Analysis and Improvement of Medium Access Control Protocols in Wireless Networks

Performance Modelling and Quality-of-Service Enhancement of IEEE 802.11e MAC in Wireless Local Area Networks under Heterogeneous Multimedia Traffic

Jia Hu

A thesis submitted for the degree of

Doctor of Philosophy

Department of Computing

School of Computing, Informatics and Media

University of Bradford

2010

Abstract

In order to efficiently utilize the scarce wireless resource as well as keep up with the ever-increasing demand for Quality-of-Service (QoS) of multimedia applications, wireless networks are undergoing rapid development and dramatic changes in the underlying technologies and protocols. The Medium Access Control (MAC) protocol, which coordinates the channel access and data transmission of wireless stations, plays a pivotal role in wireless networks.

Performance modelling and analysis has been and continues to be of great theoretical and practical importance in the design and development of wireless networks. This research is devoted to developing efficient and cost-effective analytical tools for the performance analysis and enhancement of MAC protocols in Wireless Local Area Networks (WLANs) under heterogeneous multimedia traffic. To support the MAC-layer QoS in WLANs, the IEEE 802.11e Enhanced Distributed Channel Access (EDCA) protocol has proposed three QoS differentiation schemes in terms of Arbitrary Inter-Frame Space (AIFS), Contention Window (CW), and Transmission Opportunity (TXOP). This research starts with the development of new analytical models for the TXOP scheme specified in the EDCA protocol under Poisson traffic. A dynamic TXOP scheme is then proposed to adjust the TXOP limits according to the status of the transmission queue. Theoretical analysis and simulation experiments show that the proposed dynamic scheme largely improves the performance of TXOP. To evaluate the TXOP scheme in the presence of

heterogeneous traffic, a versatile analytical model is developed to capture the traffic heterogeneity and model the features of burst transmission. The performance results highlight the importance of taking into account the heterogeneous traffic for the accurate evaluation of the TXOP scheme in wireless multimedia networks.

To obtain a thorough and deep understanding of the performance attributes of the EDCA protocol, a comprehensive analytical model is then proposed to accommodate the integration of the three QoS schemes of EDCA in terms of AIFS, CW, and TXOP under Poisson traffic. The performance results show that the TXOP scheme can not only support service differentiation but also improve the network performance, whereas the AIFS and CW schemes provide QoS differentiation only. Moreover, the results demonstrate that the MAC buffer size has considerable impact on the QoS performance of EDCA under Poisson traffic. To investigate the performance of EDCA in wireless multimedia networks, an analytical model is further developed for EDCA under heterogeneous traffic. The performance results demonstrate the significant effects of heterogeneous traffic on the total delay and frame losses of EDCA with different buffer sizes. Finally, an efficient admission control scheme is presented for the IEEE 802.11e WLANs based on analytical modelling and a game-theoretical approach. The admission control scheme can maintain the system operation at an optimal point where the utility of the Access Point (AP) is maximized with the QoS constraints of various users.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor Dr. Geyong Min, for his insightful guidance, constant encouragement and support through the course of my PhD. Without his rich expertise, deep insight, helpful advice, and willingness to provide funding, this thesis would not have been possible. His extensive knowledge and excellent analytical skills are certainly treasures to his students. His hard working attitude and dedication to top-quality research have inspired me to mature into a better researcher. Apart from supervision of my research work, he also provides warm helps on my life in UK. I feel he is not just an adviser but a good friend. It is my fortune and honour to have a supervisor like him.

I would like to thank Prof. Mike E. Woodward for taking time from his busy schedule to provide many insightful comments and suggestions on this work. I benefit greatly from his broad knowledge and invaluable advice.

I would also like to thank my fellow research students Lei Liu, Yulei Wu, Yaxiong Lei, Jianliang Gao, Shihang Yan, Lang Wang, Sha Sha, Hatem, and Junfeng Xu. I will treasure the joyful time spent with each of them, and how they have enriched my life and helped me.

I am indebted to my old friends Ming Xia, Songtao Chen, Qirui Huang, and Bin Chen. The meaningful and fun conversations with them have always served to ease my pressure and uplift my spirits.

I am eternally grateful to my mom and dad, my brother Ji Yang and his parents for their continued love and support for me during the course of my PhD career. I would like to dedicate the dissertation to my family.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACK	Acknowledgment
ACs	Access Categories
AIFS	Arbitrary Inter-Frame Space
AP	Access Point
BACK	Block Acknowledgment
BSS	Basic Service Set
CDMA	Code-Division Multiple Access
CFB	Contention Free Bursting
CRNs	Cognitive Radio Networks
CW	Contention Window
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CTMC	Continuous-Time Markov chain
CTS	Clear-to-Send
DCF	Distributed Coordination Function
DIFS	Distributed Inter-Frame Space
EDCA	Enhanced Distributed Channel Access
HCCA	Hybrid Coordinated Channel Access
HCF	Hybrid Coordination Function
HoB	Head-of-Burst
IDC	Index of Dispersion for Counts

IEEE	Institute of Electrical and Electronics Engineers
LRD	Long-Range-Dependent
MAC	Medium Access Control
MMPP	Markov-Modulated Poisson Process
NAV	Network Allocation Vector
NGHWNs	Next-Generation Heterogeneous Wireless Networks
PCF	Point Coordination Function
PHY	Physical
QoS	Quality-of-Service
RTS	Request-to-Send
SIFS	Short Inter-Frame Space
SRD	Short-Range Dependent
TCP	Transmission Control Protocol
TXOP	Transmission Opportunity
VBR	Variable-Bit-Rate
VoIP	Voice over Internet Protocol
WLANs	Wireless Local Area Networks
WMANs	Wireless Metropolitan Area Networks

Publications

1. J. Hu, G. Min, and M. E. Woodward, "Performance Analysis and Comparison of Burst Transmission Schemes in Unsaturated 802.11e WLANs", *Wiley Wireless Communications and Mobile Computing,* Accepted.

2. J. Hu, G. Min, and M. E. Woodward, "Performance Analysis of the TXOP Burst Transmission Scheme in Wireless Ad Hoc Networks with Unbalanced Stations", *Elsevier Computer Communications*, Accepted subject to minor revision.

3. J. Gao, J. Hu, G. Min, and L. Xu, "QoS Analysis of Medium Access Control in LR-WPANs under Bursty Error Channels", *Elsevier Future Generation Computer Systems,* Accepted.

4. G. Min, J. Hu, and M. E. Woodward, "Performance Modelling and Analysis of the TXOP Scheme in Wireless Multimedia Networks with Heterogeneous Stations", submitted.

5. J. Hu, G. Min, W. Jia and M. E. Woodward, "Comprehensive QoS Analysis of IEEE 802.11e EDCA", submitted.

6. J. Hu, G. Min, W. Jia, and M. E. Woodward, "Admission Control in the IEEE 802.11e WLANs based on Analytical Modelling and Game Theory", *Proc. IEEE GLOBECOM'09,* 2009

7. J. Gao, J. Hu, and G. Min, “Performance Modelling of IEEE 802.15.4 MAC in LR-WPAN with Bursty ON-OFF Traffic”, *Proc. IEEE CIT’09*, 2009

8. G. Min, J. Hu, W. Jia, and M. E. Woodward, “Performance Analysis of the TXOP Scheme in IEEE 802.11e WLANs with Bursty Error Channels”, *Proc. IEEE WCNC’09*, 2009.

9. J. Gao, J. Hu, and G. Min, “A New Analytical Model for Slotted IEEE 802.15.4 Medium Access Control Protocol in Sensor Networks”, *Proc. IEEE CMC’09*, 2009.

10. J. Hu, G. Min, M. E. Woodward, and W. Jia, “A Comprehensive Analytical Model for IEEE 802.11e QoS Differentiation Schemes under Unsaturated Traffic Loads”, *Proc. IEEE ICC’08*, pp. 241-245, 2008.

11. G. Min, J. Hu, and M. E. Woodward, “A Dynamic IEEE 802.11e TXOP Scheme in WLANs under Self-Similar Traffic: Performance Enhancement and Analysis”, *Proc. IEEE ICC’08*, pp. 2632-2636, 2008.

12. J. Hu, G. Min, M. E. Woodward, and W. Guo, “Performance Modelling and Analysis of IEEE 802.11e Contention Free Burst Scheme under Unsaturated Traffic", *Wireless Quality of Service: Techniques, Standards and Applications,* M. Ma, M. Denko, Y. Zhang (ed.), Auerbach Publications, Taylor & Francis Group, 2008.

13. G. Min, J. Hu, and M. E. Woodward, "An Analytical Model of the TXOP Scheme with Heterogeneous Classes of Stations", *Proc. IEEE GLOBECOM'08*, 2008.

14. J. Hu, G. Min, and M. E. Woodward, "Analysis and Comparison of Burst Transmission Schemes in Unsaturated 802.11e WLANs", *Proc. IEEE GLOBECOM'07*, pp. 5133-5137, 2007.

15. J. Hu, G. Min, and M. E. Woodward, "Modeling of IEEE 802.11e Contention Free Bursting Scheme with Non-Identical Stations", *Proc. IEEE/ACM MASCOTS'07*, pp. 88-94, 2007.

16. G. Min, J. Hu, and M. E. Woodward, "Performance Modelling of TXOP Differentiation in Infrastructure-Based WLANs", *Proc. IEEE LCN'07*, pp. 260-261, 2007.

Chapter 1

Introduction

The IEEE 802.11-based Wireless Local Area Networks (WLANs) [58] have become ubiquitous over recent years, especially encouraged by the success of the Internet and the proliferation of portable devices, such as the laptop computers and personal digital assistants. With the rapid deployment of WLANs, academic and industrial communities have carried out in-depth research activities by analytical or numerical means to gain insights into the Quality-of-Service (QoS) performance metrics of WLANs. The architecture of the IEEE 802.11 standard [58] includes the definition of the physical (PHY) layer and the Medium Access Control (MAC) sublayer. The original MAC sublayer employs a mandatory contention-based channel access function called Distributed Coordination Function (DCF) which is based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol [69]. This standard also specifies an optional polling-based channel access function called Point Coordination Function (PCF). However, PCF is rarely implemented in commercially available WLANs and receives little attention due to its complexity and inefficiency [88].

With the rapid growth in the popularity of multimedia applications such as Voice-over-IP and (VoIP) and video conferencing, the demand for high bandwidth and differentiated QoS in WLANs is increasing dramatically [28, 132]. To support

MAC-level QoS, an enhanced version of the IEEE 802.11 MAC protocol, namely IEEE 802.11e [59], has been standardized. The IEEE 802.11e MAC employs a channel access function called Hybrid Coordination Function (HCF) [59], which comprises the contention-based Enhanced Distributed Channel Access (EDCA) and the centrally controlled Hybrid Coordinated Channel Access (HCCA). EDCA is the fundamental and mandatory mechanism of 802.11e, whereas HCCA is optional and requires complex scheduling algorithms for resource allocation. EDCA classifies the traffic flows into four Access Categories (ACs) [59], each of which is associated to a separate transmission queue and behaves independently. These ACs are differentiated through adjusting the parameters of Arbitrary Inter-frame Space (AIFS), Contention Window (CW) and Transmission Opportunity (TXOP) limit [59]. The AIFS and CW decide the deferring time and the backoff time of the AC that contends for the channel, respectively, while the TXOP limit controls the channel occupation time of the AC that gains the channel.

The rest of this chapter is organized as follows: the motivations and challenges of this research are pointed out in Section 1.1. The research aims and major contributions of this thesis are then introduced in Section 1.2. Finally, the outline of this thesis is presented in Section 1.3.

1.1 Motivations and Challenges

Performance modelling and analysis has been and continues to be of great theoretical and practical importance in the design, development and optimization of wireless communication networks. To obtain an in-depth understanding of the performance characteristics of WLANs, significant research efforts have been devoted to developing the analytical models for the DCF and EDCA protocols. However, there are still some practical and open research issues related to the DCF and EDCA protocols that need to be further tackled. For example,

1) The realistic working conditions of WLANs are often unsaturated as only very few networks are in a situation where all nodes have frames to send anytime [87, 122]. Moreover, in the practical working environment, wireless channels are imperfect and error-prone [34]. Therefore, although the TXOP scheme specified in EDCA has been extensively studied under saturated network and perfect channel conditions, it is imperative to investigate its performance behaviors with unsaturated traffic loads and imperfect channels.

2) In the original TXOP scheme, the TXOP limit at each station is fixed [58]. However, the bursty property of multimedia traffic implies that the large bursts of packet arrivals occur frequently. It is thus desirable to dynamically adjust the TXOP limits according to the status of the transmission queue in order to meet the specific QoS requirements of multimedia applications. Furthermore, a new analytical model is needed to evaluate the efficiency and effectiveness of the dynamic TXOP scheme.

3) The existing analytical models for EDCA have been primarily focused on the AIFS, CW, and TXOP schemes, separately. In addition, these models are mostly based on the assumptions that the networks operate under saturated conditions and the MAC buffer has either a very small size or an infinite capacity. Furthermore, important performance measures for real-time applications such as end-to-end delay and delay jitter are often ignored in the existing models. Consequently, it is necessary to develop a comprehensive and accurate analytical model for EDCA in order to

   - integrate all three QoS differentiation schemes in WLANs with finite buffer capacity in the presence of unsaturated traffic loads;
   - derive the important QoS performance metrics in terms of throughput, delay, delay jitter, and frame loss probability.

4) WLANs are currently integrating a diverse range of traffic sources that significantly differ in their packet arrival patterns. Although initially successful and analytically simple for modelling the non-bursty traffic behavior, the Poisson model has proven inadequate for capturing traffic burstiness of compressed voice and video in modern communication networks, where batch arrivals, event correlations and burstiness are important factors. It is well known that Variable-Bit-Rate (VBR) VoIP generates traffic with time-varying arrival rates [77, 114]. Furthermore, many studies [8, 9, 15, 40] by means of high quality, high time-resolution measurements have demonstrated that VBR video

traffic exhibits noticeable burstiness over a wide range of time scales. This fractal behaviour of video traffic should be modelled using statistically self-similar processes [104], which have significantly different theoretical properties from those of the conventional Poisson process. Therefore, it is critical and timely to take the heterogeneous characteristics of multimedia traffic into account in order to accurately evaluate and obtain a better understanding of the performance characteristics of multimedia WLANs.

1.2 Research Aims and Contributions

This research is focused on the analysis and enhancement of the IEEE 802.11 and 802.11e MAC protocols in multimedia WLANs under practical working environments. The main objectives of this research are:

- To develop reliable, efficient, and cost-effective analytical tools for IEEE 802.11 DCF and 802.11e EDCA protocols in WLANs with heterogeneous multimedia applications.
- To use these analytical models to investigate the intrinsic performance characteristics of DCF and EDCA protocols and develop efficient resource allocation schemes for multimedia WLANs.
- To enhance the performance of the IEEE 802.11e WLANs with novel MAC schemes and admission control mechanisms in the presence of wireless multimedia traffic.

To reach these aims, the research develops new analytical tools for the performance analysis and enhancement of wireless MAC protocols in the presence of heterogeneous multimedia traffic. The accuracy of the proposed models is validated through the extensive comparison of the analytical performance results with those obtained from NS-2 [99] simulation experiments. The original contributions of this research are summarized as follows:

- A new analytical model is developed for evaluating the burst transmission schemes in terms of Contention Free Burst (CFB) and Block Acknowledgement (BACK) [59] under unsaturated traffic loads and finite buffer capacity. A thorough investigation has been conducted into the impact of the traffic load, TXOP limit, buffer size, channel data rate, minimum contention window, and number of stations on the QoS performance of the burst transmission schemes.

- A novel and comprehensive analytical model for the TXOP scheme in WLANs consisting of unbalanced stations with different traffic loads is developed. The model derives the expressions of the important QoS performance metrics including throughput, end-to-end delay, frame dropping probability, and energy consumption of the TXOP scheme.

- A new analytical model is developed for the TXOP scheme in unsaturated WLANs with bursty error channels. The transmission queue of each station is modelled by a two-state Continuous-Time Markov Chain (CTMC) [19] that captures the bursty characteristics of the wireless channel errors and the burst

transmission mechanism. This model is used to investigate the impact of traffic loads, TXOP limits, network sizes, and channel conditions on the performance of TXOP.

- A new dynamic TXOP scheme is presented, which can adjust the TXOP limits according to the status of the transmission queue and the pre-setting threshold. An original analytical performance model is developed for this dynamic TXOP scheme subject to self-similar traffic. The analytical performance results demonstrate that the dynamic TXOP scheme can achieve the better QoS performance than the original scheme under self-similar traffic.

- A versatile model is proposed to address the difficulties of queueing analysis arising from the bulk service of the TXOP scheme in the presence of heterogeneous traffic. The model validations are subject to the traffic parameters obtained from the accurate measurements of the real-world multimedia voice and video sources. The results highlight the importance of taking into account the heterogeneous traffic for the accurate evaluation of the TXOP scheme in multimedia WLANs.

- A comprehensive analytical model is developed for EDCA with the integration of the three QoS schemes in terms of AIFS, CW, and TXOP. The performance metrics including throughput, end-to-end delay, delay jitter, and frame loss probability are derived. Performance results show that the TXOP scheme can not only support service differentiation but also improve the network

performance, whereas the AIFS and CW schemes provide QoS differentiation only.

- An analytical model is proposed to investigate the effects of heterogeneous traffic on the QoS of EDCA. The traffic parameters used in the validations are obtained from the accurate measurements of the multimedia applications including the G.711 codec [43] voice sources and the VBR encoded H.263 [40] video streams. The results reveal the significant effects of heterogeneous traffic on the total delay and frame losses of EDCA with different buffer sizes.

- An admission control scheme is presented for the IEEE 802.11e WLANs based on analytical modelling and a game theoretical approach [46]. The numerical results demonstrate that the proposed admission control scheme can maintain the system operation at an optimal point where the utility of the AP is maximized with the QoS constraints of various users.

1.3 Outline of the Thesis

The rest of this thesis is organized as follows.

Chapter 2 firstly introduces the background knowledge including DCF and EDCA protocols, traffic models, and channel models. A detailed literature review on the modelling and analysis of DCF and EDCA protocols is then presented.

Chapter 3 presents analytical models for the TXOP scheme under unsaturated conditions. Specifically, the first model deals with different ACK policies; the

second one considers the unbalanced stations in WLAN; and the third one takes the bursty channel errors into account. These models are then used to conduct a thorough performance evaluation of the TXOP scheme.

Chapter 4 proposes a new dynamic TXOP scheme and develops an original analytical performance model for this dynamic TXOP scheme subject to self-similar traffic.

Chapter 5 develops an analytical model for the TXOP scheme in WLANs with heterogeneous stations. The analytical model is used to investigate the effects of the TXOP scheme on the QoS of heterogeneous stations.

Chapter 6 develops a comprehensive analytical model for the EDCA protocol with the AIFS, CW, and TXOP schemes. This model is then used as a cost-efficient tool for performance evaluation of EDCA.

Chapter 7 analyzes the performance of EDCA with heterogeneous multimedia applications. An admission control scheme is also presented for the IEEE 802.11e WLANs based on analytical modelling and a game theoretical approach.

Finally, Chapter 8 concludes the thesis and highlights the future research work.

Chapter 2

Background and Literature Review

The past few years have witnessed the fact that WLANs have become ubiquitous and revolutionized our daily life. With the exponential growth of popular multimedia applications such as VoIP, video conferencing, and online games, there is an ever-increasing demand for provisioning of QoS over WLANs. However, the scarce resource and variable transmission quality of wireless channels pose significant challenges on the QoS provisioning over wireless networks with multimedia applications. To support the MAC-level QoS, an enhance version of the IEEE 802.11 DCF, namely the IEEE 802.11e EDCA, has been proposed. To obtain an in-depth understanding of the performance attributes of DCF and EDCA, a significant number of analytical models for these protocols have been developed over the past few years. This chapter presents the background knowledge and related work on modelling the DCF and EDCA in WLANs with multimedia applications.

The rest of this chapter is organized as follows. The background knowledge including the IEEE 802.11 DCF and 802.11e EDCA MAC Protocols, traffic models, and wireless channel models is introduced in Section 2.1, 2.2, and 2.3, respectively. A detailed literature review on modelling and analysis of DCF and EDCA as well as the model-based admission control approaches is then presented in Section 2.4. Section 2.5 concludes this chapter.

2.1 Wireless Local Area Networks (WLANs)

Wireless networks are becoming an integral part of our daily life, due to the flexibility and convenience they offer. The applications of wireless networks include e-commerce, personal communications, telecommunications, monitoring environments, emergency operations and wireless Internet access. WLANs use radio waves to communicate information from one device to another without relying on any physical connection. WLANs can either extend or replace wired LANs to provide the connectivity between a backbone network and the in-building or on-campus users. Over the past few years, WLANs have become one of the most popular wireless networks and have experienced enormous market success.

The 802.11 specification defines two types of operational modes in WLANs: ad hoc (peer-to-peer) mode and infrastructure mode [58]. In ad hoc mode, the wireless nodes communicate directly with one another without the use of an access point. In infrastructure mode, the WLAN is composed of an AP and a group of wireless nodes. The AP acts as a base station in an 802.11-based WLAN and communications from all of the wireless nodes go through the access point. A fundamental WLAN infrastructure with a single AP is called a Basic Service Set (BSS).

2.2 Medium Access Control (MAC)

The MAC protocol, which coordinates the data transmission of wireless stations, plays a pivotal role in wireless networks [69]. When two or more stations transmit simultaneously over the wireless channel, a collision happens and the receivers cannot successfully decode the transmitted frames. Therefore, the MAC protocol is needed to determine when and how a particular station accesses the wireless channel. This section gives an overview on the popular IEEE 802.11 and 802.11e MAC protocols that are widely deployed in WLANs.

2.2.1 Distributed Coordination Function (DCF)

DCF is the fundamental MAC scheme ratified in the IEEE 802.11 MAC protocol [58], where a station senses the channel before attempting the transmission of frames. As shown in Fig. 2.1, if the channel is detected idle for a Distributed Inter-frame Space (DIFS), the transmission starts. Otherwise, if the channel is sensed busy (either initially or during the DIFS), the station defers until the channel is detected idle for a DIFS, and then generates a random backoff counter.

The value of the backoff counter is uniformly chosen in the range $[0, W_i - 1]$, where W_i is the current contention window size and i is the backoff stage. At the first transmission attempt, W_i is set to be the minimum contention window W. After each unsuccessful transmission, W_i is doubled, up to a maximum value $W_m = 2^m W$, where m represents the maximum backoff stage. It remains at the value W_m until the transmission succeeds or the number of retransmission attempts reaches a retry limit. The backoff counter is decreased by one for each time slot (an

interval of fixed duration specified in the protocol [58]) when the channel is idle, halted when the channel becomes busy and resumed when the channel is idle again for a DIFS. A station transmits a frame when its backoff counter reaches zero. Other stations which hear the transmission of the frame set their Network Allocation Vector (NAV) to the expected period of time in which the channel will be busy. This is called the virtual carrier sensing mechanism. If either the virtual carrier sensing or physical carrier sensing [58] indicates that the channel is busy, the station commences the backoff procedure.

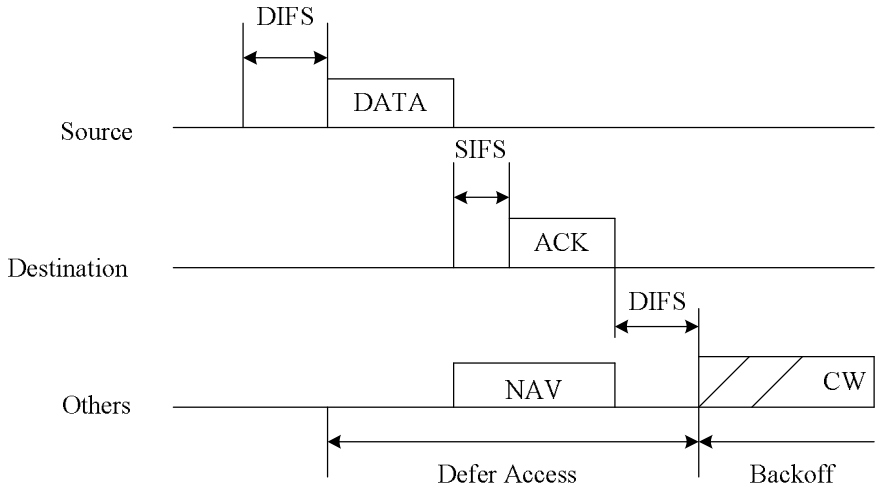


Fig. 2.1: The basic access mechanism of DCF.

Upon the successful reception of the frame, the destination station sends back an Acknowledgement (ACK) frame immediately following a Short Inter-frame Space (SIFS) interval. If the station does not receive the ACK within a timeout interval [58], it retransmits the frame. Each station maintains a retry counter that is increased by one after each retransmission. The frame is discarded if the number of retransmission attempts reaches the retry limit. Hidden terminal problems [127]

occur when a station is unable to detect a potential competitor for the channel because they are not within the transmission range of each other. To combat the hidden terminal problems, DCF also defines an optional four-way handshake scheme whereby the source and destination exchange Request-To-Send (RTS) and Clear-To-Send (CTS) messages before the transmission of actual data frame.

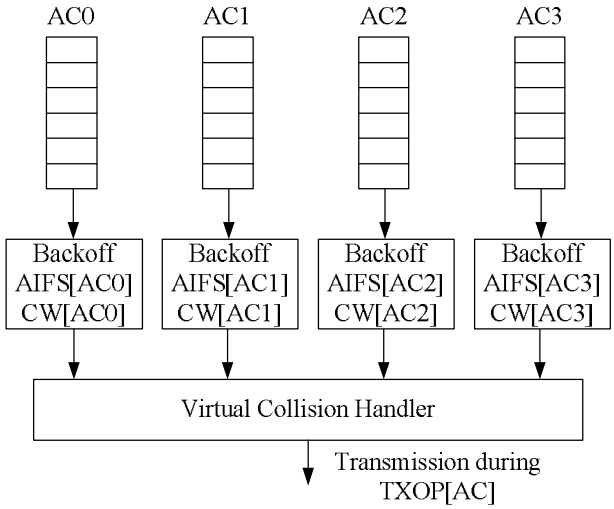


Fig. 2.2: The IEEE 802.11e MAC with four ACs.

2.2.2 Enhanced Distributed Channel Access (EDCA)

EDCA was designed to improve the performance of DCF and provide the differentiated QoS [59]. As shown in Fig 2.2, the traffic of different classes is assigned to one of four ACs, which is associated to a separate transmission queue and behaves independently. The QoS of these ACs is differentiated through assigning various EDCA parameters including AIFS values, CW sizes, and TXOP limits. Specifically, a smaller AIFS/CW leads to a larger probability of winning the

contention for the channel. On the other hand, the larger the TXOP limit is, the longer is the channel holding time of the AC winning the contention.

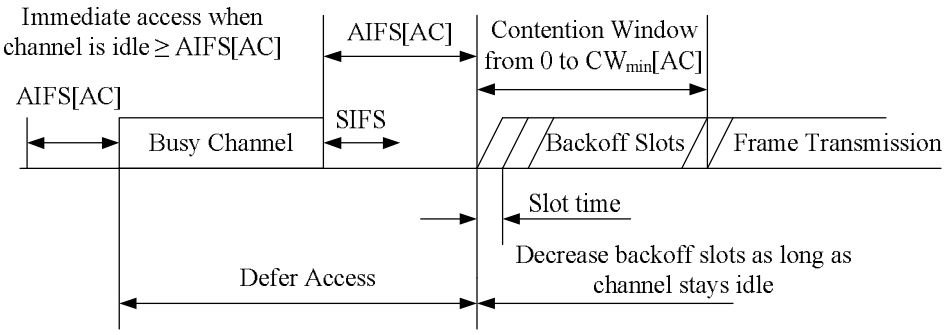


Fig. 2.3: The timing diagram of the EDCA channel access.

