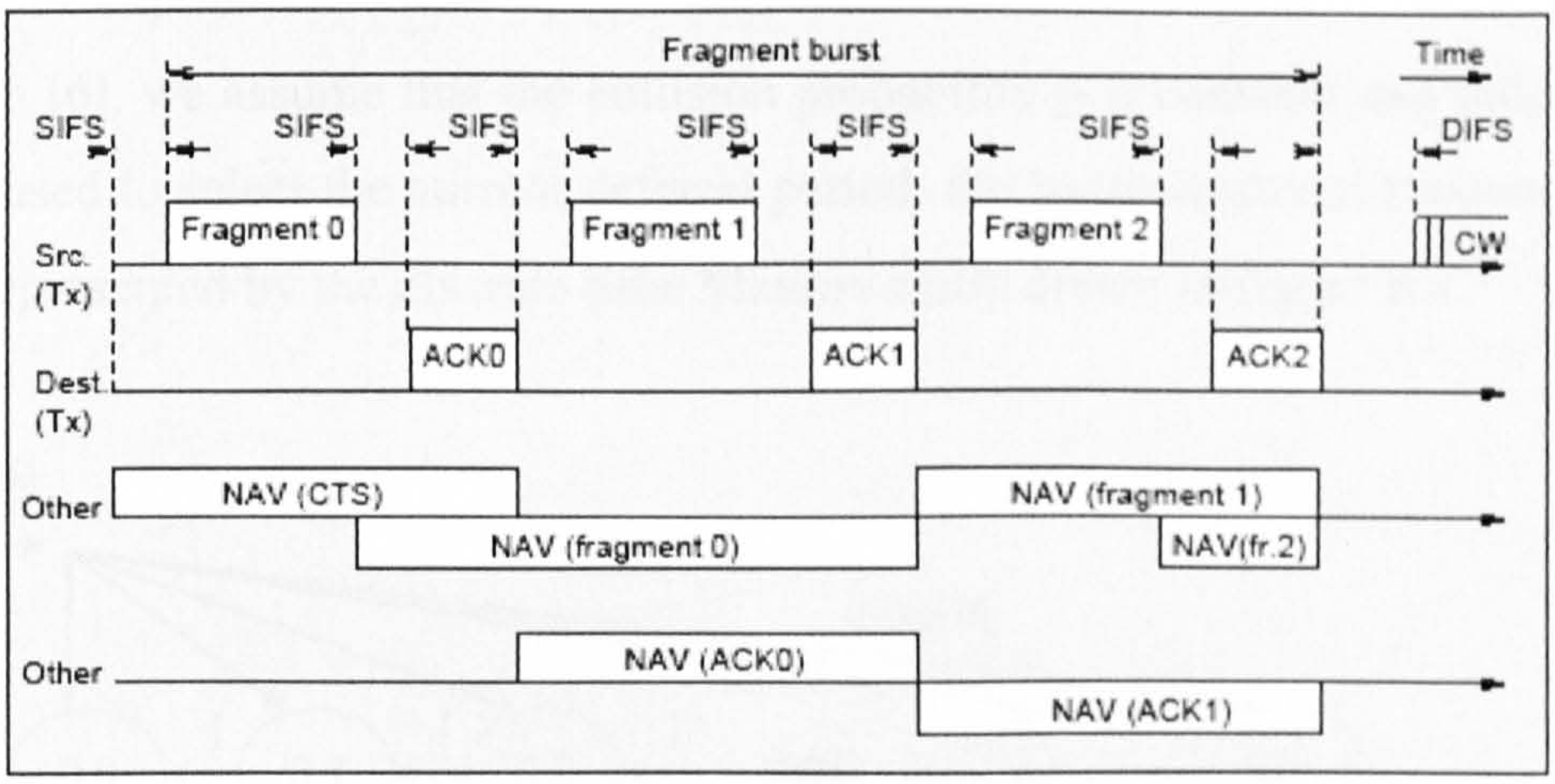


Figure A.20: Packet fragmentation

Although fragmentation can increase the reliability of packet transmission (especially against noise due to signal fading and interference), this comes at a cost of extra-added MAC overhead. At high data rates and good channel quality, fragmentation has an adverse effect on performance. Therefore, fragmentation must balance between medium reliability and medium overhead. The IEEE 802.11 standard mandates that all receivers are able to support fragmentation, but fragmentation support is optional for transmitters. In this thesis, typical MAC packets are considered smaller than the default fragmentation threshold (equal to 2346 bytes).