In the EDCA protocol, the channel is sensed before an AC attempts to transmit frames. If the channel is detected idle for an AIFS, the transmission starts. Otherwise, the AC defers until the channel is detected idle for an AIFS, and then generates a random backoff counter (Fig. 2.3). The AIFS for a given AC is defined as $AIFS_{[AC]} = SIFS + AIFSN_{[AC]} \times aSlotTime$, where $AIFSN_{[AC]}$ ($AIFSN_{[AC]} \geq 2$) represents the number of time slots in $AIFS_{[AC]}$ and $aSlotTime$ denotes the duration of a time slot [59].

The value of the backoff counter is uniformly chosen between zero and $CW_{[AC]}$, which is initially set to $CW_{\min[AC]}$ and doubled after each unsuccessful transmission until it reaches the maximum value $CW_{\max[AC]}$. It is reset to $CW_{\min[AC]}$ after the transmission succeeds or the number of retransmission attempts reaches a retry limit. The backoff counter is decreased by one for each time slot [59] when the channel is idle, halted when the channel becomes busy and resumed when the channel is idle again for an AIFS. An AC transmits when its

backoff counter becomes zero. When the backoff counters of different ACs within a station decrease to zero simultaneously, the frame from the highest priority AC among the contending ones is selected for transmission on the channel, while the others suffer from a virtual collision and invoke the backoff procedure.

The backoff rule of EDCA is slightly different from that of DCF. In DCF, the backoff counter is frozen during the channel busy period, resumed after the channel is sensed idle for a DIFS and decreased by one at the end of the first slot following DIFS. However, the backoff counter in EDCA is resumed one slot time before the end of AIFS [59]. It means that the backoff counter has already been decremented by one at the end of AIFS. In addition, after the backoff counter decrements to zero, the AC has to wait for an extra slot before transmitting.

Upon winning the access to the channel, the AC transmits the frames available in its buffer consecutively provided that the duration of transmission does not exceed the specific TXOP limit [59]. As shown in Fig. 2.4, each frame is acknowledged by an Acknowledgement (ACK) after a Short Inter-frame Space (SIFS) interval. The next frame is transmitted immediately after receiving the ACK and waiting for an SIFS. If the transmission of any frame fails the burst is terminated and the AC contends again for the channel to retransmit the failed frame. The TXOP scheme is an efficient way to improve the channel utilization of wireless MAC protocols because the contention overhead is shared between all the frames transmitted in a burst. Moreover, it enables service differentiation between multiple traffic classes by virtue of various TXOP limits.

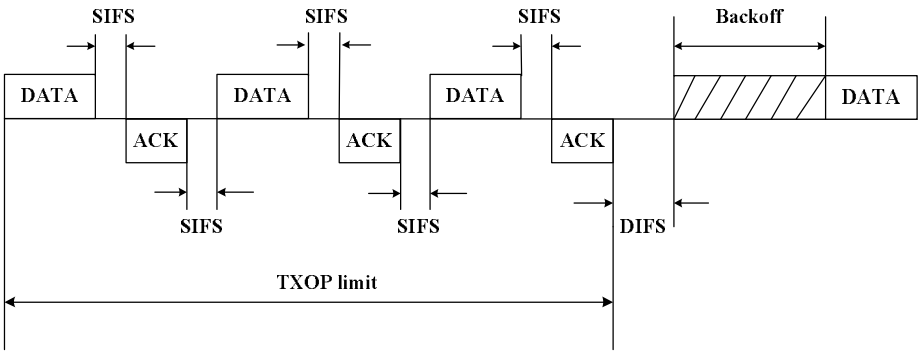


Fig. 2.4: The TXOP scheme, i. e., Contention Free Bursting (CFB).

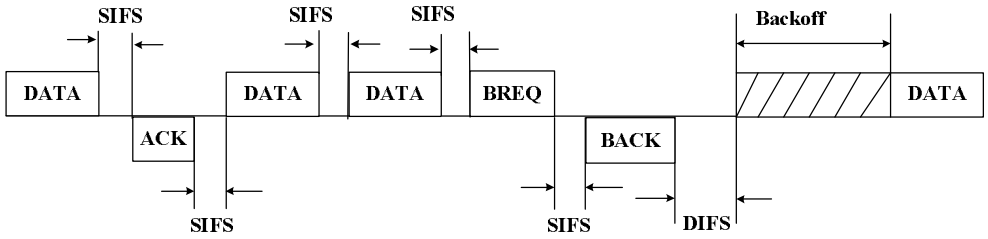


Fig. 2.5: The TXOP scheme with a new ACK policy, i. e., Block Acknowledgement (BACK).

Fig. 2.5 depicts the Block Acknowledgement (BACK) scheme, which is the TXOP scheme with a new ACK policy. This scheme improves the efficiency by aggregating several ACK frames of a burst into one single ACK frame [125]. The next frame is transmitted immediately after waiting for an SIFS. The burst of frames is acknowledged by a *BlockAck* control frame, which is requested by a *BlockAckReq* control frame from the sender. The bitmap field in the *BlockAck* control frame indicates the reception status of all frames within the burst. Those frames which are not received successfully will be retransmitted in the subsequent burst. In order to quickly identify collisions which occur during the transmission of the Head-of-Burst (HoB) frame, an immediate ACK is used to protect the HoB frame of each burst.

2.3 Traffic Models

In recently years, there has been an ever-increasing demand for provisioning of multimedia services such as VoIP and real-time video in wireless networks. In order to conduct effective analysis and evaluation of the performance attributes of wireless networks, suitable traffic models that can accurately capture the characteristics of the heterogeneous multimedia traffic are required. The following subsections present the popular traffic models including non-bursty Poisson, bursty ON-OFF and Markov Modulated Poisson Process (MMPP), and fractal Self-similar process that are used to model multimedia traffic generated by the heterogeneous stations in WLANs.

2.3.1 Poisson Process

The Poisson process has been widely used to model the traffic behaviour in many communication networks and systems [1, 3, 22, 61, 87, 91, 106, 119, 140, 144]. The inter-arrival intervals of Poisson process are independent and obey the exponential distribution. The memoryless property of the exponential distribution (*i.e.*, the future progress does not depend on its past) facilitates considerably the analysis of queueing systems. However, the number of arrivals found in this process during any finite interval depends only on the length of the interval and not on its starting point. As a result, the Poisson process is not able to model the bursty behaviour of multimedia traffic with time-varying arrival rates.

2.3.2 ON-OFF Process

The ON-OFF process has been widely used to model the bursty properties of voice traffic [43, 110, 114]. This process alternates between ON and OFF state, where the arrivals are generated with constant rate λ_p during ON state and there is no arrival during OFF state. The durations of the ON and OFF periods are exponentially distributed with means $1/v_{on}$ and $1/v_{off}$, respectively. The ON-OFF traffic source is thus a stream of deterministically distributed correlated bursts and silent periods. The mean traffic arrival rate, λ, is given by

$$\lambda = \frac{\lambda_p v_{off}}{v_{off} + v_{on}} \tag{2.1}$$

2.3.3 Markov-Modulated Poisson Process (MMPP)

The MMPP is a doubly stochastic Poisson process with the arrival rate varying according to an irreducible continuous-time Markov chain [39]. As a consequence, this stochastic process is able to capture the time-varying arrival rate. Furthermore, the superposition and splitting of MMPPs give rise to a new MMPP [39]. Such features have made the MMPP very attractive for modelling bursty traffic. In particular, a two-state MMPP can be used to model the superposition of multiple ON-OFF voice sources [114]. Thus, in this study, the aggregate voice traffic is modelled by an MMPP, which is parameterized by the infinitesimal generator $\mathbf{Q}$ of the underlying Markov chain and the rate matrix $\mathbf{\Lambda}$ as follows

$$\mathbf{Q} = \begin{bmatrix} -\delta_1 & \delta_1 \\ \delta_2 & -\delta_2 \end{bmatrix} \qquad \text{and} \qquad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{2.2}$$

where the element δ_1 is the transition rate from state 1 to 2 of the MMPP and δ_2 is the rate out of state 2 to 1. λ_1 and λ_2 are the traffic rates when the Markov chain is in state 1 and 2, respectively.

For an ON-OFF voice source, traffic is generated at a fixed rate of A (frames/sec) during talk spurts (*i.e.*, in the ON state) and no frame arrives during silences (i.e., in the OFF state). The ON and OFF periods are exponentially distributed. Let α^{-1} and β^{-1} denote the mean sojourn time in the ON and OFF states, respectively. A superposition of k ON-OFF voice sources can be modelled by a two-state MMPP whose parameters are determined by virtue of the Index of Dispersion for Counts (IDC) matching technique [114]

$$\lambda_1 = \mathrm{A}\frac{\sum_{i=0}^{k'} i\pi_i}{\sum_{j=0}^{k'} \pi_j} \qquad \text{and} \qquad \lambda_2 = \mathrm{A}\frac{\sum_{i=k'+1}^{k} i\pi_i}{\sum_{j=k'+1}^{k} \pi_j} \tag{2.3}$$

$$\delta_1 = \frac{2(\lambda_2 - \lambda_{avg})(\lambda_{avg} - \lambda_1)^2}{(\lambda_2 - \lambda_1)\lambda_{avg}(\mathrm{IDC}(\infty) - 1)} \qquad \text{and} \qquad \delta_2 = \frac{2(\lambda_2 - \lambda_{avg})^2(\lambda_{avg} - \lambda_1)}{(\lambda_2 - \lambda_1)\lambda_{avg}(\mathrm{IDC}(\infty) - 1)}$$

(2.4)

where

$$\pi_j = \frac{k!}{j!(k-j)!} q^j (1-q)^{k-j}, q = \frac{\beta}{\alpha + \beta}, \; k' = \lfloor qk \rfloor, \lambda_{avg} = k\mathrm{A}q, \text{ and } \mathrm{IDC}(\infty) =$$

$$\frac{1-(1-\alpha/\mathrm{A})^2}{(\alpha/\mathrm{A} + \beta/\mathrm{A})^2}.$$

2.3.4 Self-Similar Process

Many recent accurate measurement studies [8, 9, 15, 40] have demonstrated that the traffic generated by VBR videos exhibits noticeable bursty nature over a wide range of time scales. This fractal behaviour of network traffic can be modelled using statistically self-similar or long-range-dependent processes, which have significantly different theoretical properties from those of the conventional short-range-dependent processes, such as Poisson process and MMPP. Self-similar traffic can considerably deteriorate the user-perceived QoS and has drawn significant interests and received tremendous research efforts from both academia and industry [6, 15, 20, 32, 48, 63, 76, 89, 94, 104]. Moreover, traffic self-similarity has been shown to appear in wireless networks [60, 79, 123].

Let $\mathbf{X} = (X_t : t = 0, 1, 2, \cdots)$ be a covariance stationary stochastic process with the autocorrelation function $r(k)$. For each $m = 1, 2, 3, \cdots$, let $\mathbf{X^{(m)}} = (X_k^{(m)} : k = 1, 2, \cdots)$ denote the new covariance stationary time series (with corresponding autocorrelation function $r^{(m)}$) obtained by averaging the original series $\mathbf{X}$ over non-overlapping blocks of size m. Then the following definitions can be given [76]:

Definition 1: The process $\mathbf{X}$ is called second order self-similar if the autocorrelation function of the aggregated process $\mathbf{X^{(m)}}$ is identical to the autocorrelation function of the original process $\mathbf{X}$ in the limit of large k . *i.e.*,

$$\lim_{k \to \infty} \frac{r^{(m)}(k)}{r(k)} = 1, \quad \text{for all } m \tag{2.5}$$

Definition 2: The process $\mathbf{X}$ is called LRD if its autocorrelation function $r(k)$ decays so slowly that its sum diverges:

$$\lim_{n\to\infty}\sum_{k=1}^{n} r(k) = \infty \tag{2.6}$$

A sufficient condition for long-range dependence to occur is that the autocorrelation function drops off with a power-law:

$$r(k) \sim \frac{1}{k^{\alpha-1}}, \qquad 1 < \alpha < 2 \tag{2.7}$$

The LRD process whose autocorrelation function decays as a power-law with an exponent α, can be shown to be second order self-similar [76] with Hurst parameter $H = (3-\alpha)/2.$

2.4 Channel Models

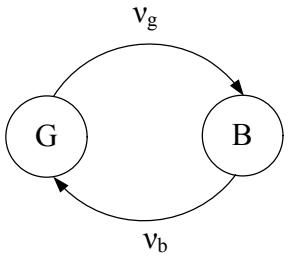


Fig. 2.6: Two-State Markov Model of Wireless Channels.

In the practical working environment, wireless channels are error-prone and the occurrences of channel errors are bursty and highly correlated due to fading and shadowing [106]. The bursty characteristics of channel errors can be modelled by a two-state Markov model [49, 102, 106, 137]. As shown in Fig. 2.6, the wireless

channel is characterized by a Markov chain alternating between a good state and a bad state. The transition rate from the good to the bad state is represented by ν_g while that from the bad to the good state is denoted as ν_b. Thus, the durations of the good state and bad state are exponential random variables with means ν_g^{-1} and ν_b^{-1}, respectively. In the good state, the channel is error free and a frame transmission would be successful if there is no collision, while in the bad state frames transmitted over the channel are corrupted. The steady-state probabilities of the channel being in the good and bad states, π_g and π_b, can be given by

$$\pi_g = \frac{\nu_b}{\nu_g + \nu_b} \tag{2.8}$$

$$\pi_b = \frac{\nu_g}{\nu_g + \nu_b} \tag{2.9}$$

2.5 Literature Review

There has been a significant amount of work on the analysis and modelling of 802.11 and 802.11e MAC protocols. In this section, the prior work relevant to analytical modeling of DCF and EDCA and the model-based admission control approaches are surveyed. How this research fills some of the important gaps left in the existing work is also discussed.

2.5.1 Analytical Modelling of DCF

Performance analysis of DCF has come under much scrutiny in recent years [1, 5, 10, 11, 13, 16, 17, 22-26, 29, 30, 33-35, 41, 45, 49, 54, 66, 72, 73, 82, 86, 87, 91, 95, 96, 103, 122, 133, 139, 140, 144, 145, 147]. Cali, Conti and Gregori [23] have presented the analysis of the saturated throughput of p-persistent CSMA/CA protocol by modelling the backoff counter value as a geometric distribution. Their study has revealed the possibility of maximizing the throughput by tuning the backoff window size at run time. Bianchi's well-known analytical model [16] has adopted a bi-dimensional discrete time Markov Chain to derive the saturation throughput for the DCF, assuming the ideal channel conditions (*i.e.*, no hidden terminals and capture effects [127]). Many subsequent studies have built upon Bianchi's work. For instance, Ziouva and Antonakopoulos [147] have improved Bianchi's model by taking account of the busy medium conditions for invoking the backoff procedure. Based on their model, the performance measures in terms of throughput and the average service time have been calculated. Wu *et al.* [130] have modified Bianchi's model to deal with the retry limit. They have also proposed a new scheme called DCF+ which is more suitable than DCF for Transmission Control Protocol (TCP). Kumar *et al.* [72] have studied the fixed point formulation based on the analysis of Bianchi's model and showed that the derivation of transmission probability can be significantly simplified by removing the Markovian assumptions.

The queueing delay becomes the dominant factor of the total delay. However, the vast majority of the saturation models have only considered throughput, access

delay, and service time, but neglected the queueing delay. For example, Carvalho and Garcia-Luna-Aceves [25] have performed the analysis of the average service time as well as jitter. Chatzimisios, Boucouvalas, and Vitsas [26] have adopted Bianchi's model to handle the retry limit and calculate the access delay (the delay seen by the frame at the head of queue) on the saturation state. Zenella and De Pellegrini [139] have derived a close-form probability generating function for the service time in saturated networks.

The aforementioned saturated models assume that all stations in the network always have frames to transmit and thus exclude any need to consider queuing dynamics or traffic models for performance analysis. However, realistic network conditions are unsaturated as very few networks are in a situation where all nodes have frames to send all the time. Therefore, it is important to develop analytical models for DCF under the unsaturated condition. There are many studies focused on modelling DCF under unsaturated working conditions [1, 5, 22, 24, 34, 35, 41, 45, 87, 91, 95, 122, 133, 140, 144]. For instance, Tickoo and Sikdar [122] have used a discrete-time G/G/1 queueing system with infinite buffer capacity to model the DCF and extended the queueing model to analyze the TXOP scheme under unsaturated traffic loads. This model is accurate at low and high loads but overestimates the delay at medium loads because of the approximation adopted to calculate the burst size. After deriving the saturation transmission probability through the average value analytical model, Tay and Chua [121] have obtained the transmission probability of finite loads by weighting the saturation transmission probability with the probability

of the transmission queue being non-empty. Medepalli and Tobagi [91] have developed a unified analytical model where the unsaturated transmission probability is also derived by weighting the saturation transmission probability with the probability of non-empty M/M/1 queue. The key features of their model possess the ability to handle hidden/exposed terminals, directional antennas, multiple channels, and arbitrary traffic matrices. Ozdemir and McDonald [100] have obtained the unsaturated performance metrics based on the M/G/1/K queueing model where the service time distribution is modelled by a Markov-modulated general distribution. Moreover, they have removed the fundamental assumption of Bianchi's model that every frame collides with a constant and independent probability regardless of the number of retransmissions it suffered. Miorandi, Khereni, and Altman [95] have taken a processor sharing view of DCF to evaluate the performance of HTTP traffic over IEEE 802.11. They have proved that setting the TCP's advertised window size to a small value leads to insensitivity of mean file transfer times to the file size distribution. Zhao, Tsang, and Sakurai [144] have developed a simple model to approximate an unsaturated 802.11 network, assuming either a very small buffer or an infinite buffer at the MAC layer.

None of the above-mentioned unsaturated models have taken into account the case of unbalanced stations. Cantieni *et al.* [24] have introduced an unsaturated model considering the multi-rate capabilities where the MAC buffer is modelled by an M/G/1 queue. In addition, they have invented a new fairness metric for general CSMA/CA multirate networks. Malone, Duffy and Leith [87] have presented an

extension of Bianchi's model to an unsaturated environment in the presence of unbalanced loads with the assumption of no more than one frame buffered at each MAC. Their model has captured several important features of unsaturated operation, for instance, predicting the maximum throughput. To investigate the effect of buffering on resource allocation, Duffy and Ganesh [35] have further extended the model developed in [87] for stations with large buffers and Poisson arrivals.

Most of the existing analytical models on DCF have been developed under the saturated conditions or non-bursty Poisson traffic. However, with the explosive growth in multimedia applications such as voice and video, it is important to take the bursty and correlated nature of multimedia traffic into account when analyzing the performance of DCF.

2.5.2 Analytical Modelling of EDCA

Modelling and performance analysis of the AIFS and CW schemes specified in EDCA have received considerable research interests [3, 18, 28, 36, 42, 43, 53, 55-57, 61, 67, 70, 71, 75, 83, 96, 108, 109, 118-120, 131, 132, 134, 146]. Most of these studies were based on Bianchi's two-dimensional (2D) Markov chain [16] under the assumption of saturated traffic conditions. For instance, Xiao [131] have extended the Markov chain proposed in [16] to model the CW differentiation scheme of EDCA. Tao and Panwar [120] have developed another three-dimensional Markov Chain model where the third dimension represents the number of time slots after the

end of the last AIFS period. Robinson and Randhawa [109] have adopted a bi-dimensional Markov chain model where the collision probability is calculated as a weighted average of the collision probabilities in different contention zones during AIFS. Zhu and Chlamtac [146] have proposed a Markov model of EDCA to calculate the saturation throughput and access delay. Kong *et al.* [70] have analyzed the AIFS and CW schemes using a 3D Markov chain where the third dimension indicates the remaining time before activating the backoff counter. Huang and Liao [55] have analyzed the performance of saturation throughput and access delay of EDCA. Hwang *et al.* [57] have presented an analytical model for EDCA with consideration of virtual collision, but the model was limited to the case that the difference between the minimum and maximum AIFS is one time slot only.

As an important QoS scheme specified in EDCA, the TXOP scheme has also attracted many research efforts, which were mainly focused on the analysis of the saturation performance [78, 125, 134]. Specifically, Tinnirello and Choi [125] have compared the saturation throughput of the TXOP scheme coupled by different ACK policies. They have further discussed the use of TXOP for provisioning of the temporal fairness in multi-rate 802.11e WLANs [126]. Peng *et al.* [105] have evaluated the saturation throughput of different Access Categories (ACs) as a function of different TXOP limits. Li, Ni, and Xiao [78] have evaluated the saturation throughput of the TXOP scheme with the block ACK policy under noisy channel conditions. Xu, Sakurai, and Vu [134] have analyzed the MAC access delay of EDCA with the AIFS, CW, and TXOP schemes under saturated conditions.

Tickoo and Sikdar [122] have extended the G/G/1 discrete-time queueing model for DCF to analyze the CFB scheme, assuming an infinite capacity of the MAC buffer. For the analytical tractability and simplicity, most of these models for TXOP are based on the assumption that the network traffic is saturated or follows a Poisson arrival process, which fails to model the characteristics of heterogeneous traffic generated by multimedia applications.

The existing models for EDCA have been primarily derived under the assumption of saturated traffic conditions where all the stations are backlogged all the time [55-57, 70, 75, 78, 108, 109, 118, 120, 125, 131, 132, 134, 146]. Since the traffic loads in the practical network environments are mainly unsaturated [81, 144], the development of analytical models for EDCA under unsaturated traffic conditions has also attracted much research attention [3, 28, 36, 43, 52, 61, 67, 83, 85, 93, 96, 119, 122]. For example, Tantra *et al*. [119] have introduced a Markov model to derive the throughput and delay of the CW scheme in EDCA, assuming that each AC has a MAC buffer with the capacity of only one frame. Engelstad and Osterbo [36] have analyzed the end-to-end delay of EDCA with the AIFS and CW schemes through modelling each AC as an M/G/1 queue of infinite capacity. Liu and Niu [83] have employed an M/M/1 queueing model to analyze the EDCA with the AIFS and CW schemes. They assumed an infinite capacity of the MAC buffer.

The performance of the AIFS, CW and TXOP schemes specified in EDCA has been primarily studied separately under unsaturated traffic conditions [3, 28, 36, 42, 43, 52, 67, 83, 93, 96, 119, 122]. However, to the best of our knowledge, there has

been only one attempt [61] to analytically modelling the combination of these three QoS schemes in EDCA under unsaturated traffic conditions. Inan, Keceli, and Ayanoglu [61] have leveraged a 3D Markov chain to capture the functionality of these three QoS schemes of EDCA, where the third dimension of the Markov chain denotes the number of backlogged frames in the transmission queue of an AC. As a result, the complexity of the solution becomes very high as the size of the transmission queue augments. Moreover, their model assumed a constant probability of packet arrival per state, which is unable to capture the stochastic properties of packet arrivals.

2.5.3 Admission Control

Although EDCA provides service differentiation among various traffic classes, the QoS constraints of the real-time applications cannot be guaranteed. This problem becomes more serious when the wireless channel is overloaded. Admission control is an important mechanism to guarantee the user-perceived QoS in WLANs and thus has received significant research efforts [2, 7, 28, 29, 44, 74, 81, 96, 106, 113]. For instance, Pong and Moors [107] have proposed to adjust the sizes of CW of different stations to fulfill the goal of admission control. Their scheme was based on the analytical model of the DCF proposed in [3] and was limited to saturated traffic conditions. Chen *et al.* [28] have proposed two admission control schemes based on the average delay and the channel occupancy ratio. However, their model only

considered the CW differentiation in EDCA. Garroppo *et al.* [44] have presented an admission control approach using the mean channel occupancy time calculated by an EDCA model proposed in [36], which did not take the TXOP differentiation into account.

In WLANs, the Access Point (AP) (*i.e.*, service provider) and new users have to cope with a limited radio resource that imposes a conflict of interests. For instance, the AP wants to increase its utility by improving the channel utilization and accommodating more new users. On the other hand, new users want to maximize its own utility by achieving the highest QoS if possible. Since the two objectives are different and often conflict with each other, the AP and new users do not have the apparent incentive to cooperate. Therefore, the non-cooperative game theory [46] has been applied to solve the admission control problem in wireless networks from the perspectives of both the service provider and new users. For example, Kuo, Wu, and Chen [74] have used a non-cooperative game theoretical approach for admission control in WLANs where the performance measures required in the game theoretical approach were obtained from simulations. Lin *et al.* [80] have proposed an integrated admission control and rate control method for Code Division Multiple Access (CDMA) [80] wireless networks based on non-cooperative game theory. Niyato and Hossain [98] have proposed a game theoretical framework for bandwidth allocation and admission control in IEEE 802.16 broadband wireless networks, where the QoS performance metrics were calculated through a queueing model.

Cost-efficient analytical models with good accuracy and lower computation complexity compared to simulation experiments can be used in the admission control schemes which need the real-time calculation and estimation of the performance metrics. To the best of our knowledge, none of the existing studies has incorporated game theory and the performance metrics derived from the analytical model for admission control in the IEEE 802.11e WLANs.

Chapter 3

Modelling the TXOP Scheme under Poisson Traffic

3.1 Introduction

The Enhanced Distributed Channel Access (EDCA) specifies an innovative burst transmission scheme, namely Transmission Opportunity (TXOP), to provide the service differentiation. The TXOP scheme allows a station that gains the channel to transmit the frames available in its buffer consecutively provided that the duration of transmission does not exceed a specified TXOP limit [59]. This scheme can achieve service differentiation by assigning the different TXOP limits to various traffic classes and can improve the utilization of the scarce wireless bandwidth because the contention overhead is amortized by all the frames transmitted within a burst.

The weaknesses of the existing analytical models for the TXOP scheme reported in the current literature are threefold. Firstly, most existing models assumed that all the stations are identical in terms of traffic loads, and thus are unable to evaluate the effects of the TXOP scheme on QoS differentiation among stations with different priorities. Secondly, existing work has primarily focused on the analysis of system throughput and access delay, but did not consider other important QoS

performance metrics, such as end-to-end delay and energy consumption. Finally, existing analytical models for the TXOP scheme were mainly based on the unrealistic assumptions of the saturated working conditions, infinite buffer of transmission queues, or perfect wireless channels. To overcome these weaknesses, this chapter aims to develop cost-effective performance evaluation tools in order to obtain a thorough understanding of the performance of the TXOP scheme under more realistic working environments.

The rest of this chapter is organized as follows: Section 3.2 presents the analytical models for the TXOP scheme with different ACK policies and unsaturated traffic loads, unbalanced stations, and bursty error channels. Section 3.3 validates the presented analytical models through NS-2 simulation experiments and conducts performance evaluation of the TXOP scheme with these models. Section 3.4 summarizes this chapter.

3.2 Analytical Models

This section presents the fundamental methodology and components to develop the analytical models for the burst transmission schemes in WLANs under practical working conditions. The transmission queue at each station is modelled as a bulk service queueing system where the arrival traffic follows a Poisson process. The service time of the queueing system is defined as the time interval from the instant that a Head-of-Burst (HoB) frame starts contending for the channel to the instant

that either the data burst is acknowledged following successful transmission or the data burst is terminated due to transmission failure. The service time is calculated through modelling the backoff procedure of the burst transmission schemes under unsaturated conditions. In the following, the term time slot denotes the time interval between the starts of two consecutive decrements of the backoff counter, while the term physical time slot represents a fixed time interval (unit time) specified in the protocol [58].

3.2.1 Modelling the TXOP scheme with Different ACK Policies

The subsection presents the analytical model for the TXOP scheme with two different ACK policies, referred to as the CFB and BACK schemes, in WLANs under unsaturated traffic conditions.

3.2.1.1 Analysis of the Backoff Procedure and the Service Time

Note that only the HoB frame needs to contend for the channel. The collision probability experienced by a transmitted HoB frame is equal to the probability that at least one of the $(n-1)$ remaining stations transmits in the time slot. Thus, the collision probability, p, can be given by

$$p = 1-(1-\tau')^{n-1} \tag{3.1}$$

where τ' denotes the probability that a station transmits under unsaturated traffic conditions. Since a station can transmit only when there are pending frames in its transmission queue, the transmission probability, τ', can be written as

$$\tau' = (1 - P_0)\tau \tag{3.2}$$

where P_0 represents the probability that the transmission queue is empty and τ is the probability that the station transmits given that the transmission queue is non-empty. τ is given by [16]

$$\tau = \frac{2(1-2p)}{(1-2p)(W+1) + pW(1-(2p)^m)} \tag{3.3}$$

where m is the maximum number of backoff stages and W denotes the minimum contention window.

The service time consists of two components: channel access delay and burst transmission delay. The former is the time interval from the instant that the frame reaches to the head of its transmission queue until it wins the contention and is ready for transmission. The latter is defined as the time duration of successfully transmitting a burst. Let $E[S_i]$, $E[A]$ and $E[B_i]$ denote the means of the service time, channel access delay, and burst transmission delay, respectively, where i represents the number of frames transmitted in a burst.

Given that an HoB frame is successfully transmitted after experiencing r collisions $(r \geq 0)$, its channel access delay consists of the delay from r unsuccessful transmissions and delay from $(r+1)$ backoff stages. Therefore, the average channel access delay is given by

$$E[A] = \sum_{r=0}^{\infty} \left(rT_c + \sigma' \sum_{h=0}^{r} \frac{W_h - 1}{2} \right) p^r (1-p) \tag{3.4}$$

where $p^r(1-p)$ is the probability that the frame is successfully transmitted after experiencing r collisions, rT_c is the collision time that the frame experiences during its transmission attempts, $(W_h-1)/2$ is the mean value of the backoff counter generated in the *h*-th backoff stage, and σ' is the average length of a time slot.

Let P_{tr} be the probability that at least one station among the remaining $(n-1)$ stations transmits during the time slot when the tagged station is in the backoff status. P_{tr} can be written as

$$P_{tr} = 1-(1-\tau')^{n-1} \tag{3.5}$$

The probability, P_s, that there is a successful transmission among the remaining $(n-1)$ stations when the tagged station is in the backoff status, can be expressed as

$$P_s = (n-1)\tau'(1-\tau')^{n-2} \tag{3.6}$$

The average length of a time slot, σ', is obtained by considering the fact that the channel is idle with probability $(1-P_{tr})$, a successful transmission occurs with probability P_s, and a collision happens with probability $(P_{tr}-P_s)$. Thus, σ' is written as

$$\sigma' = (1-P_{tr})\sigma + P_s T_s + (P_{tr}-P_s)T_c \tag{3.7}$$

where σ is the duration of a physical time slot, T_s denotes the average time for the successful transmission of a burst, and T_c represents the average collision time, respectively.

Note that only the HoB frame can be involved in the collision using burst transmission mechanism, T_c is given by

$$T_c = T_L + T_H + T_{SIFS} + T_{ACK} + T_{DIFS} + 2\Delta \tag{3.8}$$

where T_L and T_H are the average time required for transmitting the frame payload and the frame header, respectively. T_{ACK} denotes the time to transmit an ACK frame and Δ is the propagation delay.

The average time for the successful transmission of a burst, T_s, can be given by

$$T_s = \frac{\sum_{i=1}^{K} E[B_i]L_i}{1 - P_0} \tag{3.9}$$

where K denotes the maximum number of frames that can be transmitted in a TXOP limit, the denominator $(1 - P_0)$ means that the occurrence of burst transmission is conditioned on the fact that there is at least one frame in the transmission queue, L_i $(1 \le i \le K)$ is the probability of having i frames in the burst, and $E[B_i]$ represents the burst transmission delay. $E[B_i]$ is dependent on the ACK policies (CFB or BACK) and the number of frames transmitted within a burst. $E[B_i]$ can be expressed as

$$E[B_i] = \begin{cases} T_{DIFS} + T_L + T_H + T_{SIFS} + T_{ACK} + 2\Delta, & i = 1 \\ T_{DIFS} + i(T_L + T_H + 2T_{SIFS} + T_{ACK} + 2\Delta) - T_{SIFS}, & i > 1, \text{CFB} \\ T_{DIFS} + T_{ACK} + i(T_L + T_H + T_{SIFS} + \Delta) + T_{BREQ} + 2T_{SIFS} + T_{BACK} + 3\Delta, & i > 1, \text{BACK} \end{cases} \tag{3.10}$$

where $BREQ$ and $BACK$ represent the transmission time of BlockReq and BlockAck frames, respectively.

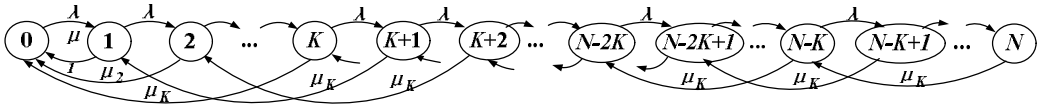


Fig. 3.1: The M/G[1,K]1//N queue state-transition-rate diagram.

3.2.1.2 Queueing Model and Performance Measures

The transmission queue at each station can be modelled as an M/G[1,K]/1/N queueing system [68] where the superscript $[1, K]$ denotes that the number of frames transmitted in a burst ranging from 1 to K, and N represents the buffer size at each station. The arrival traffic at each station follows a Poisson process with the rate λ (frames/second).

The server becomes busy when a frame reaches to the head of the transmission queue. The server becomes free after a burst of frames are acknowledged by the destination following successful transmission. The service time is dependent on the number of frames transmitted in a burst and can be characterized by an exponential distribution function with mean $E[S_i]$. Thus, the service rate, μ_i, is given by $1/E[S_i]$.

Figure 3.1 illustrates the state-transition-rate diagram of the queuing system where each state denotes the number of frames in the system. The transition rate from state v to state $v+1$ $(0 \le v \le N-1)$ is the arrival rate λ of the Poisson process. A transition out of state v to state $v-K$ $(K \le v \le N)$ implies that the burst transmission of K frames completes and the transition rate is μ_K. The change from state v to state 0 $(1 \le v \le K-1)$ denotes that all v frames in the system are transmitted within a burst and the transition rate is μ_v. The transition rate matrix, $\mathbf{G}$, of the Markov chain can be obtained by the state-transition-rate diagram.

$$\mathbf{G} = \begin{bmatrix} -\lambda & \lambda & & & & & & & & & \\ \mu_1 & -\lambda-\mu_1 & \lambda & & & & & & & & \\ \mu_2 & & -\lambda-\mu_2 & \lambda & & & & & & & \\ \vdots & & & \ddots & \ddots & & & & & & \\ \mu_{K-1} & & & & -\lambda-\mu_{K-1} & \lambda & & & & & \\ \mu_K & & & & & -\lambda-\mu_K & \lambda & & & & \\ & \ddots & & & & & \ddots & \ddots & & & \\ & & \mu_K & & & & & -\lambda-\mu_K & \lambda & & \\ & & & \mu_K & & & & & -\mu_K & & \end{bmatrix} \tag{3.11}$$

The steady-state probability vector, $\mathbf{P} = (P_v, v = 0, 1, ..., N)$ of the Markov chain satisfies the following equations: $\mathbf{PG} = 0$ and $\mathbf{Pe} = 1$, where $\mathbf{e}$ is a unit column vector. Solving these equations [39] yields the steady-state vector $\mathbf{P}$ and then we can express L_i as $L_i = P_i, 1 \le i < K$ and $L_i = \sum_{v=K}^{N} P_v, i = K$.

The end-to-end delay is the time duration from the instant that a frame enters the transmission queue of the station to the instant the frame is acknowledged after

successful transmission. It is equivalent to the queueing delay plus service time. By virtue of Little's Law [68], the end-to-end delay, $E[D]$, can be given by

$$E[D] = \frac{E[N]}{\lambda(1 - p_b)} \tag{3.12}$$

where $E[N] = \sum_{v=0}^{N} vP_v$ is the average number of frames in the queueing system. $\lambda(1 - p_b)$ is the effective rate of the traffic entering into the transmission queue since the arriving frames are discarded if the finite buffer becomes full.

The frame loss probability, p_b, which is the probability that an arriving frame finds the finite buffer full, is equivalent to P_N. Given the loss probability, p_b, the normalized system throughput, TH, can be computed by

$$TH = \frac{n\lambda E[P](1 - p_b)}{C} \tag{3.13}$$

where n denotes the number of stations, $E[P]$ represents the average frame payload length, and C is the channel data rate.

3.2.2 Modelling the TXOP scheme with Unbalanced Stations

In this subsection, we present an analytical model for evaluating the QoS metrics of the TXOP scheme in a WLAN consisting of unbalanced stations with different traffic loads. Without loss of generality, we consider a scenario of N classes of stations where Class i $(i = 1, 2, \cdots, N)$ has n_i stations. The transmission queue at each station in Class i is modelled as a bulk service

queueing system where the traffic follows a Poisson arrival process with the rate λ_i (frames/second).

3.2.2.1 The Backoff Procedure

As in [122, 140, 144], we use a simplification in our model that a weighted version of the saturated transmission probability can approximate the unsaturated transmission probability. The weight is the steady-state probability of a non-empty buffer. The saturation transmission probability, τ_i', is given by [130]

$$\tau_i' = \frac{2(1-p_i)(1-p_i^{m+1})}{2W(1-p_i-p_i(2p_i)^{m'}+2^{m'}p_i^{m+1}(2p_i-1))+(1-p_i)(1-p_i^{m+1})} \quad (3.14)$$

where m is the maximum number of backoff stages (*i.e.*, the retry limit), m' represents the maximum number of times the contention window can be doubled, and W is the minimum contention window size.

The collision probability experienced by a transmitted HoB frame, p_i, is equal to the probability that at least one of the remaining stations transmits in the considered time slot. Therefore, p_i is given by

$$p_i = 1-(1-\tau_i)^{n_i-1}\prod_{r\neq i}(1-\tau_r)^{n_r} \quad (3.15)$$

Since that a station can only transmit when there is a frame in its transmission queue, the transmission probability under unsaturated traffic loads, τ_i, can be obtained by weighting the saturation transmission probability with the probability of the non-empty transmission queue.

$$\tau_i = (1 - P_{0i})\tau_i' \tag{3.16}$$

where P_{0i} is the probability that there is no frame in the transmission queue of a station in Class i. τ_i' denotes the probability that the station transmits in a randomly chosen time slot given that its transmission queue is non-empty. From Eqs. (3.15) and (3.16), the collision probability, p_i, can be rewritten as

$$p_i = 1 - (1 - (1 - P_{0i})\tau_i')^{n_i - 1} \prod_{r \neq i} (1 - (1 - P_{0r})\tau_r')^{n_r} \tag{3.17}$$

3.2.2.2 The Service Time

The service time consists of two parts: the channel access delay and burst transmission delay. The former is the time interval from the instant that an HoB frame reaches to the head of the transmission queue until it wins the contention and is ready for transmission, or until it is discarded due to transmission failures. The latter is the time duration of successfully transmitting a burst (note that it equals zero if the HoB frame is discarded). In the case that the HoB frame is successfully transmitted, we denote $E[S_{vi}]$, $E[A_i]$ and $E[B_v]$ as the means of the service time, channel access delay, and burst transmission delay, respectively, where v represents the number of frames transmitted within the burst and i denotes that the burst is transmitted from a station of Class i. Similarly, let $E'[S_i]$, $E'[A_i]$, and $E'[B]$ denote the means of the service time, channel access delay, and burst transmission delay, respectively, when the HoB frame is discarded due to transmission failures. First, $E[B_v]$ and $E'[B]$ can be expressed as

$$\begin{cases} E[B_v] = T_{DIFS} + v(T_L + T_H + 2T_{SIFS} + T_{ACK} + 2\Delta) - T_{SIFS} \\ E'[B] = 0 \end{cases} \tag{3.18}$$

where T_L and T_H denote the transmission time for the frame payload and frame head, respectively. T_{ACK} accounts for the time to transmit an ACK frame and Δ is the propagation delay.

Given that an HoB frame is successfully transmitted after experiencing j collisions $(j \geq 0)$, its channel access delay has two components: delay from j unsuccessful transmissions and delay from $(j+1)$ backoff stages. Thus, the average channel access delay, $E[A_i]$ and $E'[A_i]$, are given by

$$\begin{cases} E[A_i] = T_c \varphi_i + \overline{\sigma}_i \delta_i \\ E'[A_i] = T_c \varphi_i' + \overline{\sigma}_i \delta_i' \end{cases} \tag{3.19}$$

where T_c is the average collision time and $\overline{\sigma}_i$ is the average length of a time slot. φ_i accounts for the average number of collisions before a successful transmission from the station in Class i while φ_i' denotes the number of collisions when the HoB frame is discarded due to transmission failures. φ_i and φ_i' are given by

$$\begin{cases} \varphi_i = \sum_{j=0}^{m} \dfrac{j p_i^j (1 - p_i)}{(1 - p_i^{m+1})} \\ \varphi_i' = m + 1 \end{cases} \tag{3.20}$$

δ_i denotes the average number of backoff counter decrements before a successful transmission from the station in Class i while δ_i' represents the average number of backoff counter decrements before the HoB frame is discarded due to transmission failures. δ_i and δ_i' can be expressed as

$$\begin{cases} \delta_i = \sum_{j=0}^{m} \sum_{h=0}^{j} \frac{W_h - 1}{2} \frac{p_i^j (1 - p_i)}{(1 - p_i^{m+1})} \\ \delta_i' = \sum_{j=0}^{m} \frac{W_j - 1}{2} \end{cases} \tag{3.21}$$

where p_i^j is the probability that the HoB frame experiences j collisions and $(W_h - 1)/2$ denotes the mean of the backoff counters generated in the h-th backoff stage. W_h is given by

$$W_h = \begin{cases} 2^h W & 0 \le h \le m' \\ 2^{m'} W & m' < h \le m \end{cases} \tag{3.22}$$

where W is the minimum contention window size.

Let P_{t_i} be the probability that at least one station among the remaining stations transmits in a considered time slot, given that the station in Class i is in the backoff procedure. P_{t_i} can be written as

$$P_{t_i} = 1 - (1 - \tau_i)^{n_i - 1} \prod_{r \ne i} (1 - \tau_r)^{n_r} \tag{3.23}$$

The probability, $P_{s_i}^r$, that a station in Class r successfully transmits among the remaining stations, given the station in Class i being in the backoff procedure, can be expressed as

$$P_{s_i}^r = \tau_r (1 - \tau_r)^{n_r - 1} (1 - \tau_i)^{-1} \prod_{k \ne r} (1 - \tau_k)^{n_k} \tag{3.24}$$

The average length of a time slot, $\bar{\sigma}_i$, is obtained by considering the fact that the channel is idle with probability $(1 - P_{t_i})$, a successful transmission occurs with

probability $\sum_{r=1}^{N} P_{s_i}^r$, and a collision happens with probability $(P_{t_i} - \sum_{r=1}^{N} P_{s_i}^r)$. Thus, $\bar{\sigma}_i$ can be written as

$$\bar{\sigma}_i = (1 - P_{t_i})\sigma + \sum_{r=1}^{N} P_{s_i}^r T_s^r + (P_{t_i} - \sum_{r=1}^{N} P_{s_i}^r) T_c \tag{3.25}$$

where σ is the length of a physical time slot, T_s^r is the average time duration for the successful transmission of a burst from the station in Class r, and T_c is the average collision time, respectively.

Note that only the HoB frame can be involved in the collision using the TXOP scheme, T_c is given by

$$T_c = T_L + T_H + T_{SIFS} + T_{ACK} + T_{DIFS} + 2\Delta \tag{3.26}$$

The average time duration, T_s^r, for the successful transmission of a burst from the station in Class r, is given by

$$T_s^r = \frac{\sum_{v=1}^{F_r} E[B_v] L_{vr}}{1 - P_{0r}} \tag{3.27}$$

where F_r denotes the maximum number of frames that can be transmitted in a TXOP limit of the station in Class r, the denominator $(1 - P_{0r})$ means that the occurrence of burst transmission is conditioned on the fact that there is at least one frame in the transmission queue of the station, L_{vr} $(1 \le v \le F_r)$ is the probability of having v frames within the burst transmitted from the station, and $E[B_v]$ is the burst transmission delay given in Eq. (3.18).

The average number, d_r, of frames within a burst transmitted from the station in Class r can be derived by

$$d_r = \frac{T_s^r + T_{SIFS} - T_{DIFS}}{T_L + T_H + 2T_{SIFS} + T_{ACK} + 2\Delta} \tag{3.28}$$

We can also derive $E[S_r]$, the average time duration for serving a burst transmitted from a station in Class r, as follows

$$E[S_r] = E[A_r] + T_s^r \tag{3.29}$$

3.2.2.3 Queueing Model

The transmission queue at the station in Class r can be modelled as an M/G$^{[1,F_r]}$/1/K queueing system [68] where the superscript $[1, F_r]$ denotes that the number of frames transmitted in a burst ranges from 1 to F_r, and K_r represents the buffer size at the Class r station.

The server becomes busy when a frame reaches to the head of the transmission queue. The server becomes free after a burst of frames are acknowledged by the destination following successful transmission, or after the HoB frame is dropped due to transmission failures. The service time is dependent on the number of frames transmitted within a burst and the class of the transmitting station. Thus, the service time of a burst with v $(1 \le v \le F_r)$ frames successfully transmitted from the Class r station can be modelled by an exponential distribution function with mean $E[S_{vr}]$, then the mean service rate, μ_{vr}, is given by $1/E[S_{vr}]$. On the other hand, when the HoB frame is discarded due to transmission failures, the service time can also be modelled by an exponential distribution with mean $E'[S_r]$, then the mean service rate, μ_r', is given by $1/E'[S_r]$.

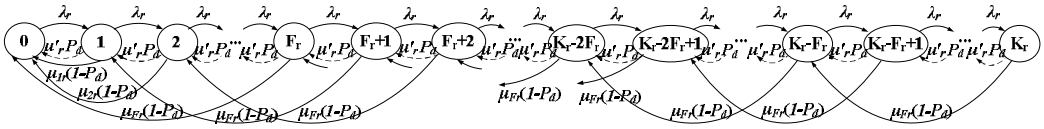


Fig. 3.2: The M/G$^{[1, Fr]}$/1/K$_r$ queue state-transition-rate diagram.

Fig. 3.2 illustrates the state-transition-rate diagram of the queuing system of the Class r station where each state denotes the number of frames in the queueing system. The transition rate from state s to state $s+1$ $(0 \le s \le K_r - 1)$ is the arrival rate λ_r of the Poisson traffic. A transition from state s to state $s-1$ $(1 \le s \le K_r)$ implies that the HoB frame is dropped due to transmission failures with probability P_d, which is given by p_r^{m+1}, and the transition rate is $\mu_r' P_d$. Thus, a transition out of state s to state $s - F_r$ $(F_r \le s \le K_r)$ represents that the burst transmission of F_r frames completes and the transition rate is $\mu_{F_r}(1-P_d)$. The change from state s to state 0 $(1 \le s \le F_r - 1)$ denotes that all s frames in the queueing system are transmitted within a burst and the transition rate is $\mu_{sr}(1-P_d)$.

We can obtain the generator matrix, $\mathbf{G}_r$, of the Markov chain in Fig. 3.2. The steady-state probability vector, $\mathbf{P}_r = (P_{sr}, s = 0, 1, \ldots, K_r)$ of the Markov chain satisfies the following equations

$$\mathbf{P}_r \mathbf{G}_r = 0 \quad \text{and} \quad \mathbf{P}_r \mathbf{e} = 1 \tag{3.30}$$

Solving these equations yields the steady-state vector as [39]

$$\mathbf{P}_r = \mathbf{u}(\mathbf{I} - \Re_r + \mathbf{e}\mathbf{u})^{-1} \tag{3.31}$$

where $\Re_r = \mathbf{I} + \mathbf{G}_r / \min\{\mathbf{G}_r(\rho,\rho)\}$, $\mathbf{u}$ is an arbitrary row vector of $\Re_r$, $\mathbf{e}$ is a unit column vector, and $\mathbf{I}$ denotes the unit matrix. Therefore, we can obtain the probability, P_{0r}, that the transmission queue of station in Class r is empty. Afterwards, the probability that v $(1 \le v \le F_r)$ frames are transmitted from the station in Class r within a TXOP limit, L_{vr}, can be given by

$$\begin{cases} L_{vr} = P_{vr}, & 1 \le v < F_r \\ L_{vr} = \sum_{s=F_r}^{K_r} P_{sr}, & v = F_r \end{cases} \tag{3.32}$$

3.2.2.4 Performance Metrics: Throughput, End-to-End Delay, and Frame Dropping Probability

The frame dropping probability, p_{br}, which is the probability that an arriving frame finds the finite buffer at the station of Class r being full, is equivalent to P_{Kr}. With the frame dropping probability, the throughput TH_r of the station in Class r can be calculated as

$$TH_r = \lambda_r E[P](1 - p_{br})(1 - P_d) \tag{3.33}$$

where $E[P]$ is the frame payload length and P_d is the probability that the frame is dropped due to transmission failures. The system throughput is also the aggregate throughput of all the stations in the WLAN, $\sum_{r=1}^{N} n_r TH_r$.

The end-to-end delay is the time duration from the instant that a frame enters the transmission queue of the station to the instant that the frame leaves the queueing system. It consists of queueing delay and service time. By virtue of Little's Law [13], the end-to-end delay, $E[D_r]$, is given by

$$E[D_r] = \frac{E[N_r]}{\lambda_r (1 - p_{br})} \tag{3.34}$$

where $E[N_r] = \sum_{s=1}^{K_r} sP_{sr}$ is the average number of frames in the queueing system of the station in Class r. $\lambda_r (1 - p_{br})$ is the effective rate of the traffic entering into the transmission queue of the station since the arriving frames are discarded if the finite buffer becomes full.

3.2.2.5 Performance Metrics: Energy Consumption

Many research efforts have been devoted to modelling of energy consumption [14, 64, 137] of wireless networks with the aim of gaining insights into the issues on how to improve the energy efficiency and prolong the battery life of wireless stations. Moreover, 802.11-based MAC protocols with some improved functions have been widely used in wireless ad hoc networks and sensor networks [27, 101, 142] where the energy consumption is a very important issue. Therefore, the study of energy consumption of the TXOP scheme is useful for the MAC design in these networks. This subsection analyzes the energy consumption per successful frame transmission in WLAN with unbalanced stations incorporated with the TXOP scheme.

Let e_{tx}, e_{rx}, e_{ov} and e_{id} denote the power required for transmitting, receiving, overhearing and being idle, respectively. The overall energy consumed by the station in Class r to successfully transmit a frame, E_r, can be decomposed into four components and given by

$$E_r = \frac{E_r^{su} + E_r^{co} + E_r^{bf} + E_r^{em}}{d_r} \tag{3.35}$$

where E_r^{su} is the energy consumption for a successful burst transmission from the station, E_r^{co} is the energy wasted in collisions before the successful transmission, E_r^{bf} is the energy spent in the backoff stages, E_r^{em} is the energy consumed when the station has no pending frame between any two consecutive transmission bursts, and d_r is the average number of frames within a burst transmitted from the station, which is given in Eq. (3.28). In what follows, these components of the energy consumption respectively are calculated.

First, the energy consumed by the station to successfully transmit a burst, E_r^{su}, can be expressed as

$$E_r^{su} = e_{tx}(T_L + T_H)d_r + e_{rx}T_{ACK}d_r + e_{id}T_{SIFS}(2d_r - 1) \tag{3.36}$$

On average, the station incurs collisions for $(\varphi_r(1-P_d) + \varphi_r' P_d)$ times before successfully transmitting a burst. Therefore, the energy consumed during the collisions, E_{co}, can be given by

$$E_r^{co} = (e_{tx}(T_L + T_H) + e_{id}(T_{SIFS} + T_{ACK} + T_{DIFS}))(\varphi_r(1-P_d) + \varphi_r' P_d) \tag{3.37}$$

As aforementioned, the time slot perceived by a station that is currently in the backoff state is considered here. In the case that the channel is sensed idle, backoff counters are decremented by one per physical time slot σ. If the channel is sensed busy either due to collisions or successful transmissions from the remaining stations,

the backoff counter is halted. Hence, with Eq. (3.25), the energy spent during the backoff process can be expressed as

$$E_r^{bf} = (e_{id}\sigma + e_{ov}(\sum_{i=1}^{N} P_{s_r}^i T_s^i + (P_{t_r} - \sum_{i=1}^{N} P_{s_r}^i)T_c))(\delta_r(1-P_d) + \delta_r' P_d) \quad (3.38)$$

where δ_r and δ_r', given in Eq. (3.21), are the average numbers of time slots that the station defers in backoff stages, given that the HoB frame is successfully transmitted or discarded, respectively.

To obtain E_r^{em}, the energy consumed when the station has no frame backlogged between any two consecutive transmission bursts, we have to derive the average time duration of the transmission queue being empty, T_r^{em}, which can be calculated as the following

$$T_r^{em} = E[S_r]\frac{P_{0r}}{1-P_{0r}} \quad (3.39)$$

where $E[S_r]$, given in Eq. (3.29), is the average service time of a burst transmitted from the station. P_{0r} is the probability that the transmission queue of the station is empty. Therefore, conditioned on the fact that the channel is idle or not, E_r^{em} can be expressed as

$$E_r^{em} = T_r^{em}(e_{id}(1-P_{t_r}) + e_{ov}P_{t_r}) \quad (3.40)$$

where P_{t_r}, given in Eq. (3.23), is the probability that at least one of the remaining stations transmits, given the station in Class r being in the backoff state.

3.2.3 Modelling the TXOP scheme with Bursty Error Channels

We model the TXOP scheme in WLANs under bursty error channels. During the service process of a burst, the channel may be in one of the three cases: good state, bad state, and mixed state (*i.e.*, the channel state varies during the service process). Therefore, the mean service time in these three cases are calculated, respectively.

3.2.3.1 Analysis of the Backoff Procedure

Let τ be the probability that the station transmits a frame in a randomly chosen time slot, given that its transmission queue is non-empty. Based on the Markov model in [130], τ is given by

$$\tau = \frac{2(1-p)}{1-p^{m+1}+W\left(\sum_{i=0}^{m'}(2p)^i(1-p)+p(2p)^{m'}(1-p^{m-m'})\right)} \tag{3.41}$$

where W is the minimum contention window, m' represents the maximum backoff stage, and m denotes the retry limit.

Let p_c denote the probability that a transmitted frame from a station encounters a collision. It is equal to the probability that at least one of the remaining stations transmits in the considered time slot and is given by

$$p_c = 1-(1-\tau')^{n-1} \tag{3.42}$$

where τ' is transmission probability given by $\tau' = (1-P_0)\tau$, where P_0 is the probability that the transmission queue of the station is empty.

Let p_e denote the transmission error probability. Since the transmission error may occur due to the corruption of DATA frame or ACK frame, p_e is given by

$$p_e = p_{ed} + (1 - p_{ed}) p_{ea} \tag{3.43}$$

where p_{ed} and p_{ea} are the probabilities of DATA frame corruption and ACK frame corruption, respectively. Given a channel being in good (or bad) state at time *t*, it will remain in this state until time $t + T$ with the probability $e^{-\nu_g T}$ (or $e^{-\nu_b T}$). Therefore, p_{ed} and p_{ea} can be expressed as

$$p_{ed} = 1 - \pi_g e^{-\nu_g (T_{DATA} + \delta)} \tag{3.44}$$

$$p_{ea} = (1 - e^{-\nu_g T_{SIFS}}) + e^{-\nu_g T_{SIFS}} (1 - e^{-\nu_g (T_{ACK} + \delta)}) \tag{3.45}$$

where T_{DATA} and T_{ACK} are the time durations required for transmitting the data and ACK frame, respectively. π_g and π_b are the steady-state probabilities of the channel being in the good and bad state, respectively. δ is the propagation delay.