B.1 Bianchi analysis

This section presents the derivation of throughput efficiency, average packet delay and packet inter arrival utilizing Bianchi's approach that does not consider packet retry limits. As mentioned before, Bianchi models the idealistic assumption that packet retransmissions are unlimited, thus, a packet is being retransmitted continuously until its successful reception.

As in [6], we assume that the collision probability p is constant and independent of the CW used to select the current deferral period; the bi-dimensional process $\{s(t), b(t)\}$ can be represented by the discrete-time Markov chain drawn in figure B.1.

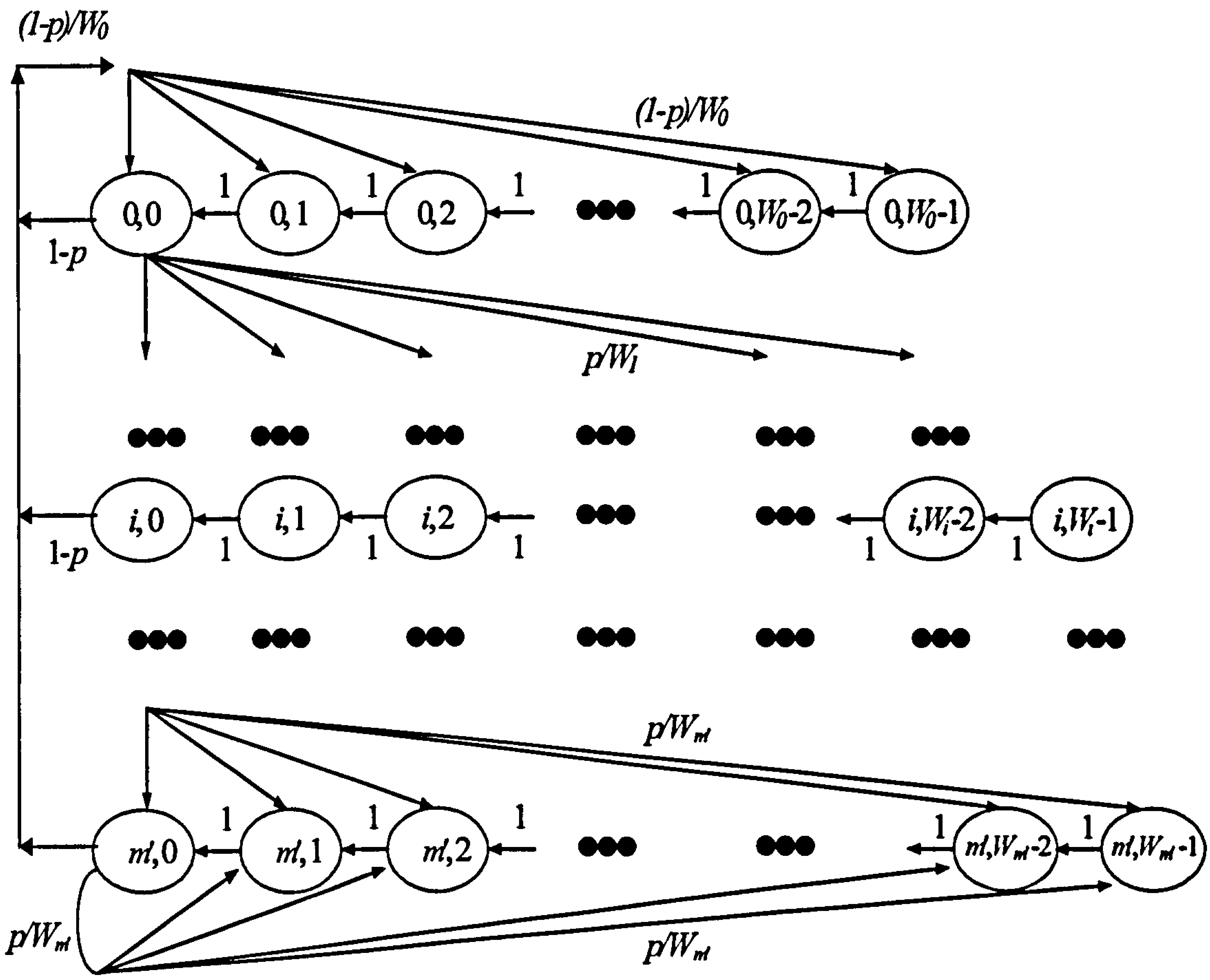


Figure B.1 Markov chain model

Prior to initiating a packet transmission, the value of the station's backoff timer is uniformly chosen in the range $[0, W_i\text{-}1]$, where W_i is the current CW size and $i \in [0, m']$ is the backoff stage. We have $W_i = 2^i \cdot W$, where $W = CW_{min} = W_0$ is the minimum CW size and m' is the maximum backoff stage such that $CW_{max} = 2^{m'} \cdot W$.

The only non-null one-step transition probabilities of the chain are (using the short notation $P\{i_1,k_1 \mid i_0,k_0\} = P\{s(t+1)=i_1,b(t+1)=k_1 \mid s(t)=i_0,b(t)=k_0\}$):

$$\begin{cases} P\{i,k \mid i,k+1\} = 1 & k \in (0,W_i-2) \quad i \in (0,m') \\ P\{0,k \mid i,0\} = (1-p)/W_0 & k \in (0,W_0-1) \quad i \in (0,m') \\ P\{i+1,k \mid i,0\} = p/W_{i+1} & k \in (0,W_{i+1}-1) \quad i \in (0,m'-1) \\ P\{m',k \mid m',0\} = p/W_{m'} & k \in (0,W_{m'}-1) \end{cases} \tag{B.1}$$