The transmission failure probability, p, under three cases is calculated. First, if the channel stays in the good state during the service process of a burst transmission, the failure probability, p, is equivalent to the collision probability p_c. Second, if the channel state varies during the service process of a burst, a frame transmission fails either due to collision, or transmission error. Thus, p is given by $p = p_c + (1 - p_c) p_e$. Finally, if the channel stays in the bad state during the service process of a burst, p is equal to 1 since any frame transmission will fail in this case. Note that in the third case, Eq. (3.41) reduces to the following one

$$\tau = \frac{(m+1)}{\sum_{s=0}^{m} \frac{(W_s - 1)}{2}} = \frac{2(m+1)}{m + 1 + W\left(2^{m'}(m - m' + 2) - 1\right)} \tag{3.46}$$

Let η, where $\eta \in \{G, B\}$, represents the channel state (*i.e.*, good (G) or bad (B) state). Let state $\alpha \in \{G, B\}$ and state $\beta \in \{G, B\}$ represent the channel states at the beginning and end of the service process, respectively. Furthermore, let the case $(\alpha, \beta) = (G, G)$ indicate that the channel stays in the good state during the service process, similarly, $(\alpha, \beta) = (B, B)$ denotes that the channel stays in the bad state during the service process and $(\alpha, \beta) = (G, B)$ or (B, G) stands for the fact that the channel state varies during the service process. Therefore, the transmission failure probability, p, can be expressed as

$$p = \begin{cases} p_c & (\alpha, \beta) = (G, G) \\ p_c + (1 - p_c) p_e & (\alpha, \beta) = (G, B) \text{ or } (B, G) \\ 1 & (\alpha, \beta) = (B, B) \end{cases} \tag{3.47}$$

3.2.3.2 Analysis of the Service Time

The service time consists of two parts: the channel access delay and burst transmission delay. The former is the time duration from the instant that an HoB frame reaches to the head of the transmission queue until it wins the contention and is ready for transmission, or until it reaches the retry limit and is to be discarded. The latter is the time duration of transmitting a burst. Let $E[S_{\alpha\beta}^{r}]$, $E[A_{\alpha\beta}]$ and $E[B_r]$ denote the means of the service time, channel access delay and burst transmission delay, respectively, where r represents the number of frames transmitted in a burst, α and β represent the channel state at the beginning and end of the service process, respectively. The mean channel access delay can be calculated as

$$E[A_{\alpha\beta}] = T_c \sum_{x=0}^{m} x p^x (1-p) + \sigma'_{\alpha\beta} \sum_{x=0}^{m} \frac{W_x - 1}{2} p^x (1-p) \tag{3.48}$$

where T_c is the average time for an unsuccessful transmission attempt, $\sigma'_{\alpha\beta}$ is the average length of a time slot perceived by the station, $p^x(1-p)$ is the probability that the frame is successfully transmitted after x failed attempts, and $(W_x - 1)/2$ is the average value of the backoff counter generated at the *x*-th backoff stage.

Let P_t be the probability that at least one station among the remaining $(n-1)$ stations transmits in a time slot. Let P_s denote the probability that only one of the remaining $(n-1)$ stations transmits in a time slot. P_t and P_s are given by

$$P_t = 1 - (1-\tau')^{n-1} \tag{3.49}$$

$$P_s = (n-1)\tau'(1-\tau')^{n-2} \tag{3.50}$$

The average length of a time slot perceived by the station, $\sigma'_{\alpha\beta}$, is obtained by considering the fact that the channel is idle with probability $(1-P_t)$, only one of the $(n-1)$ remaining stations transmits with probability P_s, and a collision happens with probability $(P_t - P_s)$.

$$\sigma'_{\alpha\beta} = (1-P_t)\sigma + P_s T_{\alpha\beta} + (P_t - P_s) T_c \tag{3.51}$$

where σ is the duration of a physical time slot, $T_{\alpha\beta}$ is the average time for the transmission of a burst, and T_c is the average time for an collision. T_c is given by

$$T_c = T_{DIFS} + T_{DATA} + T_{SIFS} + T_{ACK} + 2\delta \tag{3.52}$$

If the channel stays in the bad state during the service process, i.e., $(\alpha, \beta) = (B, B)$, only one frame can be transmitted in a burst, thus $T_{\alpha\beta}$ is equal to

T_c. In the other cases that the channel stays in the good state or varies during the service process, $T_{\alpha\beta}$ can be written as

$$T_{\alpha\beta} = \frac{\sum_{r=1}^{K} \overline{E[B_r]} L_r}{1 - P_0} \quad (\alpha, \beta) = (G,G) \text{ or } (G,B) \text{ or } (B,G) \tag{3.53}$$

where K denotes the maximum number of frames that can be transmitted in a TXOP limit of the station, the denominator $(1 - P_0)$ indicates that the occurrence of burst transmission is conditioned on the fact that there is at least one frame in the transmission queue of the station, L_r $(1 \le r \le K)$ is the probability of having at most r frames within the burst transmitted from the station, and $\overline{E[B_r]}$ is the average transmission time of a burst that has at most r frames.

Since a burst will be terminated if the transmission of any frame within the burst fails, $\overline{E[B_r]}$ can be calculated as

$$\overline{E[B_r]} = \sum_{i=2}^{r-1} E[B_i](1 - p_e)^i p_e + E[B_r](1 - p_e)^r + E[B_1]p_e \tag{3.54}$$

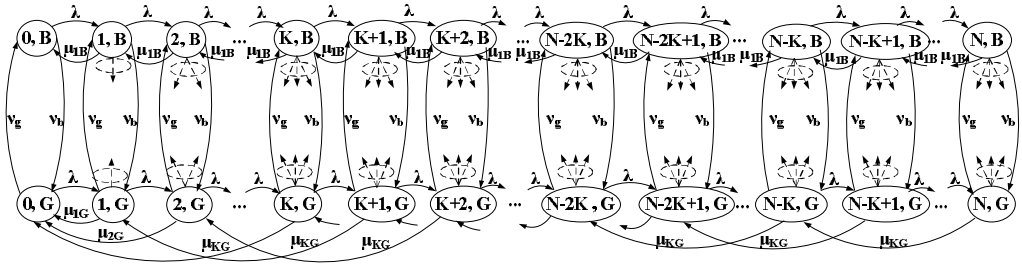
where $E[B_r]$ $(1 \le r \le K)$ is the burst transmission delay of r frames and can be given by

$$E[B_r] = T_{DIFS} + r(T_{DATA} + 2T_{SIFS} + T_{ACK} + 2\delta) - T_{SIFS} \tag{3.55}$$

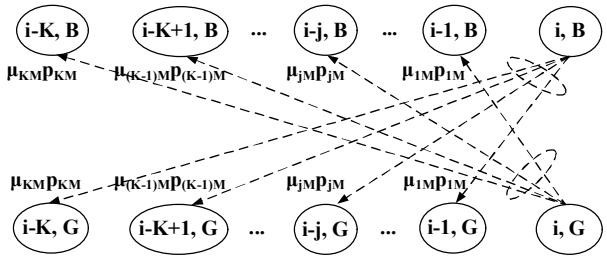
If the channel stays in the good state during the service process of a burst, *i.e.*, $(\alpha, \beta) = (G,G)$, $\overline{E[B_r]}$ reduces to $E[B_r]$ since p_e is equal to 0 in this case.

3.2.3.3 Queueing Analysis

The transmission queue at each station can be modelled as an $M/G^{[1,\,K]}/1/N$ queueing system where the superscript $[1, K]$ denotes that the number of frames transmitted in a burst ranges from 1 to K, and N represents the capacity of the queueing system.



(a) Two-state CTMC



(b) sub-Markov chain

Fig. 3.3: Markov chain model for the $M/G^{[1,\,K]}/1/N$ queueing system.

The queueing system can be characterized by a bi-dimensional CTMC as shown in Fig. 3.3, with a dashed line circle shown in detail in the sub-Markov chain below. State (s,η), where $(s = 0,1,\ldots,N)$ and $\eta \in \{G,B\}$, represents the case where there are s frames in the queueing system and the channel state is η. The service rate of the queueing system is dependent on the number of frames transmitted in a burst and the channel state η. Let $\mu_{\alpha\beta}^{r}$ be the service rate from state $\alpha \in \{G,B\}$ to state $\beta \in \{G,B\}$, with r frames transmitted in the burst. $\mu_{\alpha\beta}^{r}$ can be given by the inverse of the average service time, $E[S_{\alpha\beta}^{r}]$, which is calculated based on the

channel state. For the sake of illustration, in Fig. 3.3, μ_{rG}, μ_{rB}, and μ_{rM} are used to replace the notations of $\mu_{GG}^{r}, \mu_{BB}^{r}$, and μ_{GB}^{r} (or μ_{BG}^{r}), respectively. Note that if $(\alpha, \beta) = (G, B)$ or (B, G), the burst transmission may be terminated due to channel errors. Therefore, the transition rate from state (s, η) to state $(s - r, \eta')$, where $(\eta \neq \eta')$, is obtained by multiplying the service rate of a burst with r frames, μ_{rM}, with the probability that there are r frames transmitted during the burst transmission, p_{rM}, which is given by

$$p_{rM} = \begin{cases} p_e(1-p_e)^{r-1} & r < \min(s, K) \\ (1-p_e)^{r-1} & r = \min(s, K) \end{cases} \tag{3.56}$$

The transition rate matrix, **G**, of the CTMC can be obtained by the state-transition-rate diagram in Fig. 3.3. The steady-state probability Matrix, $\mathbf{P} = (P_{s\eta}, s = 0, 1, \ldots, N, \ \eta = \{\mathrm{G}, \mathrm{B}\})$, of the Markov chain satisfies the following equations: $\mathbf{PG} = 0$ and $\mathbf{Pe} = 1$. After obtaining the expression of **P** using the Matrix solution method [39], L_r can be expressed as $L_r = P_{rG} + P_{rB}, 1 \le r < K$ and $L_r = \sum_{i=K}^{N} (P_{iG} + P_{iB}), r = K.$

The buffer overflow probability, p_b, is given by $(P_{NG} + P_{NB})$. The frame dropping probability, p_d, is calculated conditioned on the channel state as

$$p_d = \pi_g (e^{-\nu_g E[A_{GG}]} p_c^{m+1} + (1 - e^{-\nu_g E[A_{GG}]}) p^{m+1}) + \pi_b (e^{-\nu_b E[A_{BB}]} + (1 - e^{-\nu_b E[A_{BB}]}) p^{m+1}) \tag{3.57}$$

where π_g and π_b are the probabilities that the channel is in the good state and bad state at time t, respectively. $e^{-\nu_g E[A_{GG}]}$ and $e^{-\nu_b E[A_{BB}]}$ are the probabilities

that the channel stays in the good state and bad state during the service process of the frame, respectively. p_c^{m+1} and p^{m+1} are the probabilities that the frame is dropped due to $(m+1)$ transmission failures when the channel stays in the good state and when the channel varies between good and bad state during the service process, respectively. Obtaining p_b and p_d, the throughput can be calculated as

$$TH = \lambda E[P](1-p_b)(1-p_d) \tag{3.58}$$

where $E[P]$ is the average length of the frame payload.

Table 3.1: System Parameters for the Analysis of the TXOP Scheme under Poisson Traffic

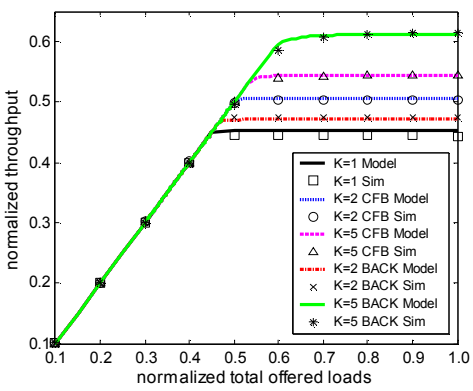
Frame payload	8000bits	PHY header	192bits
MAC header	224bits	ACK	112bits + PHY header
Channel data rate	11Mbit/s	BREQ	192bits + PHY header
Basic rate	1Mbit/s	BACK	1216bits + PHY header
Propagation delay	2μs	CW_{min}, CW_{max}	32, 1024
Slot time	20μs	Buffer size	50 frames
SIFS	10μs	DIFS	50μs

3.3 Model Validation and Performance Evaluation

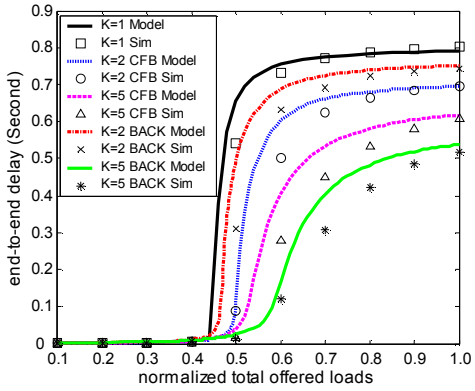
This section firstly validates the accuracy of the developed analytical models through NS-2 [99] simulation experiments. These models are then used to conduct performance evaluation of the TXOP scheme under various conditions.

NS-2 version 2.28 is used in the simulation experiments. The simulation run time is 600 seconds. The simulation results are collected after 10 seconds of warm up period. The scenario is a BSS where all stations are within the transmission range

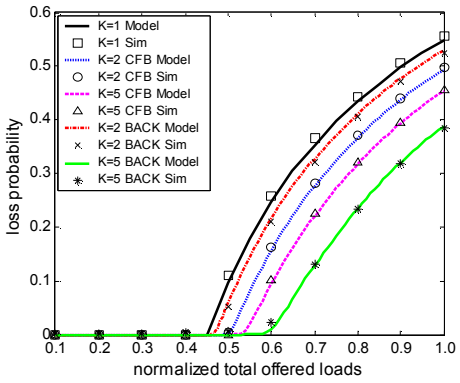
of each other. The traffic generated by each station follows the Poisson arrival process. The two-ray ground propagation model is used.



(a) Throughput



(b) End-to-end delay



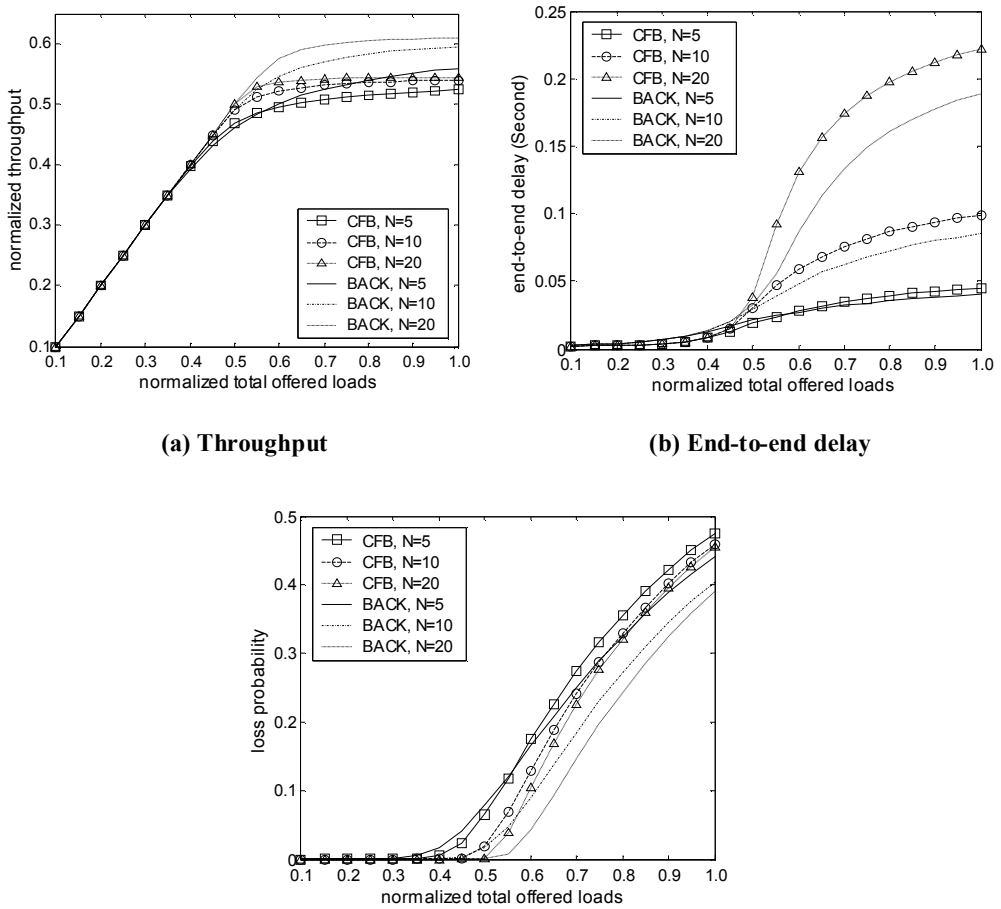
(c) Frame loss probability

Fig. 3.4: Performance measures versus normalized offered loads.

3.3.1 The TXOP scheme with Different ACK Policies

A WLAN with 10 stations located in a 150m × 150m rectangular area is considered. The system parameters follow the IEEE 802.11b standard [58] and are summarized in Table 3.1. Fig. 3.4 depicts the results of throughput, end-to-end delay, and frame loss probability of the CFB and BACK burst transmission schemes versus the normalized total traffic loads with different TXOP limits. Note that the TXOP limit $K = 1$ represents the legacy DCF. As shown in Fig. 3.4, the good degree of

agreement between the analytical and corresponding simulation results demonstrates that the analytical model possesses excellent accuracy for predicting the QoS performance metrics of throughput, end-to-end delay, and frame loss probability of the CFB and BACK schemes. The tractability and accuracy of the analytical model make it a practical and cost-effective performance evaluation tool for the burst transmission schemes in IEEE 802.11e WLANs.



(a) Throughput

(b) End-to-end delay

(c) Frame loss probability

Fig. 3.5: Analytical results versus normalized offered loads with different buffer sizes.

In the following, the developed model is adopted to investigate the performance of the burst transmission schemes under various system configurations and different

working conditions. The analytical results are obtained using the system parameters shown in Table I, unless otherwise specified.

3.3.1.1 Efficiency of the CFB and BACK Schemes

Figs. 3.4(a) and 3.4(b) reveal that both the CFB and BACK schemes can increase the throughput and reduce the end-to-end delay and loss probability over the legacy DCF when the WLAN works under the medium and high loads. As the TXOP limit increases, the network performance becomes better. Moreover, the CFB scheme outperforms the BACK scheme when the TXOP limit $K = 2$, while the BACK scheme is more efficient than the CFB scheme when $K = 5$. Specifically, the throughput of CFB is roughly 0.5 and of BACK is roughly 0.46 when $K = 2$. However, the throughput of CFB and BACK is roughly 0.53 and 0.6, respectively, when $K = 5$. Therefore, the results show that the BACK scheme is superior to CFB only when the TXOP limit goes beyond a given threshold. This threshold is obtained when the transmission delay of i frames in CFB scheme equals that in BACK scheme, so the threshold in the current scenario can be derived as 2.66 with Eq. (3.10), which means that the BACK scheme excels the CFB scheme when the TXOP limit is larger than 2.66.

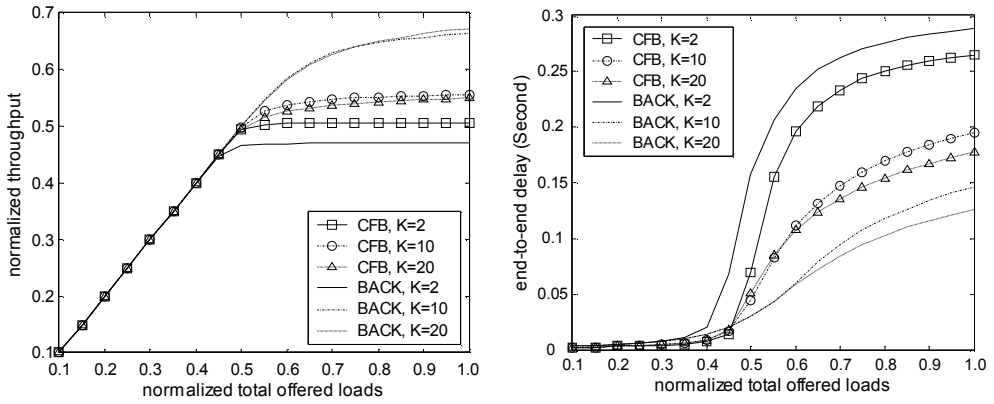
3.3.1.2 Effects of the Buffer Size

To investigate the impact of the buffer size on the performance of burst transmission schemes, Fig. 3.5 plots the throughput, end-to-end delay, and frame loss probability against the offered loads with a varying buffer size of 5, 10, and 20

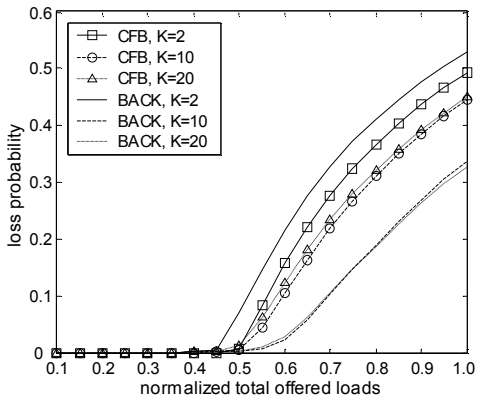
frames. The number of stations is set to 10 and the TXOP limit is fixed at 5 frames. Fig. 3.5 shows that the larger buffer achieves a higher throughput and lower loss probability while causing an increased delay. Moreover, when the buffer size increases from 10 to 20 frames, the improvement of throughput and loss probability is not significant, but there is a considerable grow in end-to-end delay. As throughput and loss probability are the most important performance metrics of delay-insensitive applications, it is desirable to set a large buffer size for these applications. However, for the delay-sensitive applications such as voice and video, a large buffer results to the high delay that may be intolerable for these inelastic applications. Thus, a small buffer is preferable for delay-sensitive applications.

3.3.1.3 Effects of the TXOP Limit

The effects of the TXOP limit on the network performance are further evaluated. The number of stations is set to 10 and the buffer size is fixed at 20 frames. Fig. 3.6 clearly shows that the throughput, end-to-end delay, and frame loss probability improve significantly as the TXOP limit increases from 2 to 10. However, the performance enhancement becomes trivial when the TXOP limit increases from 10 to 20. As setting a large TXOP limit leads to a high delay oscillation for stations, it is best to set the TXOP limit to 10 in this case. Again, it is shown that the BACK scheme outperforms the CFB scheme when the TXOP limit becomes larger.



(a) Throughput **(b) End-to-end delay**

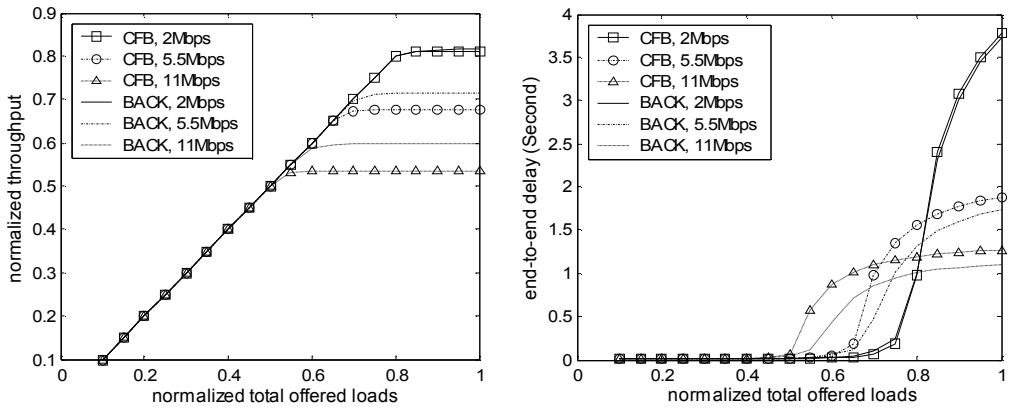


(c) Frame Loss probability

Fig. 3.6: Analytical results versus normalized offered loads with different TXOP limits.

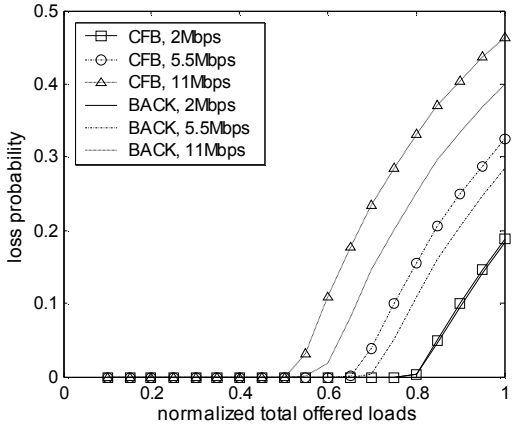
3.3.1.4 Effects of the Channel Date Rate

Fig. 3.7 shows the performance measures of the CFB and BACK schemes versus the offered loads with different channel data rates. The number of stations is set to 20. The buffer size and TXOP limit are fixed at 50 and 5 frames, respectively. Figs. 3.7(a) and (c) show that the throughput decreases and loss probability grows as the channel data rate increases. Because the ACK frames are transmitted at the basic rate while the data frames are transmitted at the channel date rate, the ACK overhead would lead to a



(a) Throughput

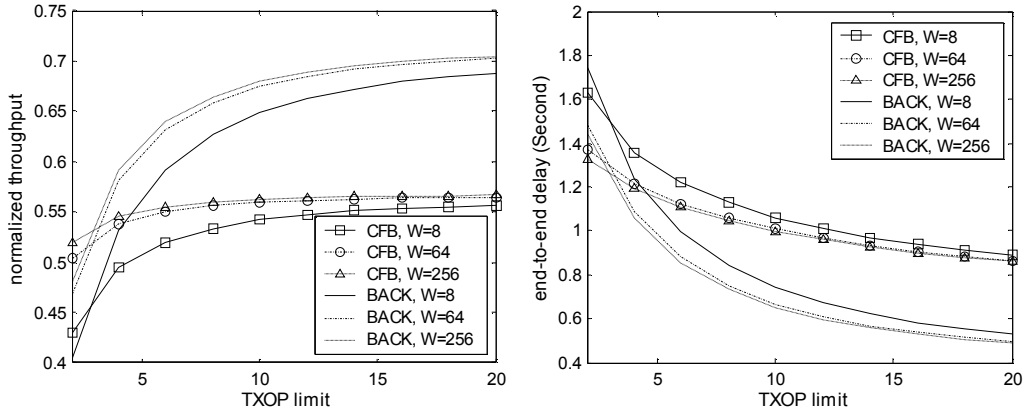
(b) End-to-end delay



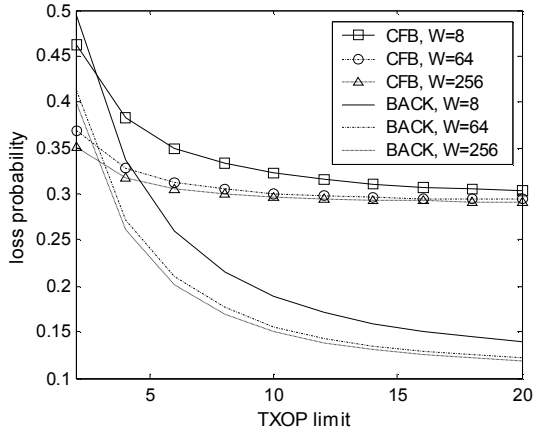
(c) Frame loss probability

Fig. 3.7: Analytical results versus normalized offered loads with different data rates.

lower channel utilization as the date rate increases. In Fig. 3.7(b), it is observed that the end-to-end delay decreases with the increasing channel data rate. On the other hand, it is shown that the throughputs of the CFB and BACK schemes are almost the same when the channel data rate is at 2 Mbps. However, the throughput of the BACK scheme is larger than that of the CFB scheme when the channel data rate is at 5.5 or 11 Mbps, which reveals that the BACK scheme is efficient with a high data rate.



(a) Throughput **(b) End-to-end delay**



(c) Frame loss probability

Fig. 3.8: Analytical results versus TXOP limit with different minimum contention windows.

3.3.1.5 Effects of the Contention Window

Fig. 3.8 evaluates the effects of the minimum contention window, W, on the performance of the CFB and BACK schemes with the varying TXOP limits. The number of stations is set to 20, the maximum number of backoff stages equals 5, and the normalized total offered loads are 0.8. It is observed that the network performance is significantly improved when W increases from 8 to 32, while it only introduces negligible improvement when W goes from 32 up to 256. Moreover, the performance improvement caused by the growing W becomes

insignificant as the TXOP limit increases. For instance, when W rises from 8 to 32, the enhancements of throughput for the CFB and BACK schemes are roughly 0.07 and 0.06, respectively, if the TXOP limit is 2. However, those figures drop to 0.01 and 0.02, respectively, if the TXOP limit rises to 16. Furthermore, it is noted that the threshold of TXOP limit beyond which the BACK scheme outperforms the CFB scheme is insensitive to W.

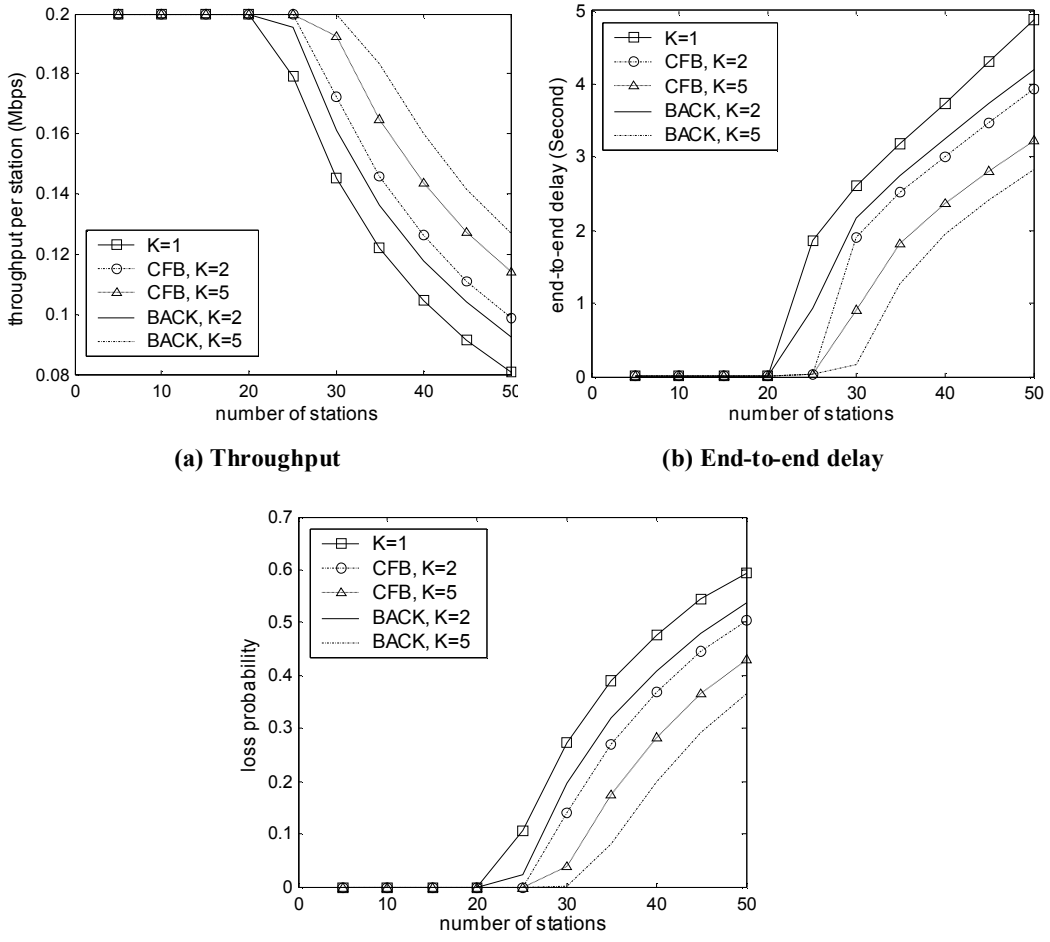


Fig. 3.9: Analytical results versus the number of stations with different TXOP limits.

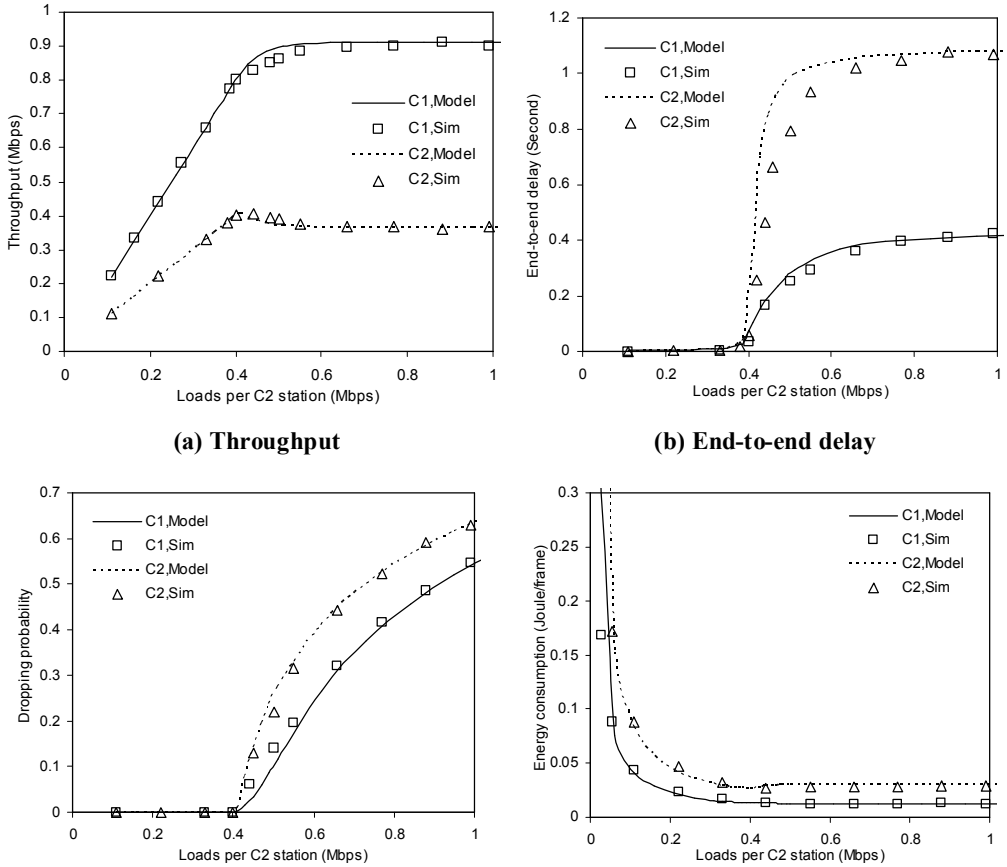
3.3.1.6 Effects of the Network Size

Fig. 3.9 plots the performance metrics of the burst transmission schemes as a function of the network size. The traffic rate and buffer size at each station are set to 0.2 Mbps and 50 frames, respectively. When the number of stations is small, the burst transmission has little impact on the performance since the network is under light loads. When the number of stations reaches 20, the network incorporated with the legacy DCF ($K = 1$) becomes saturated while that adopted the burst transmission schemes are still under unsaturated working conditions. Furthermore, the network can sustain more stations as the TXOP limit increases.

3.3.2 The TXOP scheme with Unbalanced Stations

A WLAN is considered where all the stations are located in a rectangular grid with dimension 150m × 150m and are within the transmission range of each other. This chapter presents the validation results of the proposed model with two typical scenarios. In Scenario 1, the stations are classified into two classes - Class 1 (C1) and Class 2 (C2) - where the traffic arrival rate of stations in C1 is double of that in C2. There are 4 C1 stations and 6 C2 stations, which are assigned with the TXOP limits of 5 and 2 (frames), respectively. In Scenario 2, the stations are classified into three classes - Class 1 (C1), Class 2 (C2), and Class 3 (C3) - where the traffic arrival rate of these stations follows a ratio of 3 : 2 : 1. There are 4 stations in each class and the TXOP limits of stations in Cl, C2 and C3 are 5, 3 and 1, respectively. The traffic generated by each station follows the Poisson arrival process. The power required for transmitting, receiving, overhearing and being idle is 1.65, 1.4, 1.4, and 1.15 W,

respectively. The retry limit is 10. The system parameters follow the IEEE 802.11b standard [58] and are summarized in Table 3.1.



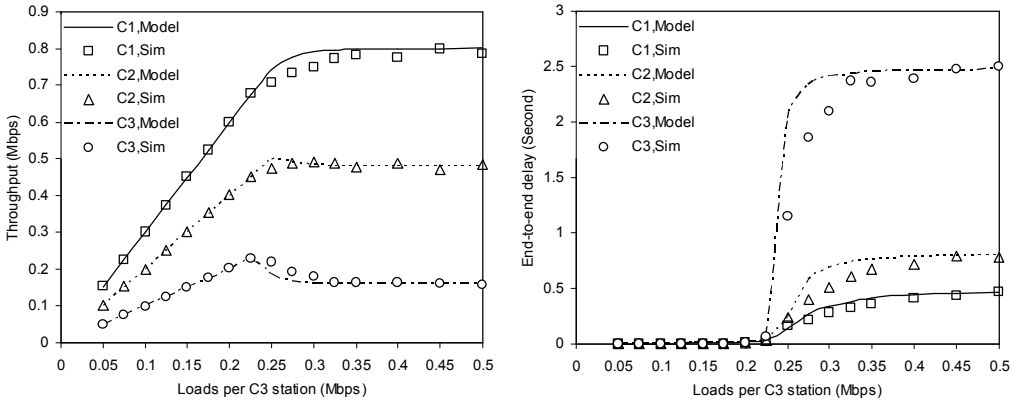
(a) Throughput **(b) End-to-end delay**

(c) Frame dropping probability **(d) Energy Consumption**

Fig. 3.10: Performance metrics versus offered loads per C2 station in scenario 1.

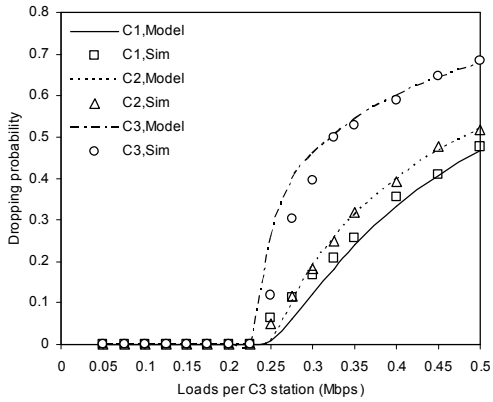
Figs. 3.10 and 3.11 plot the results of the throughput (Figs. 3.10(a), 3.11(a)), end-to-end delay (Figs. 3.10(b), 3.11(b)), frame dropping probability (Figs. 3.10(c), 3.11(c)), and energy consumption per successful transmission (Figs. 3.10(d), 3.11(d)), respectively, versus the traffic loads. As shown in Figs. 3.10 and 3.11, the good degree of agreement between the analytical results and the simulation experiments in both scenarios demonstrates the accuracy and capability of the model for evaluating the QoS performance of the TXOP scheme. Moreover, it is shown that

the throughput and end-to-end delay first increase then stabilize as the traffic loads grow while the frame dropping probability keep rising with the loads. It is also worth mentioning that the energy consumption decreases with the traffic loads because much energy is wasted in the idle period when the loads are low.

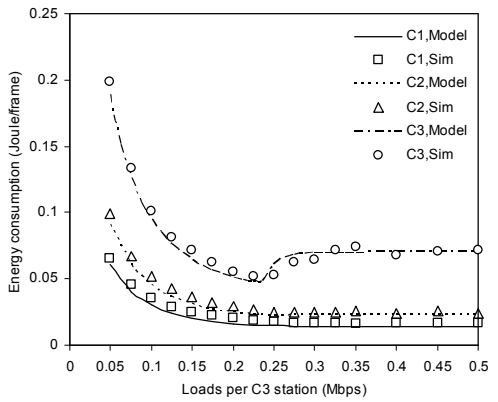


(a) Throughput

(b) End-to-end delay

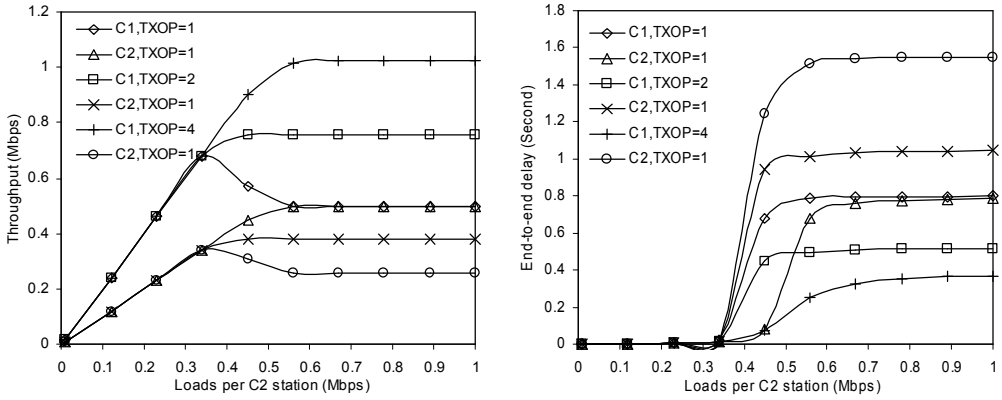


(c) Frame dropping probability



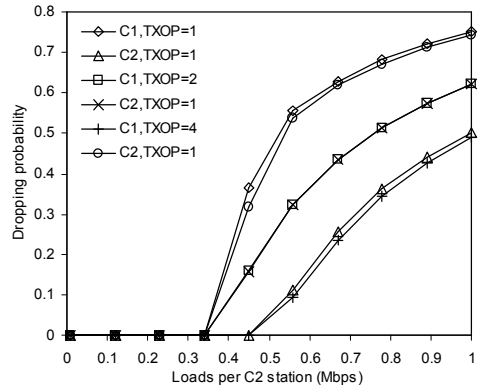
(d) Energy Consumption

Fig. 3.11: Performance metrics versus offered loads per C3 station in scenario 2.

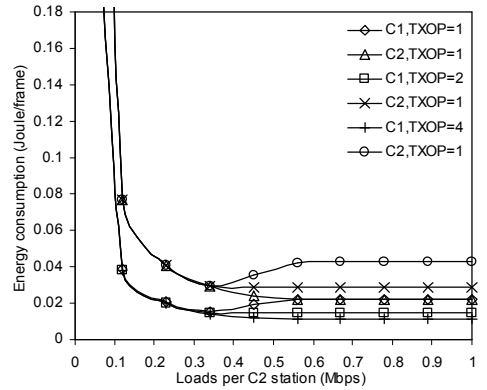


(a) Throughput

(b) End-to-end delay



(c) Frame dropping probability



(d) Energy Consumption

Fig. 3.12: Analytical results versus offered loads per C2 station (TXOP limit of C2 station = 1).

After the validation of the developed analytical model, this model will be used as a performance evaluation tool to investigate the effects of the TXOP scheme and analyze the impact of the buffer size and the number of stations.

3.3.2.1 Efficiency of the TXOP Differentiation Scheme

This study aims to show that the TXOP scheme can be used as an efficient QoS differentiation solution. For the stations with different traffic arrival rates, if they are configured with the same backoff parameters and TXOP limits because they belong to the same traffic category, the stations with high traffic rates will experience more

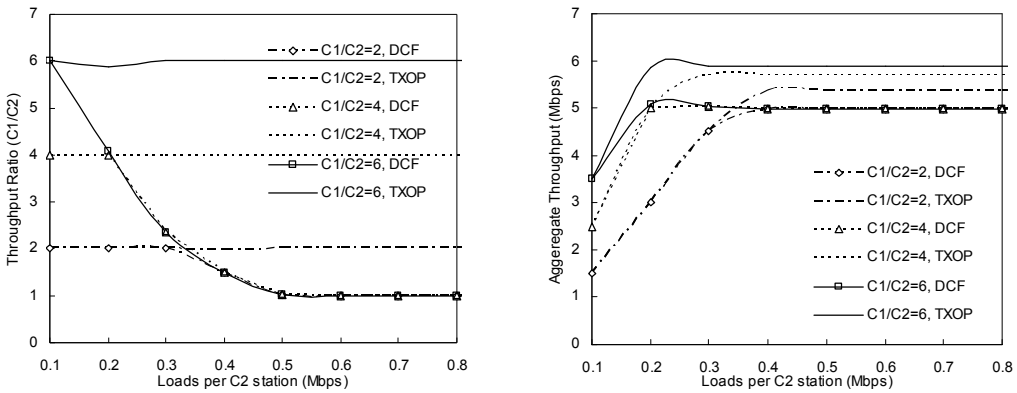
frame losses and larger delay than those with low traffic rates. However, if the TXOP limits of stations are set proportional to their traffic arrival rates, which means that the stations with higher rates can send more traffic each time they gain the channel, the stations belonging to the same traffic category can be provided similar loss rate although the network loads are unbalanced between them.

In Fig. 3.12, the TXOP limits of C1 and C2 stations are differentiated to investigate their effects on the performance of the WLAN with unbalanced stations. The TXOP limit of C2 stations is fixed at 1 and that of C1 stations varies from 1, 2, to 4 (frames). The traffic arrival rate of stations in C1 is double of that in C2. For the sake of clarity, the loads per C2 station (from 0 to 1Mbps) are divided into three regions where 0-0.35Mbps is the light load region, 0.35-0.55Mbps is the medium load region, and 0.55-1Mbps is the heavy load region. Firstly, the case where the TXOP differentiation is not adopted is considered. As shown in Fig. 3.12, under the light load region, the end-to-end delay and frame dropping probability of C1 stations are almost equal to those of C2 stations, while the throughput and energy consumption of C1 stations are double of those of C2 stations. Under the medium load region, the end-to-end delay and frame dropping probability of C1 stations are larger than those of C2 stations, while the throughput and energy consumption of C1 stations approach to those of C2 stations. Under the heavy load region, the throughput, end-to-end delay, and energy consumption of C1 stations are equal to those of C2 stations, while the frame dropping probability of C1 stations is double of that of C2 stations.

For the case of the TXOP differentiation, the performance of C1 stations improves as the TXOP limit of C1 stations increases under both the medium and heavy load regions. More specifically, when the TXOP limit of C1 stations is double of that of C2 stations, it can be observed that the frame dropping probability at C1 stations is equal to that at C2 stations, the throughput of C1 stations is double of that of C2 stations, while the end-to-end delay and the energy consumption of C1 stations are half of those of C2 stations. These results demonstrate that the use of the TXOP differentiation between the stations with different workloads is an efficient method to achieve the differentiated QoS. When the TXOP limit of C1 stations increases to 4, the performance of C1 stations further improves while that of C2 stations is deteriorated. Consequently, it is desirable to set the appropriate TXOP limits for the different stations according to their QoS constraints. On the other hand, when the network is under the light load region, the TXOP differentiation between C1 and C2 stations has little impact on their performance. The reason is straightforward as the transmission queues have few backlogged frames and thus there is always only one frame within a transmitting burst, when the network works under the light traffic loads.

3.3.2.2 Effects of the TXOP Limit

To investigate the effects of the TXOP scheme on QoS differentiation, Fig. 3.13 plots the throughput ratio between C1 and C2 stations and the aggregate throughput for

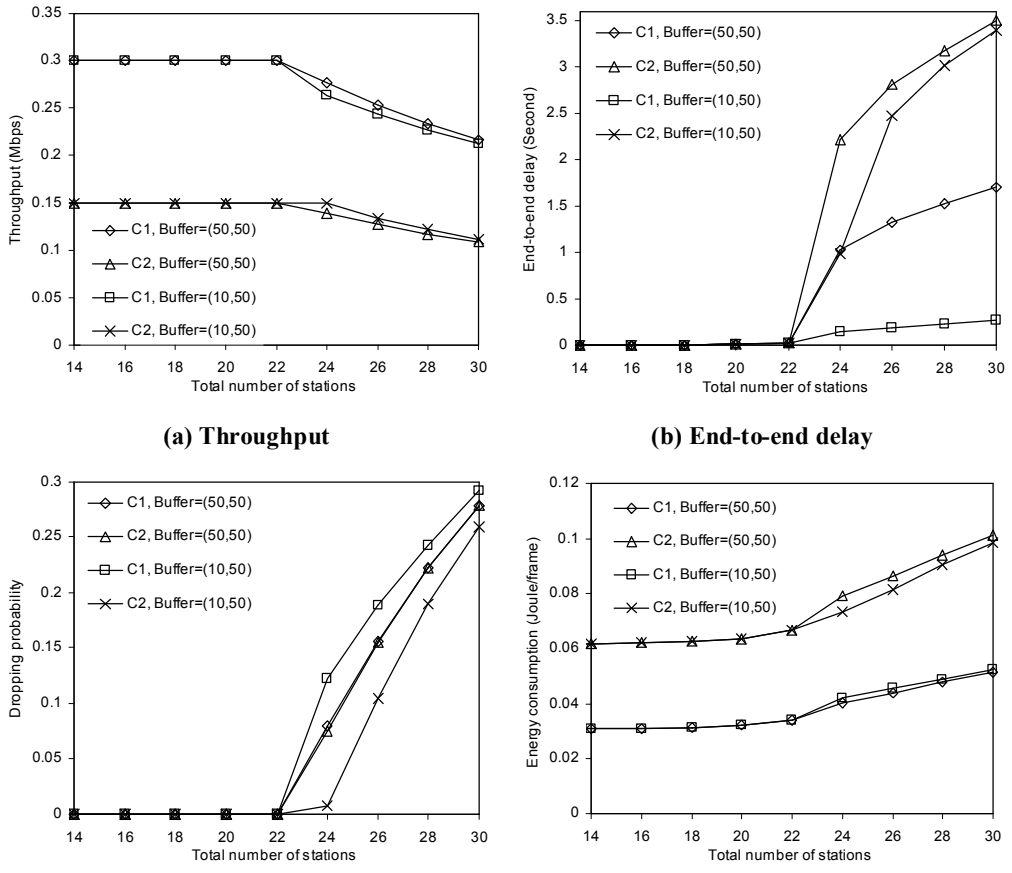


(a) Throughput ratio between C1 and C2 **(b) Aggregate throughput**

Fig. 3.13: Comparison between the DCF and TXOP scheme with the varying traffic loads.

the TXOP and DCF schemes, respectively, with the varying traffic loads per C2 station. The system consists of 10 stations divided into two classes (C1 and C2), each of which has 5 stations with the same QoS requirement. The traffic arrival rate ratio between C1 and C2 stations is set to 2, 4, and 6. The required throughput of stations in these two classes is assumed to be in direct ratio to their traffic arrival rates. To achieve this goal, the TXOP limits of stations in different classes are set proportional to their traffic arrival rates. Specifically, the TXOP limits of C1 stations are set to 2, 4, and 6 while that of C2 stations are fixed at 1. As shown in Fig. 3.13(a), for the DCF, the throughput ratio first decreases and then stabilizes at one as the traffic loads per C2 station increases. However, the throughput ratio is equal to the traffic arrival rate ratio for the TXOP scheme under any traffic loads. Fig. 3.13(b) reveals that the aggregate throughput improves when the TXOP scheme is adopted. Moreover, the aggregate throughput increases as the TXOP limits of C1 stations increase. These are because the burst transmission mechanism can improve the

efficiency of the wireless channel, since the contention overhead is amortized by all the frames transmitted within a burst.



(a) Throughput **(b) End-to-end delay**

(c) Frame dropping probability **(d) Energy Consumption**

Fig. 3.14: Analytical results versus the total number of stations with different buffer sizes.

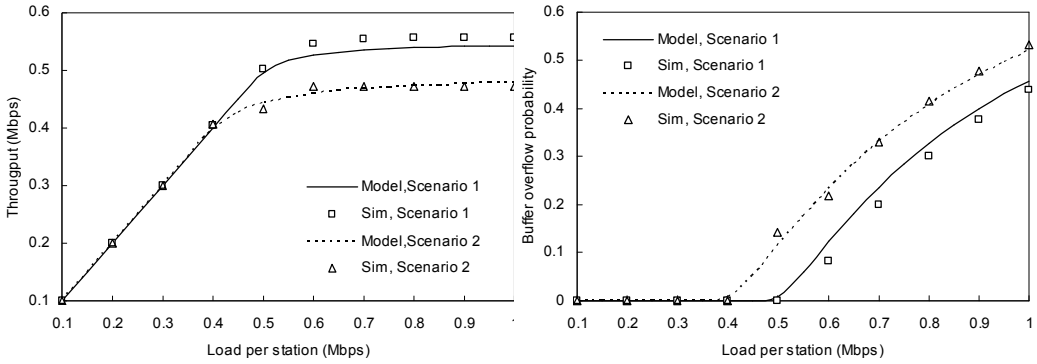
3.3.2.3 Impact of the Buffer Size and the Number of Stations

In Fig. 3.14, the impact of the buffer size and the number of stations on the performance of the TXOP scheme is investigated. Two scenarios are considered where the buffer sizes of all stations are set to 50 (frames) in Scenario 1, while the buffer sizes of C1 and C2 stations are set to 10 and 50 in Scenario 2, respectively. The traffic arrival rates at C1 and C2 stations are fixed at 0.3 Mbps and 0.15 Mbps, and the TXOP limits of C1 and C2 stations are 2 and 1, respectively. The number of

stations in each class is set to be identical. It can be observed that the network works under the unsaturated conditions when the total number of stations is not larger than 22. Particularly, the frame dropping probability is zero and the end-to-end delay is very low while the throughput keeps constant and the energy consumption increases slowly. As the number of stations increases, the throughput decreases and the end-to-end delay as well as the frame dropping probability increase dramatically since the transmission queues start to build up due to network congestion. Meanwhile, the energy consumption increases sharply due to the large amount of energy wasted on collisions.

Next, the effects of the buffer size on the QoS performance are evaluated. By comparing the performance metrics of the stations in Scenario 1 and Scenario 2, it can be observed that the buffer size has a small impact on the throughput and energy consumption. The decrease of the buffer size of C1 stations causes more frame losses and smaller throughput at C1 stations, as a result, the throughput at C2 stations increases. Since the energy consumption per successful frame transmission depends on the throughput as shown in Figs. 3.10-3.12, the buffer size that affects the throughput also has an impact on the energy consumption. On the other hand, it is shown that the buffer size has a significant impact on the end-to-end delay and frame dropping probability. More specifically, when the buffer size of C1 stations decreases from 50 to 10, it can be seen that the end-to-end delay at both C1 and C2 stations reduces considerably. On the other hand, the dropping probability of C1 stations largely increases while that of C2 stations considerably decreases. The

above observation suggests that the TXOP limit and buffer size can be utilized jointly for providing the QoS differentiation in WLANs.



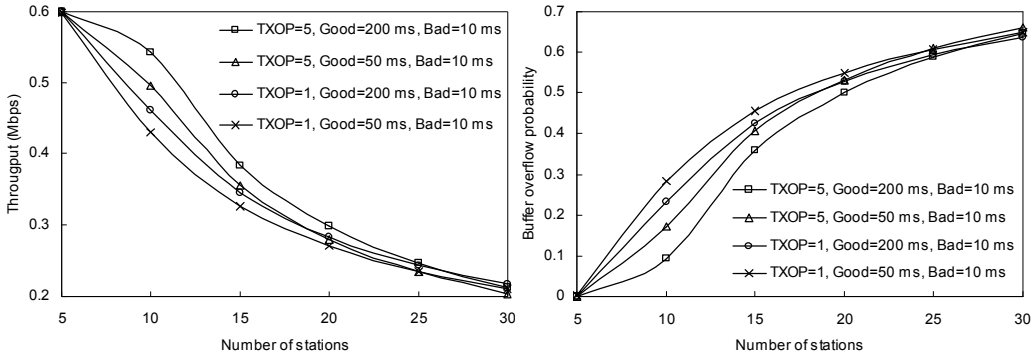
(a) Throughput **(b) Buffer overflow probability**

Fig. 3.15: Analytical and simulation results versus traffic load per station in two scenarios, Scenario 1: TXOP limit = 5, good duration = 100ms, bad duration = 10 ms, Scenario 2: TXOP limit = 2, good duration = 50 ms, bad duration = 10 ms.