The first equation in (B.1) accounts for the fact that the backoff time counter is decreased at the beginning of each time slot until it reaches zero. The second equation represents the fact that a new packet following a successful packet transmission starts with backoff stage 0 and, thus, the backoff is initially uniformly chosen in the range $[0,W_0\text{-}1]$. The third and fourth equations consider the case of an unsuccessful packet transmission. In particular, the third equation models the case when an unsuccessful transmission occurs at backoff stage i; the backoff stage increases and the new initial backoff value is uniformly chosen in the range $[0,W_{i+1}\text{-}1]$. Finally, the fourth case models the fact that once the maximum backoff stage m' is reached, the backoff stage is not increased further in subsequent packet collisions.

We define that $b_{i,k} = \lim_{t\to\infty} P\{s(t)=i,b(t)=k\}$, $i \in (0,m')$, $k \in (0,W_i-1)$ represents the stationary distribution of the chain. First note that:

$$b_{i,0} = p \cdot b_{i-1,0} = p^i \cdot b_{0,0} \qquad 0 < i < m' \tag{B.2}$$

$$b_{m',0} = \frac{p^{m'}}{1-p} b_{0,0} \tag{B.3}$$

Owing to chain regularities, for each $k \in (0,W_i-1)$,

$$b_{0,k} = \frac{W_0-k}{W_0}\left((1-p)\sum_{j=0}^{m'} b_{j,0}\right), \qquad i=0$$

$$b_{ik} = \frac{W_i-k}{W_i}\left(p \cdot b_{i-1,0}\right), \quad 0 < i < m' \tag{B.4}$$

$$b_{m',k} = \frac{W_{m'}-k}{W_{m'}}\left(p\,b_{m'-1,0} + p\,b_{m',0}\right), \quad i = m'$$

After some algebra, we have:

$$b_{0k} = \frac{W_0 - k}{W_0} b_{00}$$

$$b_{ik} = \frac{W_i - k}{W_{m'}} b_{i0} \tag{B.5}$$

$$b_{m'k} = \frac{W_{m'} - k}{W_{m'}} b_{m'0}$$

Equations (B.4) can be rewritten as:

$$b_{i,k} = \frac{W_i - k}{W_i} b_{i,0}, \; i \in [0, m'], k \in [0, W_i - 1] \tag{B.6}$$