3.3.3 The TXOP scheme with Bursty Error Channels

Table 3.1 lists the system parameters used in the analysis and simulation. A Basic Service Set (BSS) of WLANs with 10 stations located in a 150m × 150m rectangular grid is considered. Fig. 3.15 compares the analytical and simulation results of the throughput and buffer overflow probability of the TXOP scheme under different traffic loads and scenarios. In the first scenario, the TXOP limit is 5 (frames) and the channel exhibits bursty error with the mean good and bad state durations of 100 ms and 10 ms, respectively. In the second scenario, the TXOP limit is 2 (frames) and the mean good and bad state durations are 50 ms and 10 ms, respectively. The retry limit is 7. It is shown that the analytical results have a good

degree of match with the simulation results in both scenarios. This observation validates the accuracy of the analytical model.



(a) Throughput **(b) Buffer overflow probability**

Fig. 3.16: Analytical performance results versus the number of stations with different TXOP limits and channel conditions.

Fig. 3.16 depicts the throughput and buffer overflow probability as a function of the number of stations with different TXOP limits and channel conditions. The traffic load per station is set to 0.6 Mbps. Note that the TXOP limit $K = 1$ represents that the legacy DCF scheme is employed. It can be seen that when the number of stations is 5, the network works under light loads and the buffer overflow probability is 0, and the TXOP scheme exhibits the same performance as the DCF. However, the advantage of the TXOP scheme is demonstrated as the number of station increases. For instance, the throughput increases by 17% and the buffer overflow probability decreases by 60% under certain channel conditions $(\nu_g^{-1} = 200\,ms, \nu_b^{-1} = 10\,ms)$ when the number of station grows to 10. As the number of stations further increases, however, it can be observed that the performance improvement of the TXOP scheme decreases under both channel conditions.

Next, the effects of bursty channel errors on the network performance are evaluated. When the number of station is 5, the network works under the light loads and thus the throughput is almost equal to the traffic load as most frames can be successfully delivered to their destinations. As the number of stations grows to 10, the performances of both the DCF and TXOP schemes degrade when the duration of the good channel state decreases (*i.e.*, the channel condition turns worse). As the number of stations further increases, the collision probability generated by the high network loads plays a more important role than the channel errors. As a consequence, the difference of throughput under different channel conditions decreases. For instance, when the number of stations increases to 25, the throughput and buffer overflow probability of the TXOP scheme are around 0.23 Mbps and 0.60, for $\nu_g^{-1} = 50\,ms, \nu_b^{-1} = 10\,ms,$ and are around 0.24 Mbps and 0.59 for $\nu_g^{-1} = 200\,ms, \nu_b^{-1} = 10\,ms,$ respectively.

3.4 Summary

Analytical models have been presented for the TXOP scheme with different ACK policies, unbalanced stations, and bursty error channels. The accuracy of these models has been verified through comparing the analytical results with extensive NS-2 simulation experiments. These models derive the expressions of the important QoS performance metrics including throughput, end-to-end delay, frame dropping probability, and energy consumption. A thorough investigation into the impact of the traffic load, TXOP limit, buffer size, channel data rate, minimum contention

window, and number of stations on the QoS performance of the burst transmission schemes has been conducted. The efficiency of the TXOP scheme for QoS differentiation and the effects of the TXOP limit on the network performance in the presence of unbalanced stations have also been evaluated. Furthermore, the impact of channel errors on the performance of the TXOP scheme has been investigated.

The performance results have shown that the burst transmission schemes can substantially improve the QoS performance. In addition, the BACK scheme outperforms the CFB scheme when the TXOP limit exceeds a certain threshold. Moreover, it has been observed that the TXOP differentiation between stations has little impact on their QoS performance when the network is under the light load region. However, as the network loads become moderate and heavy, the stations with the larger TXOP limits perceive the better QoS than those with the smaller ones. It is also shown that the desirable throughput differentiation for various stations can be achieved by setting the appropriate TXOP limits. Furthermore, the analytical results have shown that the performance of the TXOP scheme degrades as the channel condition becomes worse. Moreover, The TXOP scheme outperforms the DCF under different channel conditions.

Chapter 4

A Dynamic TXOP Scheme under Self-Similar Traffic

4.1 Introduction

Most existing analytical models for TXOP have been developed under the assumption of saturated traffic loads and thus excluded any need to consider queuing or traffic models for performance analysis. However, many recent accurate measurement studies [60, 79, 123] have shown that realistic traffic in wireless and mobile networks exhibits self-similar nature (*i.e.*, extreme burstiness over a wide range of time scales) and the conventional Poisson model fails to capture the actual traffic properties. Actually, traffic self-similarity has been discovered to be a ubiquitous phenomenon in communication networks and multimedia systems [8, 9, 15, 20, 32, 48, 60, 76, 79, 104, 123] and has considerable effects on queueing performance and user-perceived QoS. Consequently, it is imperative to take the self-similar nature of network traffic into account when designing and evaluating WLANs under multimedia applications with QoS requirements.

The bursty property of self-similar traffic implies that the large bursts of packet arrivals occur frequently. Thus it is desirable to dynamically adjust the TXOP limits according to the status of the transmission queue in order to meet the specific QoS

requirements. To this end, a dynamic TXOP scheme that can adjust the TXOP limits according to the current status of the transmission queue and the pre-setting threshold is presented.

Due to the fractal-like nature of self-similar traffic, performance modelling of MAC protocols under self-similar traffic exhibits higher complexity than that under the traditional non-bursty Poisson traffic. As a result, there is no any analytical model reported for IEEE 802.11e MAC schemes under self-similar traffic in the current literature. To fill this gap, this chapter develops a new analytical model to evaluate the dynamic TXOP scheme in WLANs under self-similar traffic.

The rest of this chapter is organized as follows. Section 4.2 presents the dynamic TXOP scheme. Section 4.3 develops the analytical model for the proposed scheme under self-similar traffic. Section 4.4 validates the analytical model and conducts the performance analysis. Section 4.5 concludes the chapter.

4.2 Dynamic TXOP Scheme

In the original TXOP scheme, the TXOP limits at wireless stations are fixed. In order to accommodate the bursty property of self-similar traffic, the new TXOP scheme proposes to dynamically adjust the TXOP limit according to the status of the current transmission queue. As shown in Fig. 4.1, when the queue length is below a certain threshold, the TXOP limit is fixed at the default value. However, if the queue length exceeds the threshold, the TXOP limit is augmented to a new value larger

than the default one. This new TXOP value cannot be set too large because a large TXOP often causes high performance oscillations and poor fairness. In what follows, the default TXOP value, new TXOP value, and threshold are denoted as K_l, K_h, and Th, respectively, with $K_l \le K_h \le Th$.

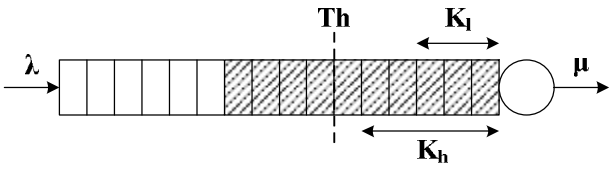


Fig. 4.1: A queue instance for the dynamic TXOP scheme.

4.3 System Model

A group of wireless stations incorporated with the dynamic TXOP scheme under self-similar traffic is considered. The transmission queue of each station is modelled as a bulk service queueing system. Firstly, the backoff procedure is analyzed and the mean service time of the queueing system is derived. Secondly, a Markovian approach is employed to model traffic self-similarity over a number of time scales for the analytical tractability. Finally, a bi-variate Markov chain is used to model the queueing system and derive the desired performance measures.

4.3.1 Analysis of the Backoff Procedure

Let p denote the probability that a transmitted frame from the station encounters a collision, let τ be the probability that the station transmits a frame in

a randomly chosen time slot, given that its transmission queue is non-empty. Based on the Markov chain proposed in [16], τ is given by

$$\tau = \frac{2(1-2p)}{(1-2p)(W+1)+pW(1-(2p)^m)} \tag{4.1}$$

where W is the minimum contention window and m represents the maximum backoff stage.

The collision probability, p, is the probability that at least one of the remaining stations transmits in the considered time slot and is given by

$$p = 1-(1-(1-\pi_0)\tau)^{n-1} \tag{4.2}$$

where π_0 is the probability that the transmission queue of the station is empty.

The service time is defined as the duration from the instant that a Head-of-Burst (HoB) frame starts contending for the channel to the instant that the burst is acknowledged following successful transmission. The service time includes two parts: the channel access delay and burst transmission delay. The former is the time duration from the instant that the frame reaches to the head of the transmission queue until it wins the contention and is ready for transmission. The latter is the time duration of successfully transmitting a burst. Let $E[S_i]$, $E[A]$ and $E[B_i]$ denote the means of the service time, channel access delay and burst transmission delay, respectively, where i represents the number of frames transmitted in a burst. The mean channel access delay can be calculated as

$$E[A] = T_c\varphi + \sigma'\omega \tag{4.3}$$

where T_c is the average collision time, σ' is the average length of time slot perceived by the station. φ and ω account for the number of collisions before a successful transmission from the station and the average number of time slots that the station defers in backoff stages, respectively.

$$\varphi = \sum_{v=0}^{\infty} v p^v (1-p) = \frac{p}{1-p} \tag{4.4}$$

$$\omega = \sum_{v=0}^{\infty} \sum_{s=0}^{v} \frac{W_s - 1}{2} p^v (1-p) \tag{4.5}$$

where $p^v(1-p)$ is the probability that the frame is successfully transmitted after v collisions, and $(W_s - 1)/2$ is the average value of the backoff counter generated at the *s*-th backoff stage.

Let P_t be the probability that at least one station among the remaining $(n-1)$ stations transmits in a time slot. Let P_s denote the probability that there is a successful transmission among the $(n-1)$ remaining stations. P_t and P_s are given by

$$P_t = 1 - (1 - (1-\pi_0)\tau)^{n-1} \tag{4.6}$$

$$P_s = (n-1)(1-\pi_0)\tau(1-(1-\pi_0)\tau)^{n-2} \tag{4.7}$$

The average length of time slot perceived by the station, σ', is obtained by considering the fact that the channel is idle with probability $(1-P_t)$, a successful transmission occurs with probability P_s, and a collision happens with probability $(P_t - P_s)$.

$$\sigma' = (1-P_t)\sigma + P_s T_s + (P_t - P_s) T_c \tag{4.8}$$

where σ is the duration of an empty time slot, T_s is the average time for the successful transmission of a burst, and T_c is the average collision time, respectively.

Note that only the HoB frame can be involved in the collision using the burst transmission scheme. T_c is given by

$$T_c = T_{DIFS} + T_L + T_H + T_{SIFS} + T_{ACK} \tag{4.9}$$

where T_L and T_H are the average times required for transmitting the frame payload and frame header, respectively. The average time for the successful transmission of a burst, T_s, can be written as

$$T_s = \frac{\sum_{i=1}^{K_l} E[B_i]L_i + E[B_{K_h}]L_{K_h}}{(1-\pi_0)} \tag{4.10}$$

where K_l and K_h denote the maximum number of frames that can be transmitted in the default and new TXOP limits, respectively, dominator $(1-\pi_0)$ indicates that the occurrence of burst transmission is conditioned on the fact that there is at least one frame in the transmission queue, L_i $(1 \le i \le K_l$ or $i = K_h)$ is the probability of having i frames in the burst, and $E[B_i]$ $(1 \le i \le K_l$ or $i = K_h)$ is the burst transmission delay, which is dependent on the number of frames transmitted within a burst and can be given by

$$E[B_i] = T_{DIFS} + i(T_L + T_H + 2T_{SIFS} + T_{ACK}) - T_{SIFS} \tag{4.11}$$

4.3.2 Queueing Analysis

Many models, such as chaotic maps and Fractional Brownian Motion [9], have been proposed to model traffic self-similarity. However, queueing theoretical techniques developed in the past are hardly applicable for these models. Anderson and Nielson [6] have proposed a method to model self-similarity using a superposition of two-state Markov-Modulated Poisson Processes (MMPPs), each of which is able to capture traffic burstiness over a given time scale. The parameters of these MMPPs are determined so as to match the mean and autocorrelation of the self-similar process over several different time scales. This simplified modeling method enables the existing analytical tools of Markov models to be available for deriving the desired performance measures [94]. Moreover, this method is practical because the measurement studies [15, 76, 104] have shown that traffic burstiness appears typically over four or five orders of time scales.

Traffic generated by each station follows self-similar process with the mean arrival rate λ, Hurst parameter H, autocorrelation at lag 1 $r(1)$, and the number of time scales, ϕ, over which the burstiness appears. Following the algorithm [6], traffic self-similarity can be modelled by the superposition of M two-state MMPPs, typically $M = 4$. For clarity, the MMPPr with subscript r is used to denote the rth two-state MMPP ($1 \le r \le M$), which can be parameterized by the infinitesimal generator, $\mathbf{Q}_r$, and the rate matrix, $\mathbf{\Lambda}_r$ as

$$\mathbf{Q}_r = \begin{bmatrix} -\delta_{1r} & \delta_{1r} \\ -\delta_{2r} & \delta_{2r} \end{bmatrix} \qquad \text{and} \qquad \mathbf{\Lambda}_r = \begin{bmatrix} \lambda_{1r} & 0 \\ 0 & \lambda_{2r} \end{bmatrix} \tag{4.12}$$

The parameters $\delta_{1r}, \delta_{2r}, \lambda_{1r}, \lambda_{2r}$ for each MMPPr ($1 \le r \le M$) are derived by the fitting algorithm presented in [6]. The superposition of the MMPPs gives rise to a new MMPP with 2^M states and its parameter matrices, $\mathbf{Q}$ and $\mathbf{\Lambda}$, is given as follows (the symbol "⊕" denotes the Kronecker sum [39])

$$\begin{aligned} \mathbf{Q} &= \mathbf{Q}_1 \oplus \mathbf{Q}_2 \oplus \cdots \oplus \mathbf{Q}_M \\ \mathbf{\Lambda} &= \mathbf{\Lambda}_1 \oplus \mathbf{\Lambda}_2 \oplus \cdots \oplus \mathbf{\Lambda}_M \end{aligned} \tag{4.13}$$

The resulting multi-sate MMPP is then used to characterize the self-similar traffic generated by the stations. The transmission queue at each station with the self-similar arrival process can be characterized as an MMPP/$G^{[1,K_h]}$/1/N system, where the superscript $[1, K_h]$ denotes that the actual number of frames transmitted in a burst ranges from 1 to K_h, and N represents the buffer size at each station.

The service time for transmitting i frames in a burst can be modelled by an exponential distribution function with mean $E[S_i]$. Thus, the service rate, μ_i, is given by $1/E[S_i]$.

The queueing system is modelled by a bi-variate Markov chain with $(N+1)\times 2^{\mathrm{M}}$ states where state (i,η), $(i = 0, 1, ..., N)$ and $(1 \le \eta \le 2^M)$, denotes that there are i frames in the system and the multi-state MMPP is at state η. For the sake of clarity, Fig. 4.2 illustrates a simple version of the Markov chain when the MMPP has only two states. The transition rate out of state (i,η) to state $(i+1,\eta)$ is λ_η, which is the traffic arrival rate when the MMPP is at state η. A transition from state (i,η) to state $(i-K_h,\eta)$, $(Th \le i \le N)$, is μ_{K_h}, from state (i,η) to state $(i-K_l,\eta)$, $(K_l \le i < Th)$, is μ_{K_l} and from state (i,η) to state $(0,\eta)$,

$(1 \le i \le K_l - 1)$, is μ_i. The transition rate out of state (i, η) to state (i, η'), where $(1 \le \eta, \eta' \le 2^M)$ and $(\eta \ne \eta')$, is the same as the rate from state η to state η' of the MMPP, i.e., $\delta_\eta = \mathbf{Q}(\eta, \eta')$.

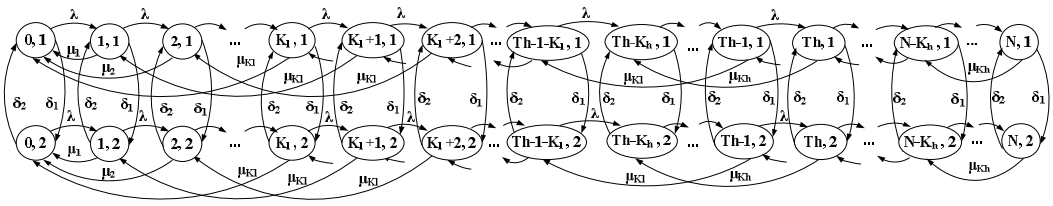


Fig. 4.2: The MMPP/G$^{[1,Kh]}$/1/N queue state-transition-rate diagram.

The transition rate matrix, $\mathbf{G}$, of the bi-variate Markov chain can be obtained by the state-transition-rate diagram. The steady-state probability vector, $\mathbf{P} = (P_{i,\eta}) = (\mathbf{P}_0, \mathbf{P}_1, \cdots, \mathbf{P}_N)$ where $\mathbf{P}_i = (P_{i,0}, P_{i,1}, ..., P_{i,M})$ satisfies the following equations: $\mathbf{PG} = 0$ and $\mathbf{Pe} = 1$. After obtaining $\mathbf{P}$, L_i can be expressed as

$$\begin{cases} L_i = \sum_{\eta=1}^{M} P_{i\eta},\ 1 \le i < K_l, L_i = \sum_{z=K_l}^{Th-1} \sum_{\eta=1}^{\mathrm{M}} P_{z\eta}, & i = K_l \\ L_i = \sum_{z=T_h}^{N} \sum_{\eta=1}^{\mathrm{M}} P_{z\eta}, & i = K_h \end{cases} \tag{4.14}$$

Let $\boldsymbol{\pi}$ denote the steady-state distribution of the number of frames in the queueing system. The steady-state probability that there are i frames in the system is

$$\pi_i = \mathbf{P}_i \mathbf{e}, \quad \text{for } 0 \le i \le N \tag{4.15}$$

Thus the probability, π_0, that the transmission queue is empty can be obtained.

Let $\boldsymbol{\pi}'$ represent the steady-state distribution of the number of frames in the queueing system as seen by an arbitrary arriving frame. The steady-state probability that there are i frames in the system as seen by the arriving frame is [92]

$$\pi_i' = \left(\sum_{i=0}^{N} \mathbf{P}_i \boldsymbol{\Lambda} \mathbf{e} \right)^{-1} \mathbf{P}_i \boldsymbol{\Lambda} \mathbf{e}, \quad \text{for } 0 \le i \le N \tag{4.16}$$

where $\boldsymbol{\Lambda}$ is given in Eq. (4.13).

Therefore the loss probability, p_b, that an arriving frame finds the finite buffer full, is given by π_N'. Given the loss probability p_b, the throughput TH of each station can be computed by $TH = \lambda E[P](1 - p_b)$, where $E[P]$ is the average frame payload length. λ is the mean traffic arrival rate.

The end-to-end delay is the time duration from the time instant a frame enters the transmission queue of the station to the time instant the frame is removed from the transmission queue after its successful transmission. It equals to the queueing delay plus service time and can be given by

$$E[D] = \frac{E[N]}{\lambda(1 - p_b)} \tag{4.17}$$

where $E[N] = \sum_{z=1}^{N} z\pi_z$ is the average number of frames in the queueing system. $\lambda(1 - p_b)$ is the effective arrival rate to the transmission queue since the arriving frames are discarded if finding the finite buffer full.

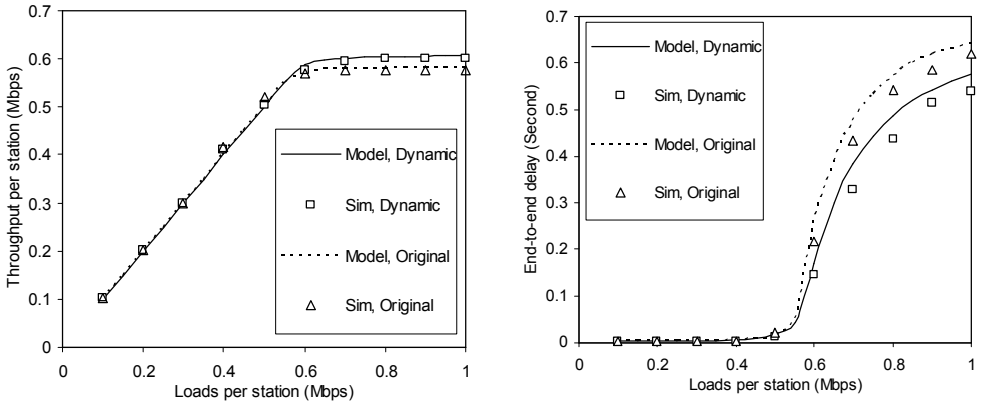
Table 4.1: System Parameters for the Analysis of the TXOP Scheme under Self-Similar Traffic

Frame payload	8000 bits	PHY header	192 bits
MAC header	224 bits	ACK	112 bits + PHY header
Channel data rate	11 Mbps	CW_{min}	32
Basic rate	1 Mbps	CW_{max}	1024
Buffer size	50 frames	DIFS	50 μs
Slot time	20 μs	SIFS	10 μs

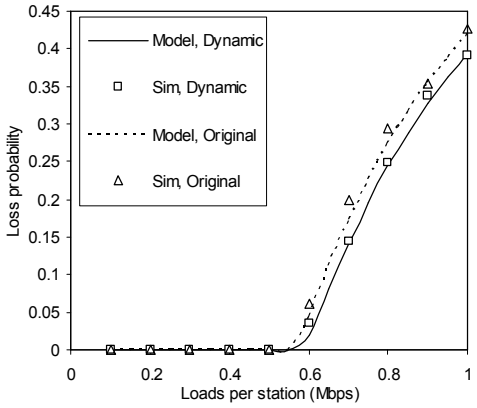
4.4 Model Validation and Performance Analysis

The analytical model is validated via NS-2 simulation experiments which run in a scenario of 10 stations located within a Basic Service Set (BSS). The traffic generated at each station is self-similar. The threshold of the queue length that triggers the switch from the default TXOP limit to the new one is 10 frames. The new TXOP limit is set to be double of the default one, e.g., $K_l = 3$ and $K_h = 6$. Other parameters used in the model and simulations are shown in Table 4.1.

Fig. 4.3 depicts the throughput, end-to-end delay, and frame loss probability versus the mean offered loads for both the original and new dynamic TXOP schemes. The Hurst parameter H, and the autocorrelation at lag 1 of the self-similar traffic are shown in the caption of the figure. Traffic burstiness is modelled over five time scales.



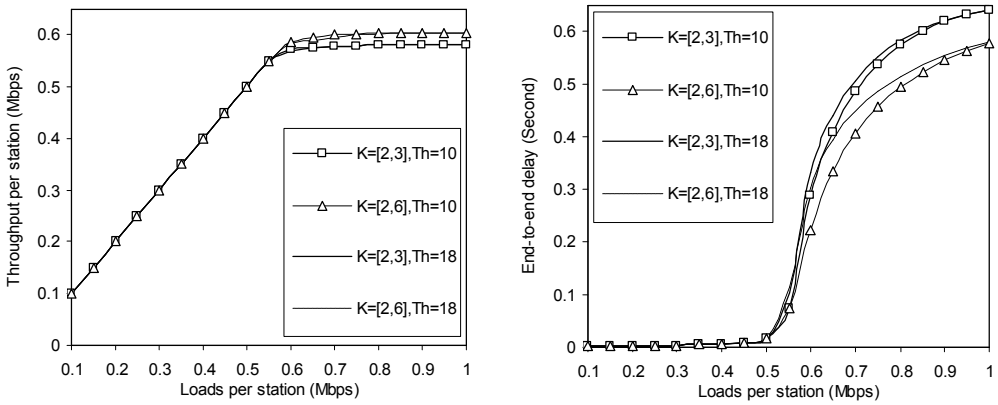
(a) Throughput **(b) End-to-end delay**



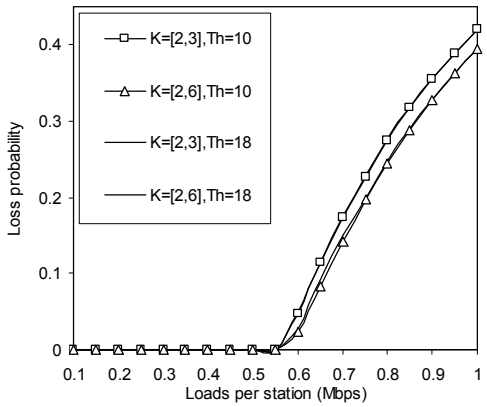
(c) Frame loss probability

Fig. 4.3: Performance measures versus offered loads per station for the original TXOP scheme and dynamic one. Self-Similar parameters: H = 0.85, r(1) = 0.6.

As shown in Fig. 4.3, the analytical results closely match with those obtained from simulation experiments, which validates the accuracy of the proposed model. It is also observed that the dynamic TXOP scheme achieves the larger throughput and lower delay and frame loss probability than the default one, thus demonstrating the effectiveness of the dynamic TXOP scheme.



(a) Throughput **(b) End-to-end delay**



(c) Frame loss probability

Fig. 4.4: Analytical results versus offered loads per station for different values of new TXOP limits (3 and 6) and thresholds (10 and 18). Self-Similar parameters: H = 0.85, r(1) = 0.6.

In Fig. 4.4, the impacts of the TXOP limit and threshold on the network performance are investigated. Firstly, it is noted that although the threshold has minor impacts on the system throughput and frame loss probability, the end-to-end delay declines as the threshold increases from 10 to 18 frames. It is also shown that the throughput, delay, and loss probability performance improve as the value of the new TXOP limit increases from 3 to 6 frames. However, the further increase of the new TXOP limit only leads to little performance improvement. Moreover, a too large TXOP limit can cause the large performance jitter and poor fairness. Therefore,

the optimal value for the new TXOP limit of the dynamic TXOP scheme is around 6 under this case. The above analysis demonstrates that the developed new analytical model is an effective and efficient tool for obtaining the optimal settings of the dynamic TXOP scheme.

4.5 Summary

In order to adapt to the bursty nature of self-similar traffic, a dynamic TXOP scheme which adjusts the TXOP limits according to the transmission queue length has been presented. An analytical model has been developed for this new scheme under self-similar traffic. The QoS performance metrics in terms of throughput, end-to-end delay, and frame loss probability have been derived with the model and validated against NS-2 simulation experiments. The performance results have shown that the dynamic TXOP scheme outperforms the original one specified in the IEEE 802.11e standard. The impact of the TXOP limit and threshold on the network performance has also been investigated. The results have demonstrated that the developed model is an effective tool for obtaining the optimal settings of the dynamic TXOP scheme.

Chapter 5

Modelling the TXOP Scheme in Multimedia WLANs

5.1 Introduction

WLANs are currently integrating a diverse range of traffic sources that significantly differ in their packet arrival patterns. Although initially successful and analytically simple for modelling the non-bursty traffic behavior, the Poisson model has proven inadequate for capturing traffic burstiness of compressed voice and video in modern communication networks, where batch arrivals, event correlations and burstiness are important factors. It is well known that VBR VoIP generates traffic with time-varying arrival rates [77, 114]. Furthermore, many studies by means of high quality, high time-resolution measurements [8, 9, 15, 40] have demonstrated that VBR video traffic exhibits noticeable burstiness over a wide range of time scales. This fractal behaviour of video traffic can be modelled using statistically self-similar processes, which have significantly different theoretical properties from those of the conventional Poisson process. Therefore, it is critical and timely to take the heterogeneous characteristics of multimedia traffic into account in order to accurately evaluate and obtain a better understanding of the performance characteristics of multimedia WLANs. To this end, this chapter proposes a new

analytical model for the IEEE 802.11e TXOP scheme in WLANs with heterogeneous stations.

The rest of this chapter is organized as follows. The analytical model for the TXOP scheme in the presence of heterogeneous stations is developed in Section 5.2. Section 5.3 validates the accuracy of the model and conducts the performance evaluation. Finally, Section 5.4 concludes the chapter.

5.2 Analytical Model

A WLAN comprising C different classes of stations where Class $i\,(i = 1, 2, \cdots, C)$ has n_i stations is considered. The network supports multimedia applications including voice, background data, real-time and non-real-time video. To capture the characteristics of burst transmission of the TXOP scheme, the transmission queue at each station is modelled as a bulk service queueing system where the service time is obtained by analyzing the backoff procedure and burst transmission mechanism.