As a result, equations (B.2), (B.4) and (B.6) express all $b_{i,k}$ values as a function of $b_{0,0}$ and conditional collision probability p. By applying the normalization factor:

$$1 = \sum_{i=0}^{m'} \sum_{k=0}^{W_i - 1} b_{i,k} = \sum_{i=0}^{m'} b_{i,0} \sum_{k=0}^{W_i - 1} \frac{W_i - k}{W_i} = \sum_{i=0}^{m'} b_{i,0} \frac{W_i + 1}{2} = \frac{b_{0,0}}{2} \left(\sum_{i=0}^{m'} p^i \cdot W_i + \sum_{i=0}^{m'} p^i \right)$$

and by substituting W_i, we have:

$$1 = \frac{b_{0,0}}{2} \left[W \left(\sum_{i=0}^{m'-1} (2p)^i + \frac{(2p)^{m'}}{1-p} \right) + \frac{1}{1-p} \right] = \frac{b_{0,0}}{2} \left[W \left(\frac{1-(2p)^{m'+1}}{1-(2p)} + \frac{(2p)^{m'}}{1-p} \right) + \frac{1}{1-p} \right] \tag{B.7}$$

which can solved to[8]:

$$b_{0,0} = \frac{2 \cdot (1-2p) \cdot (1-p)}{(W+1)(1-2p) + pW(1-(2p)^{m'})} \tag{B.8}$$

Equation (B.8) expresses $b_{0,0}$ as a function of the collision probability p, the initial contention window size W and the number of backoff stages m'. This analysis allows us to evaluate the station's transmission probability τ. Considering that a station transmits when the backoff timer reaches zero:

$$\tau = \sum_{i=0}^{m'} b_{i,0} = b_{0,0} \cdot \frac{1}{(1-p)} \tag{B.9}$$

---

[8] Note that the term $2p$ that occurs in equation (B.8) is related to the fact that the contention window is doubled (increased by the factor 2) upon unsuccessful transmission attempts.

Substituting the value of $b_{0,0}$ from equation (B.8) to equation (B.9), τ is given by:

$$\tau(p):\tau=\frac{2(1-2p)}{(W+1)(1-2p)+pW(1-(2p)^{m'})} \qquad \text{(B.10)}$$

Equation (B.10) expresses probability τ as a function of the probability p that it is still unknown. Assuming that the number of transmitting stations n is constant and that all contending stations 'see' the discrete-time Markov chain drawn in figure B.1 at the steady state and transmit with probability τ, the transmission collision probability p can be expressed as the probability that at least an extra one of the remaining (n-1) stations transmit at the chosen slot time:

$$p=1-(1-\tau)^{n-1} \qquad \text{(B.11)}$$

As expected, in the case of only one station i.e. n=1, the collision probability is p=0. Equations (B.10) and (B.11) form a non-linear system with the unknowns τ and p, which can be solved by employing numerical methods. It is easy to demonstrate that there is only one solution to the non-linear system. Finally, the probabilities τ and p are evaluated for a certain W, m' and n combination.

A. Saturation throughput

Following the same reasoning with Chapter 3, the saturation throughput S can be expressed by dividing the successfully transmitted payload information in a time slot, with the average length of a time slot:

$$S=\frac{P_{tr}P_S l}{E[slot]}=\frac{P_{tr}P_S l}{(1-P_{tr})\sigma+P_{tr}P_S T_S+P_{tr}(1-P_S)T_C} \qquad \text{(B.12)}$$

This expression is exactly the same with the one given in section 3.2 (equation 3.14) for the case of finite retry limits. However, as been shown previously, the two models (with or without packet retry limits) reach different equations for τ due to the fact that the Markov chain transitions are different.

B. Packet inter arrival time

The packet inter arrival time is defined as the time interval between two successful packet receptions at the receiver. For both cases of infinite and finite retry limits, the average packet inter arrival time $E[D_{inter}]$ can be simply obtained from throughput:

$$E[D_{inter}] = \frac{n}{S/l} \tag{B.13}$$

C. Packet delay

Next, we calculate the average delay $E^B[D]$ for a successfully transmitted packet utilizing a similar approach with the one in section 3.2 but for the case of no retry limits. We have:

$$E^B[D] = E^B[X]\ E[slot] \tag{B.14}$$

where $E^B[X]$ is the average number of time slots needed for a successful transmission. $E^B[X]$ is calculated by multiplying the number of time slots d_i the packet is delayed in each backoff stage by the probability q_i for the packet to arrive at this backoff stage:

$$E^B[X] = \sum_{i=0}^{m'} d_i\, q_i \tag{B.15}$$

where d_i is given by:

$$d_i = \frac{W_i + 1}{2} \quad ,\ i \in [0, m'] \tag{B.16}$$

and q_i depends on the consideration or not of retry limits. For the case of infinite retry limits, the probability q_i is calculated as:

$$q_i^B = \begin{cases} p^i & ,\ i \in [0, m'-1] \\ \dfrac{p^{m'}}{1-p} & ,\ i = m' \end{cases} \tag{B.17}$$

After some algebra, $E^B[X]$ is given by:

$$E^B[X] = \frac{(1-2p)(W+1) + pW(1-(2p)^{m'})}{2(1-2p)(1-p)} = \frac{1}{\tau(1-p)} \tag{B.18}$$

If we substitute E^B [X] from (B.18) and $E[slot]$ into equation (B.14), the average packet delay $E^B[D]$ can be calculated and as expected, in the case of no retry limits, the packet delay coincides with the packet inter arrival time $E[D_{inter}]$ (obtained directly from throughput in (B.12)):

$$E^B[D] = \frac{E[slot]}{\tau(1-p)} = \frac{n}{S/l} = E[D_{inter}] \tag{B.19}$$

C.1 (proof for Wu's analysis)

Let $b_{i,k} = \lim_{t\to\infty} P\{s(t)=i, b(t)=k\}$ be the stationary distribution of this Markov chain, where $i \in [0,m]$, $k \in [0, W_i - 1]$. Based on the two-dimensional Markov chain illustrated in figure 4.1 and by considering that $b_{1,0} = p \cdot b_{0,0}$ and $b_{2,0} = p \cdot b_{1,0} = p^2 \cdot b_{0,0}$, we have the following relation for $b_{i,0}$:

$$b_{i,0} = p\, b_{i-1,0} = p^i b_{0,0} \qquad 0 < i \le m \tag{C.1}$$

Owing to chain regularities and by means of equation (C.1), all $b_{i,k}$ values are expressed as a function of $b_{0,0}$ and p as:

$$b_{i,k} = \frac{W_i - k}{W_i} \cdot b_{i,0} \qquad 0 \le i \le m \quad , \quad 0 \le k \le W_i - 1 \tag{C.2}$$

Applying the normalization condition for this stationary distribution:

$$1 = \sum_{i=0}^{m} \sum_{k=0}^{W_i-1} b_{i,k} = \sum_{i=0}^{m} b_{i,0} \cdot \sum_{k=0}^{W_i-1} \frac{W_i - k}{W_i}$$

$$= \sum_{i=0}^{m} b_{i,0} \cdot \frac{W_i + 1}{2} = \sum_{i=0}^{m} p^i \cdot b_{0,0} \cdot \frac{W_i + 1}{2}$$

$$= \frac{b_{0,0}}{2} \cdot \left(\sum_{i=0}^{m} p^i \cdot W_i + \sum_{i=0}^{m} p^i \right)$$

from which:

$$b_{0,0} = \frac{2}{\left(\sum_{i=0}^{m} p^i \cdot W_i + \sum_{i=0}^{m} p^i \right)} \tag{C.3}$$

By utilizing the Markov chain model, the probability τ that a station transmits a packet in a randomly chosen slot time is equal to:

$$\tau = \sum_{i=0}^{m} b_{i,0} = \sum_{i=0}^{m} p^i \cdot b_{0,0} = b_{0,0} \cdot \frac{1 - p^{m+1}}{(1-p)} \tag{C.4}$$

and $b_{0,0}$ can be acquired from equation (C.3). From equation (C.4) we observe that the transmission probability τ depends on the conditional probability p, which is defined as the probability that a transmitted packet collides and is given by:

$$p = 1 - (1 - \tau)^{n-1} \tag{C.5}$$

As we stated before, equations (C.4) and (C.5) represent a non-linear system with two unknowns τ and p. This non-linear system, which has a unique solution, can be solved utilizing numerical methods evaluating τ and p for a certain W, m and m' combination. Since the system of the two equations is different from the one in [6], a detailed proof of the uniqueness of this solution is derived next.