5.2.1 Analysis of the Service Time

In the following, the term *time slot* is used to denote the time interval between the starts of two consecutive decrements of the backoff counter, while the term *physical time slot* represents a fixed time interval (unit time) specified in the IEEE 802.11 standard [58]. Firstly, the analytical model [16] is extended to derive the burst

transmission probability, τ_i, of a station in Class i under the unsaturated traffic conditions.

Since a station transmits only when its transmission queue is non-empty, the burst transmission probability, τ_i, can be obtained by weighting the saturation transmission probability with the probability of the non-empty transmission queue. τ_i can be written as

$$\tau_i = (1 - Pi_0)\tau_i' \tag{5.1}$$

where Pi_0 is the probability that the transmission queue of the station in Class i is empty and will be derived in Section 4.2. τ_i' is the probability that the station transmits when its transmission queue is non-empty and can be given by [16]

$$\tau_i' = \frac{2(1-2p_i)}{(1-2p_i)(W+1) + p_i W(1-(2p_i)^m)} \tag{5.2}$$

where m represents the maximum backoff stage and W is the minimum contention window.

The collision probability, p_i, equals to the probability that at least one of the remaining stations transmits in a given time slot. p_i, can be written as

$$p_i = 1 - (1-\tau_i)^{n_i - 1} \prod_{r \neq i} (1-\tau_r)^{n_r} \tag{5.3}$$

Note that only the Head-of-Burst (HoB) frame needs to contend for the channel. The service time of a burst transmission is defined as the time interval from the instant that a HoB frame starts contending for the channel to the instant that the

whole burst is acknowledged following successful transmission. The service time consists of two parts: the channel access delay and burst transmission delay. The former is the time interval from the instant that the HoB frame reaches the head of its transmission queue and starts contending for the channel, until it wins the contention and is ready for transmission. The latter is the time interval of successfully transmitting the burst. Let $E[S_{vi}]$, $E[A_i]$ and $E[B_v]$ denote the mean service time, channel access delay and burst transmission delay, respectively, where v represents the number of frames transmitted in a burst and i denotes that the burst transmission is from a station in Class i. $E[S_{vi}]$ can be written as

$$E[S_{vi}] = E[A_i] + E[B_v] \tag{5.4}$$

The average channel access delay, $E[A_i]$, can be calculated as

$$E[A_i] = T_c \varphi_i + \sigma_i' \delta_i \tag{5.5}$$

where T_c is the average collision time and σ_i' is the average length of a time slot. φ_i accounts for the average number of collisions before a successful transmission from the station and is given by

$$\varphi_i = \sum_{j=0}^{\infty} j p_i^j (1 - p_i) = \frac{p_i}{1 - p_i} \tag{5.6}$$

δ_i denotes the average number of time slots that the station defers during backoff stages and can be expressed as

$$\delta_i = \sum_{j=0}^{\infty} \sum_{h=0}^{j} \frac{W_h - 1}{2} p_i^j (1 - p_i) \tag{5.7}$$

where $p_i^j(1-p_i)$ is the probability that the HoB frame is successfully transmitted after j collisions and $(W_h-1)/2$ denotes the mean of the backoff counters generated in the *h*-th backoff stage.

Let P_{t_i} represent the probability that at least one of the remaining stations transmits in a given time slot when a station in Class i is in the backoff procedure. P_{t_i} is equal to the collision probability, p_i, when a station in Class i transmits, as given in Eq. (5.3). The probability, $P_{s_i}^r$, that a station in Class r $(1 \le r \le C)$ successfully transmits when the station in Class i is in the backoff procedure, can be expressed as

$$P_{s_i}^r = \tau_r(1-\tau_r)^{n_r-1}(1-\tau_i)^{n_i-1}\prod_{k\neq r,i}(1-\tau_k)^{n_k} \tag{5.8}$$

The average size of a time slot, σ_i', when the station in Class i is in the backoff procedure is obtained by considering the fact that the channel is idle with probability $(1-P_{t_i})$, a successful transmission occurs with probability $\sum_{r=1}^{C} P_{s_i}^r$, and a collision happens with probability $(P_{t_i} - \sum_{r=1}^{C} P_{s_i}^r)$. Thus, σ_i' can be written as

$$\sigma_i' = (1-P_{t_i})\sigma + \sum_{r=1}^{C} P_{s_i}^r T_s^r + (P_{t_i} - \sum_{r=1}^{C} P_{s_i}^r)T_c \tag{5.9}$$

where σ is the duration of a physical time slot [58], T_s^r is the average time for the successful transmission of a burst from the station in Class r, and T_c is the average collision time.

Note that only the HoB frame can be involved in the collision using the TXOP

scheme, T_c is given by

$$T_c = T_{DIFS} + T_L + T_H + T_{SIFS} + T_{ACK} + 2\Delta \tag{5.10}$$

where T_L and T_H are the average time required for transmitting the frame payload and the frame header, respectively. Δ denotes the propagation delay. The average time, T_s^r, for the successful transmission of a burst from the station in Class r can be written as

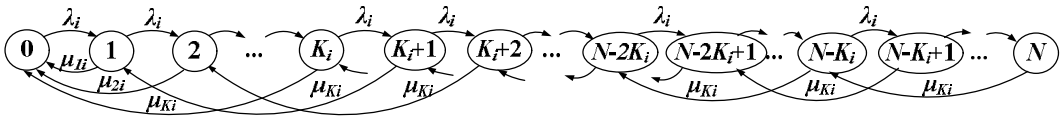
$$T_s^r = \frac{\sum_{v=1}^{K_r} E[B_v] L_{vr}}{1 - Pr_0} \tag{5.11}$$

where K_r denotes the maximum number of frames that can be transmitted in a TXOP limit of the station in Class r, the denominator $(1 - Pr_0)$ means that the occurrence of burst transmission is conditioned on the fact that there is at least one frame in the transmission queue of the station, L_{vr} is the probability that v $(1 \le v \le K_r)$ frames are transmitted from the station within a TXOP limit. Let Pr_s represent the probability that there are s $(0 \le s \le N)$ frames in the transmission queue where N is the system capacity of the station. The probability, L_{vr}, can be written as

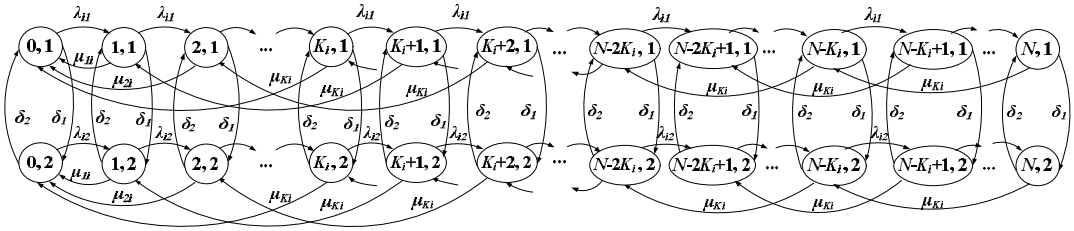
$$\begin{cases} L_{vr} = Pr_v & 1 \le v < K_r \\ L_{vr} = \sum_{s=K_r}^{N} Pr_s & v = K_r \end{cases} \tag{5.12}$$

$E[B_v]$ is the burst transmission delay which is dependent on the number of frames transmitted within a burst and can be expressed as

$$E[B_v] = T_{DIFS} + v(T_L + T_H + 2T_{SIFS} + T_{ACK} + 2\Delta) - T_{SIFS} \tag{5.13}$$



(a)



(b)

Fig. 5.1: State-transition-rate diagram (a) M/G$^{[1, Ki]}$/1/N queue (b) MMPP/G$^{[1, Ki]}$/1/N queue.

5.2.2 Queueing Analysis

This subsection elaborates on modelling and analysis of transmission queues at stations. For the queueing system of a station in Class *i*, the service time can be modelled by an exponential distribution function with mean $E[S_{vi}]$ as the service time distribution obtained from simulation experiments was compared with various standard distributions and was found that the exponential distribution gives a good approximation to the service time [140]. Thus, the mean service rate, μ_{vi}, is given by $1/E[S_{vi}]$. In what follows, the queueing systems for the stations with Poisson, two-state MMPP, and self-similar traffic will be presented, respectively.

5.2.2.1 Stations with Poisson Traffic

The transmission queue at the station with non-bursty Poisson traffic is modelled as an M/G$^{[1,K_i]}$/1/N queueing system, where the superscript $[1, K_i]$ denotes that the number of frames transmitted during a TXOP ranges from 1 to

K_i, and N represents the system capacity.

Fig. 5.1(a) illustrates the state-transition-rate diagram of the queuing system where the states denote the numbers of frames in the system. The transition rate from state s to state $s+1$, $(0 \le s \le N-1)$, is the arrival rate λ_i of the Poisson process. A transition out of state s to state $s-K_i$, $(K_i \le s \le N)$, implies that the burst transmission of K_i frames completes and the transition rate is μ_{Ki}. The change from state s to state 0, $(1 \le s \le K_i - 1)$, denotes that all s frames in the system are transmitted within a burst and the transition rate is μ_{si}.

The transition rate matrix, $\mathbf{G}$, of the Markov chain can be obtained in Fig. 5.1(a). The steady-state probability vector, $\mathbf{P} = (Pi_s, s = 0, 1, \ldots, N)$ of the Markov chain satisfies the following equations

$$\mathbf{PG} = 0 \quad \text{and} \quad \mathbf{Pe} = 1 \tag{5.14}$$

Solving these equations yields the steady-state vector as [39]

$$\mathbf{P} = \mathbf{u}(\mathbf{I} - \Re + \mathbf{eu})^{-1} \tag{5.15}$$

where $\Re = \mathbf{I} + \mathbf{G} / \min\{\mathbf{G}(\rho, \rho)\}$, $\mathbf{u}$ is an arbitrary row vector of $\Re$ and $\mathbf{e}$ is a unit column vector. Thus, the probability, Pi_0, that the transmission queue of the station in Class i is empty is obtained.

5.2.2.2 Stations with MMPP Traffic

The transmission queue at the station with the bursty MMPP traffic is modelled as an MMPP/$G^{[1,K_i]}$/1/N queueing system, which is characterized by a bivariate

Markov chain as shown in Fig. 5.1(b). State (s,η), where $(s = 0, 1, \ldots, N)$ and $(\eta = 1, 2)$, represents the case that there are s frames in the queueing system and the two-state MMPP characterizing the traffic on the station is at state η. The transition rate out of state (s,η) to $(s+1,\eta)$ is $\lambda_{i\eta}$, which is the traffic arrival rate when the MMPP is at state η. The transition rate out of state (s,η) to $(s-K_i,\eta)$, $(K_i \le s \le N)$, is μ_{Ki} and that from state (s,η) to $(0,\eta)$, $(1 \le s \le K_i - 1)$, is μ_{si}. The transition rate out of state (s,η) to (s,η'), where $(\eta' = 1, 2)$ and $(\eta' \ne \eta)$ is the same as the transition rate from state η to η' of the MMPP, *i.e.*, $\delta_\eta = \mathbf{Q}(\eta,\eta')$. Following the above analysis, the transition rate matrix, $\mathbf{G}$, of the bivarite Markov chain can be obtained in Fig. 5.1(b). Using Eqs. (5.14) and (5.15) the steady-state probability vector, $\mathbf{P} = (Pi_{s,\eta}) = (\mathbf{P}_0, \mathbf{P}_1, \cdots, \mathbf{P}_N)$ where $\mathbf{P}_s = (Pi_{s,0}, Pi_{s,1}),\ 0 \le s \le N$ can be derived.

5.2.2.3 Stations with Self-Similar Traffic

Due to the fractal nature, modelling of self-similar traffic poses greater challenges and exhibits more complexity than the traditional short-range-dependent traffic. Many models, such as chaotic maps and Fractional Brownian Motion [104], have been proposed to capture traffic self-similarity. However, queueing theoretical techniques developed in the past are hardly applicable for these models. Anderson and Nielson [6] have proposed a method for modelling self-similarity using a superposition of different two-state MMPPs, each of which is able to capture traffic burstiness over a given time scale. The parameters of these MMPPs are determined so as to match the mean and autocorrelation of the self-similar process over several

different time scales. This simplified modeling method enables the existing analytical tools of Markovian models applicable for deriving the desired performance measures [94]. Moreover, this method is practical because the measurement studies [15, 76, 104] have shown that traffic burstiness appears typically over four or five orders of time scales.

Traffic generated by stations with the self-similar arrival process can be characterized by the mean arrival rate λ, Hurst parameter H, autocorrelation at lag 1 $r(1)$, and the number of time scales, ϕ, over which the burstiness appears. Traffic self-similarity can be modelled by the superposition of L two-state MMPPs, typically $L = 4$, [76, 104]. For clarity, the MMPP_j with subscript j is used to denote the j-th two-state MMPP $(1 \le j \le L)$, which can be parameterized by the infinitesimal generator, $\mathbf{Q}_j$, and the rate matrix, $\mathbf{\Lambda}_j$ as follows

$$\mathbf{Q}_j = \begin{bmatrix} -\delta_{1j} & \delta_{1j} \\ \delta_{2j} & -\delta_{2j} \end{bmatrix} \qquad \text{and} \qquad \mathbf{\Lambda}_j = \begin{bmatrix} \lambda_{1j} & 0 \\ 0 & \lambda_{2j} \end{bmatrix} \tag{5.16}$$

The parameters $\delta_{1j}, \delta_{2j}, \lambda_{1j}, \lambda_{2j}$ for each MMPP_j $(1 \le j \le L)$ are derived by the fitting algorithm presented in [6]. The superposition of the MMPPs gives rise to a new MMPP with 2^L states and its parameter matrices, $\mathbf{Q}$ and $\mathbf{\Lambda}$, can be calculated as follows (the symbol "$\oplus$" denotes the Kronecker sum [39])

$$\begin{aligned} \mathbf{Q} &= \mathbf{Q}_1 \oplus \mathbf{Q}_2 \oplus \cdots \oplus \mathbf{Q}_L \\ \mathbf{\Lambda} &= \mathbf{\Lambda}_1 \oplus \mathbf{\Lambda}_2 \oplus \cdots \oplus \mathbf{\Lambda}_L \end{aligned} \tag{5.17}$$

The resulting multi-sate MMPP is then used to characterize the self-similar traffic. The bivariate Markov chain shown in Fig. 5.1(b) can be extended to have

$(N+1)\times 2^{\mathrm{L}}$ states where state (s,η), $(s=0,1,\ldots,N)$ and $(1\le\eta\le 2^{L})$, denotes that there are s frames in the system and the multi-state MMPP is at state η. Obtaining the transition rate matrix, $\mathbf{G}$, of this Markov chain, Eqs. (5.14) and (5.15) can be used to compute the steady-state probability vector, $\mathbf{P}=(Pi_{s,\eta})=(\mathbf{P}_0,\mathbf{P}_1,\cdots,\mathbf{P}_N)$ where $\mathbf{P}_s=(Pi_{s,0},Pi_{s,1},\ldots,Pi_{s,\eta})$, $0\le s\le N$.

5.2.2.4 Performance Measures

Let π_{si} denote the steady-state probability that there are s frames in the queueing system of the station in Class i, π_{si} is given by

$$\pi_{si}=\mathbf{P}_s\mathbf{e},\qquad \text{for } 0\le s\le N \tag{5.18}$$

Thus, the probability, Pi_0, that the transmission queue of the station in Class i is empty, can be given by π_{0i}.

Let π'_{si} represent the steady-state probability that there are s frames in the queueing system of the station in Class i as seen by the arriving frame. In the case of MMPP or self-similar traffic, π'_{si} can be written as [92]

$$\pi'_{si}=\frac{\mathbf{P}_s\mathbf{\Lambda e}}{\sum_{s=0}^{N}\mathbf{P}_s\mathbf{\Lambda e}},\qquad \text{for } 0\le s\le N \tag{5.19}$$

where $\mathbf{\Lambda}$ is given in Eq. (5.17). The loss probability, p_{bi}, which is the probability that an arriving frame finds the buffer of the station in Class i full, can be given by π'_{Ni}.

In the case that the traffic follows a Poisson process, the steady-state probability distribution, $\boldsymbol{\pi}'_i$, of the number of frames in the queueing system as

seen by an arbitrary arriving frame is equivalent to the steady-state probability distribution, $\boldsymbol{\pi}_i$, of the number of frames in the queueing system. Thus, the loss probability, p_{bi}, is equal to Pi_N.

The throughput, TH_i, of the station in Class i can be further computed by

$$TH_i = \lambda_i E[P](1 - p_{bi}) \tag{5.20}$$

where $E[P]$ is the frame payload length and λ_i is the mean traffic arrival rate of the station in Class i.

The end-to-end delay is the time interval from the instant that a frame enters the transmission queue of the station, to the instant that the frame is acknowledged after successful transmission. By virtue of Little's Theorem, the average end-to-end delay, $E[D_i]$, is given by

$$E[D_i] = \frac{E[N_i]}{\lambda_i(1 - p_{bi})} \tag{5.21}$$

where $E[N_i] = \sum_{s=1}^{N} s\pi_{si}$ is the average number of frames in the queueing system of the station in Class i. $\lambda_i(1 - p_{bi})$ is the effective arrival rate to the transmission queue of the station since the arriving frames are discarded if the finite buffer becomes full.

5.3 Model Validation and Performance Evaluation

This section first investigates the accuracy of the proposed analytical model through simulation experiments and then adopts the model to conduct performance

evaluation.

5.3.1 Model Validation

The accuracy of the analytical model has been validated through extensive NS-2 [99] simulation experiments. A Basic Service Set (BSS) of WLANs is considered where stations are located in a rectangular grid with dimension 100m $\times$ 100m and are classified into four categories [59]. Specifically, stations generating voice, real-time video, non-real-time video, and background data traffic are associated to Classes 1, 2, 3, and 4. All stations are equipped with the 802.11b Physical Layer (PHY) [58] using 11 Mbps channel data rate and 1 Mbps basic rate (slot time = 20 μs, SIFS = 10 μs, DIFS = 50 μs). The data frame payload, ACK frame payload, MAC header, and PHY header sizes are 3840, 112, 224, and 192 bits, respectively. The propagation delay is 2 μs and $CW_{\min}$=32, $CW_{\max}$ = 1024. The TXOP limits of the stations in Classes 1, 2, 3, and 4 are set to 4, 3, 2, and 1 (frame), respectively, for the purpose of QoS differentiation.

To demonstrate the effectiveness of the analytical model for the TXOP scheme in practical multimedia WLANs, the traffic parameters used in the validation are obtained from the accurate measurements of the real-world multimedia applications including the G.711 codec voice sources and the VBR encoded H.263 video streams. The details are given as follows:

Voice: The G.711 voice codec [22, 43, 77] is a popular international standard for VoIP encoding, which generates 480-byte encoded voice frames with a rate of 64

kbps given the sample period (*i.e.*, frame interval) of 60 ms. For a single ON-OFF voice source, the frame arrival rate of G.711 encoded voice traffic is $64000/(480*8) \approx 16.67$ (frames/sec) during talk spurts (in the ON state) and no frame arrives during silences (in the OFF state). The ON and OFF periods are exponentially distributed with the means of 352 ms and 650 ms, respectively [114]. The superposition of multiple ON-OFF voice sources can be modelled by a two-state MMPP [114] with the parameters that can be determined by Eqs. (2.2) – (2.4).

Video: The parameters of the video traffic are obtained from the real H.263 frame traces [40]. H.263 is an international video coding standard which has found many applications on wireless multimedia networks. It has been shown that the VBR encoded H.263 traces exhibit a high degree of long-range-dependence [40]. Burstiness of self-similar traffic has been modelled over five time scales since noticeable traffic bursts are typically present over four or five time scales [104].

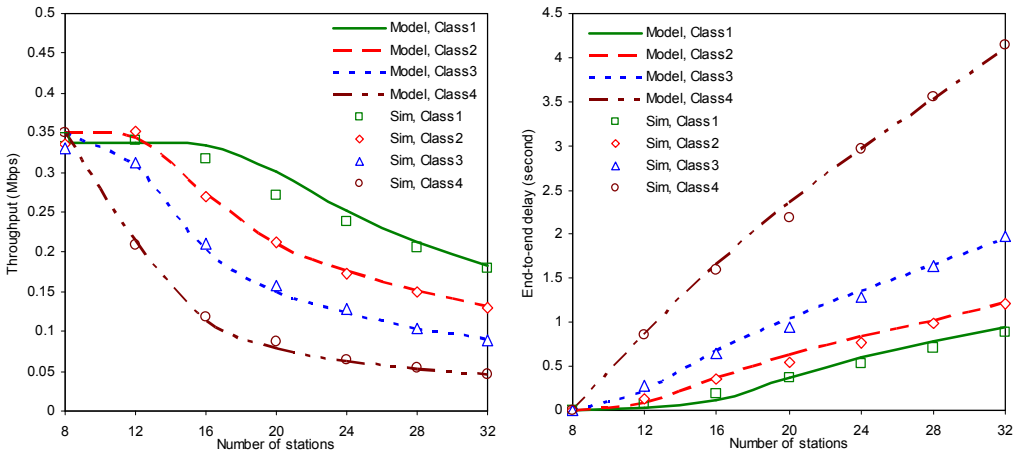
The accuracy of the analytical model has been validated under various working conditions and reached consistent performance conclusions. For the sake of illustration and without loss of generality, this section presents the verification results of two typical scenarios with the traffic parameters listed in Table 5.1. Figs. 5.2 and 5.3 depict the throughput, end-to-end delay, and frame loss probability for different classes of stations as a function of the network size in the two scenarios, respectively. The figures reveal a good degree of match between the analytical and simulation results, therefore validating the accuracy of the proposed model.

Table 5.1: Traffic Parameters for the Analysis of the TXOP Scheme under Multimedia Traffic

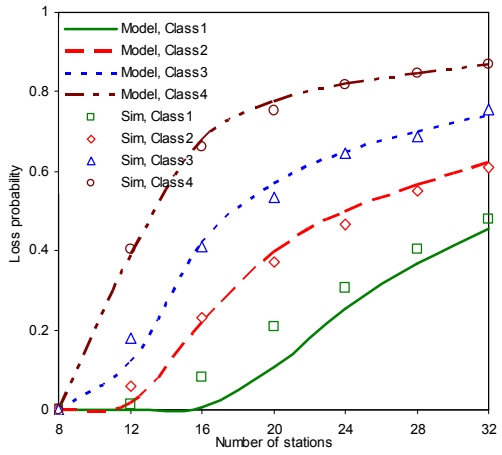
Scenarios	Voice	Video				Data
	Number of VoIP sources	H.263 Video traces	Hurst parameter	Autocorrelation at lag 1	Mean arrival rate (kbps)	Mean arrival rate (kbps)
1	15	Starship Troopers	0.844	0.82	350	350
2	25	Formula 1	0.840	0.75	530	530

5.3.2 Performance Evaluation

Figs. 5.2 and 5.3 also show that the stations with the larger TXOP limit can achieve the better performance than those with the smaller one. The results reveal the efficiency of using the TXOP scheme as a service differentiation method. It can be seen that the throughput, end-to-end delay, and frame loss probability of stations in different classes are rarely influenced by the TXOP scheme when the number of stations is less than a certain value (*e.g.*, 8 in Fig. 5.2), showing that TXOP has little impact on the network performance under light loads. However, as the number of stations increases, the station with the larger TXOP limit achieves the higher throughput, smaller end-to-end delay, and lower loss probability than those with the smaller TXOP limits. This observation clearly demonstrates the effects of the TXOP scheme.



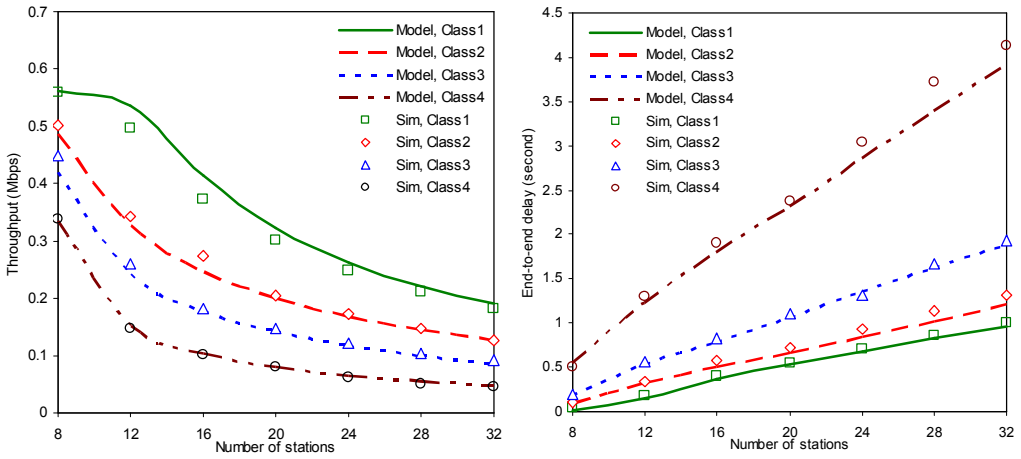
(a) Throughput **(b) End-to-end delay**



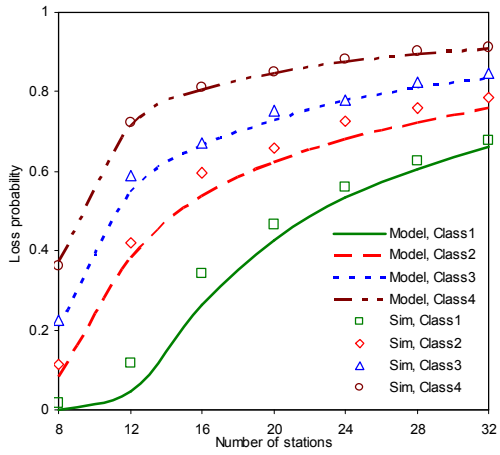
(c) Frame loss probability

Fig. 5.2: Performance measures versus the number of stations in Scenario 1.

To investigate the impact of heterogeneous traffic on the network performance, the throughput, end-to-end delay, and frame loss probability attained from the proposed model under different classes of stations are compared to those obtained from the existing model under the homogeneous stations where the traffic arriving from all stations is characterized by the Poisson process. The numerical results in Fig. 5.4 show the significant impact of traffic characteristics on the network performance. Specifically, in



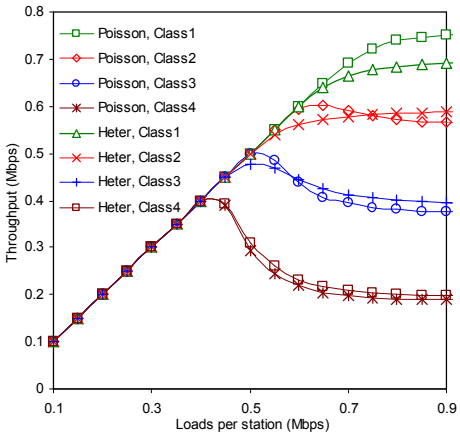
(a) Throughput **(b) End-to-end delay**



(c) Frame loss probability

Fig. 5.3: Performance measures versus the number of stations in Scenario 2.

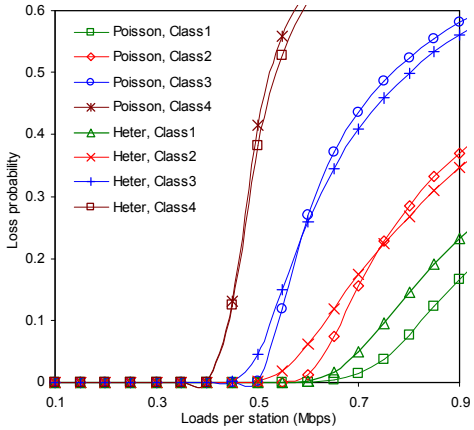
the presence of heterogeneous traffic, the performance of stations in Class 1 declines when the traffic loads per station exceed 0.65 Mbps, while the performance of stations in Class 4 is only slightly affected. For example, the throughput of stations in Class 1 is 8% smaller and their loss probability is 50% larger when the loads are 0.75 Mbps. For the stations in Classes 2 and 3, the performance degradation caused by heterogeneous traffic appears when the loads are between 0.55-0.65 and 0.45-0.55 Mbps, respectively.