Equation (C.5) can be rewritten as:

$$\tau^{*}(p): \tau = 1-(1-p)^{1/(n-1)} \tag{C.6}$$

The function $\tau^{*}(p)$ is a continuous and monotone increasing function in the range $p \in (0,1)$. It increases from $\tau^{*}(0)=0$ to $\tau^{*}(1)=1$. The function $\tau(p)$ given by equation (C.4) is also continuous in the same range[9]; continuity in correspondence of the critical value p=1/2 is simply proven by using equation (C.3) as follows:

$$b_{0,0} = \frac{2}{\sum_{i=0}^{m}\left(\frac{1}{2}\right)^{i} W_i + \sum_{i=0}^{m}\left(\frac{1}{2}\right)^{i}}$$

$$b_{0,0} = \frac{2}{\left(\sum_{i=0}^{m'}\left(\frac{1}{2}\right)^{i}\left(2^{i}W\right) + \sum_{i=m'+1}^{m}\left(\frac{1}{2}\right)^{i}\left(2^{m'}\cdot W\right) + \frac{1-\left(\frac{1}{2}\right)^{m+1}}{1-\frac{1}{2}}\right)}$$

$$b_{0,0} = \frac{2}{\left(\sum_{i=0}^{m'} W + \left(2^{m'}\cdot W\right)\sum_{i=m'+1}^{m}\left(\frac{1}{2}\right)^{i} + \frac{1-\left(\frac{1}{2}\right)^{m+1}}{\frac{1}{2}}\right)}$$

$$b_{0,0} = \frac{2}{\left(W\left(m'+1\right) + \left(2^{m'}\cdot W\right)\frac{1-\left(\frac{1}{2}\right)^{m-m'}}{1-\frac{1}{2}}\left(\frac{1}{2}\right)^{m'+1} + \frac{1-\left(\frac{1}{2}\right)^{m+1}}{\frac{1}{2}}\right)}$$

$$b_{0,0} = \frac{2}{\left(W\left(m'+1\right) + W\frac{\frac{2^{m-m'}-1}{2^{m-m'}}}{\frac{1}{2}}\frac{1}{2} + \frac{\frac{2^{m+1}-1}{2^{m+1}}}{\frac{1}{2}}\right)}$$

[9] Note that if p=1 or p=1/2, the expression for τ in equation (C.4) cannot be used.

$$b_{0,0}=\frac{2}{\left(W\left(m'+1\right)+W\frac{2^{m-m'}-1}{2^{m-m'}}+\frac{2^{m+1}-1}{2^{m}}\right)} \tag{C.7}$$

Therefore, when p=1/2 equation (C.4) becomes:

$$\tau(1/2)=\sum_{i=o}^{m} b_{i,0}=\sum_{i=o}^{m}\left(\frac{1}{2}\right)^{i} b_{0,0}=\frac{2^{m+1}-1}{2^{m}} b_{0,0}=\frac{2^{m+1}-1}{2^{m-1}\left(W\left(m'+1\right)+W\frac{2^{m-m'}-1}{2^{m-m'}}+\frac{2^{m+1}-1}{2^{m}}\right)} \tag{C.8}$$

Moreover, when p=1 and by means of equation (C.3), we have:

$$b_{0,0}=\frac{2}{\left(\sum_{i=0}^{m} W_{i}+\sum_{i=0}^{m} 1^{i}\right)}$$

$$b_{0,0}=\frac{2}{\left(\sum_{i=0}^{m'}\left(2^{i}\cdot W\right)+\sum_{i=m'+1}^{m}\left(2^{m'}\cdot W\right)+(m+1)\right)}$$

$$b_{0,0}=\frac{2}{\left(\sum_{i=0}^{m'}\left(2^{i}\cdot W\right)+2^{m'}\cdot W\left(m-m'\right)+(m+1)\right)}$$

$$b_{0,0}=\frac{2}{W\frac{1-2^{m'+1}}{1-2}+2^{m'} W\left(m-m'\right)+(m+1)}$$

$$b_{0,0}=\frac{2}{W\left(2^{m'+1}-1\right)+2^{m'} W\left(m-m'\right)+(m+1)} \tag{C.9}$$

Therefore, when p=1 equation (C.4) becomes:

$$\tau(1)=\sum_{i=o}^{m} b_{i,0}=\sum_{i=o}^{m} b_{0,0}=(m+1) b_{0,0}=\frac{2(m+1)}{W\left(2^{m'+1}-1\right)+2^{m'} W\left(m-m'\right)+(m+1)} \tag{C.10}$$

Function $\tau(p)$ is continuous and monotone decreasing in the range $p \in (0,1)$ since it decreases from $\tau(0)=2/(W+1)$ to $\tau(1)$ given by equation (C.10). Uniqueness of the solution is proven by considering that $\tau(0)>\tau^{*}(0)$ and $\tau(1)<\tau^{*}(1)$.