(a) Throughput

*****

(b) End-to-end delay



(c) Frame loss probability

Fig. 5.4: Comparison of the analytical results between heterogeneous stations and homogeneous ones (Poisson Traffic). Two-state MMPP parameters: $\lambda_1 = 6\lambda_2$, $\sigma_1 = 0.2$/s, $\sigma_2 = 0.8$/s, Self-Similar parameters: H = 0.9, r(1) = 0.75, Network size: 12 stations.

As the traffic loads further increase, the network becomes heavily congested and the frame loss probability at these stations increases to an unacceptable level. The above observations stress the great need to develop a comprehensive analytical model that can quantitatively and accurately capture the impact of heterogeneous traffic on the design and performance of multimedia WLANs. Moreover, it is worth noting that the maximum throughput of stations in Classes 3 and 4 is much larger than their saturation throughput, which emphasizes the importance of analyzing the TXOP scheme under unsaturated traffic loads.

5.4 Summary

This chapter has proposed an analytical model for the TXOP scheme in the presence of heterogeneous stations with the non-bursty Poisson, bursty MMPP, and fractal self-similar traffic, respectively. QoS performance measures in terms of throughput, end-to-end delay, and frame loss probability have been derived. The accuracy of the proposed model has been validated through extensive NS-2 simulation experiments. The traffic parameters used in the validation were obtained from the accurate measurements of the real-world multimedia applications including the G.711 codec voice sources and the VBR encoded H.263 video streams. The results have shown the efficiency of adopting TXOP as a service differentiation scheme between heterogeneous stations. Moreover, the importance of considering the heterogeneous traffic for the accurate performance evaluation of the TXOP scheme in the presence of wireless multimedia applications has been revealed.

Chapter 6

Comprehensive QoS Analysis of the EDCA Protocol under Poisson Traffic

6.1 Introduction

The IEEE 802.11e Enhanced Distributed Channel Access (EDCA) is a promising Medium Access Control (MAC) protocol for provisioning of differentiated Quality-of-Service (QoS) in Wireless Local Area Networks (WLANs). This protocol specifies three important QoS differentiation schemes including Arbitrary Inter-Frame Space (AIFS), Contention Window (CW) and Transmission Opportunity (TXOP). Significant research efforts have been devoted to developing the analytical performance model for the QoS differentiation schemes specified in EDCA [3, 28, 36, 43, 52, 53, 55-57, 61, 67, 70, 78, 83, 96, 108, 109, 118-120, 122, 125, 131, 132, 134, 146]. The majority of these existing models were developed under the assumption of saturated working conditions where all the ACs have frames for transmission anytime [55-57, 70, 78, 108, 109, 118, 120, 125, 131, 132, 134, 146]. Since the traffic loads in the practical network environments are mainly unsaturated [81, 144], it is imperative to develop analytical models for EDCA in the

presence of unsaturated traffic conditions. However, the existing analytical models reported in the current literature have been primarily focused on the AIFS, CW, and TXOP schemes, separately [3, 28, 36, 43, 52, 67, 83, 93, 96, 119, 122]. In addition, these models are based on the assumptions that the MAC buffer has either a very small size [67, 119] or an infinite capacity [3, 28, 36, 83, 122]. These assumptions cannot capture the realistic working conditions of practical WLANs. With the aim of conducting a thorough and deep investigation of the QoS performance of EDCA, this chapter proposes a comprehensive analytical model to accommodate the combination of all three QoS schemes of EDCA under unsaturated traffic loads.

The comprehensive analytical model presented in this chapter is distinguishing from those reported in the current literature in various aspects. This model is able to incorporate the three QoS schemes of EDCA (*i.e.*, AIFS, CW and TXOP) simultaneously in WLANs with finite buffer capacity in the presence of unsaturated traffic loads. Unlike the model reported in [61], an approach combining the Markov chain and queueing theory is employed to analyze the backoff procedure and the burst transmission procedure of EDCA. As a result, the proposed analytical model holds the following advantages: 1) it can handle a large number of MAC buffer size without heavily increasing the complexity of the solution; 2) it can derive the delay jitter that is an important performance measure for delay-sensitive applications.

Moreover, the new 3D Markov chain proposed in this chapter to analyze the backoff procedure of EDCA is more general and can capture the behavior of EDCA more accurately than existing EDCA models based on a 3D Markov chain where the

third dimension denotes the number of remaining time slots to complete the AIFS period [57, 70, 118]. Because this proposed 3D Markov chain takes into account the AIFS differentiation, virtual collision, frame retry limit, and details of the backoff rule, which were not properly or comprehensively captured in existing models. For instance, the models reported in [118] were limited to the case where the difference between the minimum AIFS ($AIFS_{\min}$) and maximum AIFS ($AIFS_{\max}$) is one time slot only. In [70], the channel busy probability is not differentiated between the countdown and deferring periods of the backoff counter. Moreover, this model assumed that the backoff counter of an AC is immediately activated after the ACK timeout interval when the AC encounters a collision. However, in the IEEE 802.11e standard [59], the backoff counter is invoked if the channel is detected to be idle for the AIFS period after the ACK timeout interval. To capture this behavior, the 3D Markov chain proposed in this chapter transits to the deferring state when the backoff stage increases due to an unsuccessful transmission.

The remainder of this chapter is organized as follows. Section 6.2 elaborates on the derivation of the analytical model. Section 6.3 validates the accuracy of the model and conducts performance analysis. Section 6.4 summarizes this chapter.

6.2 Analytical Model

This section elaborates on the proposed analytical model for EDCA which can be used to evaluate the QoS performance metrics including throughput, end-to-end

delay, delay jitter, and frame loss probability. The ACs from the lowest to highest priority are denoted by subscripts 0, 1, 2, $\cdots, N$. The transmission queue at each AC is modelled as a bulk service queueing system where the arrival traffic follows a Poisson process with rate λ_v (frames/second, $v = 0, 1, 2, \cdots, N$). The service rate of the queueing system, μ_v, is derived by analyzing the backoff and burst transmission procedures of the AC_v. With λ_v and μ_v, the bulk-service queueing system can be solved to obtain the QoS performance measures of ACs in the IEEE 802.11e WLANs.

6.2.1 Modelling of the Backoff Procedure

This subsection presents a new 3D discrete-time Markov chain to analyze the backoff procedure of EDCA. In this Markov chain, a time slot is referred to as the variable time interval between the starts of two consecutive decrements of the backoff counter. To avoid any ambiguity, in what follows, the fixed time interval (unit time) specified in the IEEE 802.11e standard [59] is called the *physical time slot*. Let $s(t)$ and $b(t)$ denote the stochastic processes representing the backoff stage and the backoff counter for a given AC, respectively. The third dimension, $c(t)$, represents the number of the remaining time slots required to complete the AIFS period of the AC ($AIFS_v$) after $AIFS_{\min}$. The 3D process $\{s(t), b(t), c(t)\}$ can be modelled as a discrete-time Markov chain as shown in Fig. 6.1 (a), with a dashed line box shown in detail in the sub-Markov chain of Fig. 6.1 (b).

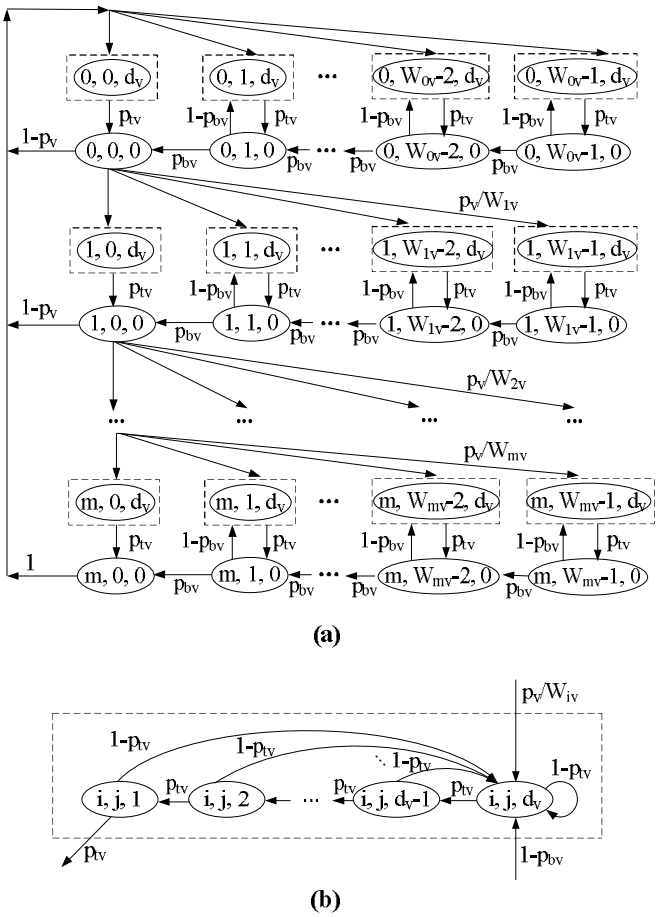


Fig. 6.1: The 3D Markov chain for modelling the backoff procedure of the AC_v.

The state transition probabilities of the 3D Markov chain are described as follows:

$$\begin{cases} P\{i,j,0 \mid i,j+1,0\} = p_{bv}, & 0 \le i \le m,\ 0 \le j \le W_{iv}-2 \quad (1.a) \\ P\{i,j,d_v \mid i,j,0\} = 1-p_{bv}, & 0 \le i \le m,\ 1 \le j \le W_{iv}-1 \quad (1.b) \\ P\{i,j,0 \mid i,j,1\} = p_{tv}, & 0 \le i \le m,\ 0 \le j \le W_{iv}-1 \quad (1.c) \\ P\{i,j,k \mid i,j,k+1\} = p_{tv}, & 1 \le k \le d_v-1 \quad (1.d) \\ P\{i,j,d_v \mid i,j,k\} = 1-p_{tv}, & 1 \le k \le d_v \quad (1.e) \\ P\{i,j,d_v \mid i-1,0,0\} = p_v/W_{iv}, & 1 \le i \le m,\ 0 \le j \le W_{iv}-1 \quad (1.f) \\ P\{0,j,d_v \mid i,0,0\} = (1-p_v)/W_{0v}, & 0 \le i \le m-1, 0 \le j \le W_{iv}-1 \quad (1.g) \\ P\{0,j,d_v \mid m,0,0\} = 1/W_{0v}, & 0 \le j \le W_{iv}-1 \quad (1.h) \end{cases} \tag{6.1}$$

where p_v denotes the collision probability of the Head-of-Burst (HoB) frame transmitted from the AC_v (note that only the HoB frame needs to contend for the channel), p_{bv} is the probability that the channel is idle in a time slot after the AIFS period of the AC_v, and p_{tv} is the probability that the channel is idle in a time slot during the AIFS period of the AC_v after $AIFS_{\min}$. d_v represents the difference in the number of physical time slots between $AIFS_{\min}$ and $AIFS_v$, *i.e.*, $d_v = AIFSN_v - AIFSN_{\min}$. m denotes the retry limit (i.e., the maximum backoff stage) and W_{iv} $(0 \le i \le m)$ accounts for the CW after i unsuccessful transmissions. W_{iv} is given by

$$W_{iv} = \begin{cases} 2^i W_{0v} & 0 \le i \le m' \\ 2^{m'} W_{0v} & m' < i \le m \end{cases} \tag{6.2}$$

where W_{0v} and $2^{m'} W_{0v}$ represent the $CW_{\min}$ and $CW_{\max}$ of the AC_v, respectively.

The equations of the state transition probabilities account, respectively, for: Eq. (6.1.a) The backoff counter is decreased by one after an idle time slot; (6.1.b) The backoff counter is frozen if sensing a busy channel; (6.1.c) The backoff counter is activated after the AIFS period; (6.1.d) The remaining number of time slots in the AIFS period is decreased by one if the channel is detected idle in a time slot; (6.1.e) The AC has to go through the AIFS period again if the channel is sensed busy during the AIFS period; (6.1.f) The backoff stage increases after an unsuccessful transmission and the AC defers for the AIFS period before activating the backoff

counter; (6.1.g) After a successful transmission, the CW is reset to $CW_{\min}$; (1.h) The CW is reset to $CW_{\min}$ if the retransmission attempts reach the retry limit.

Let $b_{i,j,k}$ be the stationary distribution of the 3D Markov chain with state $\{i, j, k\}$ where $i \in [0, m], j \in [0, W_{iv} - 1]$, and $k \in [0, d_v]$. First, the steady-state probabilities, $b_{i,0,0}$, satisfies

$$b_{i,0,0} = p_v^i \, b_{0,0,0} \qquad 0 \le i \le m \tag{6.3}$$

Because of the chain regularities, for each $j \in [0, W_{iv} - 1]$, $b_{i,j,0}$ satisfies

$$b_{i,j,0} = \frac{b_{i,0,0}(W_{iv} - j)}{W_{iv}} \qquad 0 \le i \le m \tag{6.4}$$

From the balance equations in the sub-Markov chain, the following relations exist

$$\begin{cases} b_{i,j,k} = b_{i,j,d_v} \; p_{tv}^{d_v - k}, & 1 \le k \le d_v \\ p_{tv} \, b_{i,j,d_v} = (1 - p_{bv}) \, b_{i,j,0} + \dfrac{p_v b_{i-1,0,0}}{W_{iv}} + \sum_{k=1}^{d_v - 1} b_{i,j,k}(1 - p_{tv}), & \\ \qquad\qquad 1 \le i \le m, \quad 0 \le j \le W_{iv} - 1 & \\ p_{tv} \, b_{0,j,d_v} = (1 - p_{bv}) \, b_{0,j,0} + \dfrac{(1 - p_v)\sum_{i=0}^{m-1} b_{i,0,0} + b_{m,0,0}}{W_{0v}} + \sum_{k=1}^{d_v - 1} b_{0,j,k}(1 - p_{tv}), & \\ \qquad\qquad 0 \le j \le W_{iv} - 1 & \end{cases} \tag{6.5}$$

With Eqs. (1), (3) and (4), $b_{0,0,0}$ can be finally determined by imposing the following normalization condition

$$1 = \sum_{i=0}^{m} \sum_{j=0}^{W_{iv} - 1} b_{i,j,0} + \sum_{i=0}^{m} \sum_{j=0}^{W_{iv} - 1} \sum_{k=1}^{d_v} b_{i,j,k} \tag{6.6}$$

Solving the 3D Markov chain, the initial state, $b_{0,0,0}$, is given by

$$b_{0,0,0} = \left[\frac{(1-p_{tv}^{d_v})}{(1-p_{tv})p_{tv}^{d_v}} \left((1-p_{bv}) \sum_{i=0}^{m} \frac{W_{iv}-1}{2} p_v^i + \frac{1-p_v^{m+1}}{1-p_v} \right) + \sum_{i=0}^{m} \frac{W_{iv}-1}{2} p_v^i + \frac{1-p_v^{m+1}}{1-p_v} \right]^{-1} \tag{6.7}$$

Note that for the ACs with the minimum AIFS period, d_v equals zero. Therefore, Eq. (6.7) can be reduced to

$$b_{0,0,0} = \left[\sum_{i=0}^{m} \frac{W_{iv}-1}{2} p_v^i + \frac{1-p_v^{m+1}}{1-p_v} \right]^{-1} \tag{6.8}$$

The probability that an AC_v transmits in a randomly chosen time slot, τ'_v, when its transmission queue is non-empty, is calculated as

$$\tau'_v = \sum_{i=0}^{m} b_{i,0,0} = b_{0,0,0} \sum_{i=0}^{m} p_v^i = \frac{b_{0,0,0}(1-p_v^{m+1})}{(1-p_v)} \tag{6.9}$$

Note that an AC can transmit only when there are pending frames in its transmission queue. Thus, the transmission probability, τ_v, that the AC_v transmits under unsaturated traffic conditions can be derived by

$$\tau_v = \tau'_v(1-P_{0v}) \tag{6.10}$$

where P_{0v} is the probability that the transmission queue of the AC_v is empty and will be derived in Section 4.3.

Taking virtual collision into account, an HoB frame transmitted from an AC will collide with HoB frames sent from other stations, or from the higher priority ACs in the same station. Therefore, the collision probability, p_v, of the AC_v is given by

$$p_v = 1 - \prod_{x=0}^{N} (1-\tau_x)^{n-1} \prod_{x>v}^{N} (1-\tau_x) \tag{6.11}$$

where n denotes the number of stations.

The probability, p_{bv}, that the channel is idle in a time slot after the AIFS period of the AC_v, is the probability that none of the other ACs is transmitting in the given slot

$$p_{bv} = (1-\tau_v)^{n-1} \prod_{x \neq v} (1-\tau_x)^n \tag{6.12}$$

During the AIFS period of the AC_v, the ACs with the priorities lower than or equal to AC_v cannot transmit. As a result, the probability p_{tv}, that the channel is sensed idle in a time slot during the AIFS period of the AC_v after $AIFS_{\min}$, is given by

$$p_{tv} = \prod_{x>v}^{N} (1-\tau_x)^n \tag{6.13}$$

6.2.2 Analysis of the Service Time

The service time is defined as the time interval from the instant that an HoB frame starts contending for the channel to the instant that the burst is acknowledged following successful transmission or the instant that the HoB frame is discarded due to transmission failures. The service time is composed of two parts: channel access delay and burst transmission delay. The former is the time interval from the instant that the HoB frame reaches to the head of the transmission queue, until it wins the

contention and is ready for transmission, or until it is discarded due to transmission failures. The latter is defined as the time duration of transmitting a burst. $E[S_{sv}]$, $E[A_v]$, and $E[B_{sv}]$ are denoted as the means of the service time, channel access delay, and burst transmission delay, respectively, where v represents that the burst is transmitted from an AC_v and s denotes the number of frames successfully transmitted within the burst. Similarly, let $E'[S_v]$, $E'[A_v]$, and $E'[B_v]$ denote the means of the service time, channel access delay, and burst transmission delay, respectively, when the HoB frame is discarded due to transmission failures. $E[B_{sv}]$ and $E'[B_v]$ can be expressed as

$$\begin{cases} E[B_{sv}] = AIFS_v + s(T_L + T_H + 2T_{SIFS} + T_{ACK}) - T_{SIFS} \\ E'[B_v] = 0 \end{cases} \tag{6.14}$$

where T_L and T_H denote the transmission time for the frame payload and frame head, respectively. Next, $E[A_v]$ and $E'[A_v]$ are given by

$$\begin{cases} E[A_v] = T_{cv} \sum_{i=0}^{m} \dfrac{i\, p_v^i (1-p_v)}{1-p_v^{m+1}} + \bar{\sigma}_v \sum_{i=0}^{m} \sum_{j=0}^{i} \dfrac{W_{jv}-1}{2} \dfrac{p_v^i (1-p_v)}{(1-p_v^{m+1})} \\ E'[A_v] = T_{cv}(m+1) + \bar{\sigma}_v \sum_{i=0}^{m} \dfrac{W_{iv}-1}{2} \end{cases} \tag{6.15}$$

where T_{cv} is the average collision time and $\bar{\sigma}_v$ is the average length of a time slot. p_v^i is the probability that the backoff counter reaches stage i, p_v^{m+1} represents the probability that the HoB frame is discarded due to multiple collisions, $(W_{iv}-1)/2$ denotes the mean of the backoff counters generated in the *i*-th backoff stage.

Let PT_v denote the probability that at least one of the remaining ACs transmits in a given time slot and PS_x represent the probability that an AC_x successfully transmits, respectively, provided that the AC_v is in the backoff procedure. PT_v and PS_x can be expressed as

$$PT_v = 1 - p_{bv} = 1 - (1-\tau_v)^{n-1} \prod_{x \neq v} (1-\tau_x)^n \tag{6.16}$$

$$PS_x = n\tau_x (1-\tau_v)^{n-2} \prod_{y \neq v} (1-\tau_y)^{n-1} \prod_{y>x}^{N} (1-\tau_y) \tag{6.17}$$

The average length of a time slot, $\overline{\sigma}_v$, when the AC_v is in the backoff procedure is obtained by considering the fact that the channel is idle with probability $(1-PT_v)$, a successful transmission from an AC_x occurs with probability PS_x, and a collision happens with probability $(PT_v - \sum_{x=0}^{N} PS_x)$. Thus, $\overline{\sigma}_v$ can be calculated as

$$\overline{\sigma}_v = (1-PT_v)\sigma + \sum_{x=0}^{N} PS_x T_{sx} + (PT_v - \sum_{x=0}^{N} PS_x) T_{cv} + E[X_v] PT_v \tag{6.18}$$

where σ is the duration of a physical time slot [59]. T_{sv} denotes the average time of a successful burst transmission from the AC_v and T_{cv} accounts for the average collision time, respectively. $E[X_v]$ represents the total time spent on deferring the AIFS period of the AC_v. Recall that an AC has to go through the AIFS period again if the channel is sensed busy during the deferring procedure, as shown in the sub-Markov chain of Fig. 6.1, $E[X_v]$ may consist of several attempts for deferring the AIFS period of the AC_v and can be given by

$$E[X_v] = \sum_{u=1}^{\infty} p_{tv}^{d_v} (1 - p_{tv}^{d_v})^{u-1} \, u T_{av} \tag{6.19}$$

where u is the number of attempts for deferring the AIFS period of the AC_v, T_{av} is the average time spent on each attempt, and $p_{tv}^{d_v}$ is the probability of an successful attempt. T_{av} is given by

$$T_{av} = \sum_{x>v}^{N} PS'_x T_{sx} + (PT' - \sum_{x>v}^{N} PS'_x) T_{cv} + \sum_{s=1}^{d_v - 1} p_{tv}^{s} s\sigma \tag{6.20}$$

where the first and second terms correspond to the "frozen time" of the backoff counter of the AC_v caused by the transmission from the higher priority ACs with the smaller AIFS. The third term is the average time spent on a failed attempt for down-counting the remaining time slots during the AIFS period of the AC_v after $AIFS_{\min}$. PT'_v and PS'_x denote the probabilities that at least one AC transmits and the AC_x successfully transmits in a given time slot, respectively, when the AC_v is deferring in the AIFS period. PT'_v and PS'_x can be written as

$$PT'_v = 1 - p_{tv} = 1 - \prod_{x>v}^{N} (1 - \tau_x)^n \tag{6.21}$$

$$PS'_x = n\tau_x \prod_{y>v}^{N} (1 - \tau_y)^{n-1} \prod_{y>\max\{x,v\}}^{N} (1 - \tau_y) \tag{6.22}$$

Note that only the HoB frame is involved in the collision, the collision time, T_{cv}, is given by

$$T_{cv} = T_L + T_H + T_{SIFS} + T_{ACK} + AIFS_v \tag{6.23}$$

The average time, T_{sv}, for a successful burst transmission from the AC_v can

be written as

$$T_{sv} = \frac{\sum_{s=1}^{F_v} E[B_{sv}]L_{sv}}{1 - P_{0v}} \tag{6.24}$$

where F_v denotes the maximum number of frames that can be transmitted in a TXOP limit of the AC_v, the denominator $(1 - P_{0v})$ means that the occurrence of burst transmission is conditioned on the fact that there is at least one frame in the transmission queue, L_{sv} is the probability that s ($1 \le s \le F_v$) frames are transmitted from the AC_v within a TXOP limit, and $E[B_{sv}]$ is the burst transmission delay given in Eq. (6.14).

6.2.3 Queueing Model

The transmission queue at the AC_v can be modelled as an M/G$^{[1,F_v]}$/1/K queueing system where the superscript $[1, F_v]$ denotes that the number of frames transmitted during a TXOP ranges from 1 to F_v, and K represents the system capacity.

The server becomes busy when a frame reaches to the head of the transmission queue. The server becomes free after a burst of frames are acknowledged by the destination following successful transmission, or after the HoB frame is dropped due to transmission failures. The service time is dependent on the number of frames transmitted within a burst and the class of the transmitting AC. Thus, the service time of a burst with s $(1 \le s \le F_v)$ frames successfully transmitted from the AC_v can be modelled by an exponential distribution function with mean $E[S_{sv}]$, then

the mean service rate, μ_{sv}, is given by $1/E[S_{sv}]$. When the HoB frame is discarded due to transmission failures, the mean service rate, μ'_v, is given by $1/E'[S_v]$.

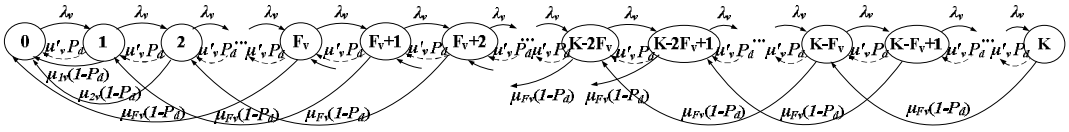


Fig. 6.2: The state-transition-rate diagram of the M/G$^{[1,Fv]}$/1/K queueing system.

Fig. 6.2 illustrates the state-transition-rate diagram of the queueing system of the AC_v where each state denotes the number of frames in the system. The transition rate from state r to state $r+1$ $(0 \le r \le K-1)$ is the arrival rate λ_v of the Poisson traffic. A transition from state r to state $r-1$ $(1 \le r \le K)$ implies that the HoB frame is dropped due to transmission failures with the probability P_d, which is given by p_v^{m+1}, and the transition rate is $\mu'_v P_d$. Then a transition out of state r to state $r-F_v$ $(F_v \le r \le K)$ represents that the burst transmission of F_v frames completes and the transition rate is $\mu_{F_v}(1-P_d)$. The change from state r to state 0 $(1 \le r \le F_v - 1)$ denotes that all r frames in the queueing system are transmitted within a burst and the transition rate is $\mu_{rv}(1-P_d)$.

The transition rate matrix, $\mathbf{G}_v$, of the Markov chain can be obtained from Fig. 6.2. The steady-state probability vector, $\mathbf{P}_v = (P_{rv}, r = 0, 1, \ldots, K)$, of the Markov chain satisfies the following equations

$$\mathbf{P}_v \mathbf{G}_v = 0 \quad \text{and} \quad \mathbf{P}_v \mathbf{e} = 1 \tag{6.25}$$

where $\mathbf{e}$ is a unit column vector. After obtaining $\mathbf{P}_v$, the probability that s $(1 \le s \le F_v)$ frames are transmitted from the AC_v within a TXOP limit, L_{sv}, can be given by

$$\begin{cases} L_{sv} = P_{sv}, & 1 \le s < F_v \\ L_{sv} = \sum_{r=F_v}^{K} P_{rv}, & s = F_v \end{cases} \tag{6.26}$$

The frame loss probability is defined as the probability that an arriving frame finds the finite buffer full, which is given by P_{Kv}. The throughput, TH_v, of the AC_v can be computed by

$$TH_v = \lambda_v E[P](1 - P_{Kv})(1 - p_v^{m+1}) \tag{6.27}$$

where $E[P]$ is the frame payload length and p_v^{m+1} is the probability that the frame is dropped due to $(m+1)$ transmission failures.

The end-to-end delay, which consists of queueing delay and service time, is the time interval from the instant that the frame enters the transmission queue to the instant that the frame is acknowledged after its successful reception. Let T_{Dv} denote the random variable for the end-to-end delay of the frame in AC_v and $P_{Dv}(t)$ denote its Cumulative Distribution Function (CDF). When an arriving frame enters the transmission queue and finds that r frames are in the queueing system, to ensure that the service of this frame completes during the time interval $[0, t]$, all $(r+1)$ frames including itself must have been served by time t. Therefore, $P_{Dv}(t)$, can be expressed as

$$P_{Dv}(t) = \Pr\{T_{Dv} \le t\} = \sum_{r=0}^{K-1} \left[\begin{array}{l} \Pr\{\, r+1 \text{ completions in } t \mid \text{arrival finds } r \text{ in the} \\ \text{system of } AC_v \,\} \cdot P'_{rv} \end{array} \right] \qquad (6.28)$$

where P'_{rv} is the probability that an arriving frame enters the transmission queue and finds r frames in the queueing system of AC_v. When the arriving frame finds the queueing system full, it will be dropped from the system. Therefore, P'_{rv}, is given by

$$P'_{rv} = \frac{P_{rv}}{1 - P_{Kv}}, \qquad r \le K-1 \qquad (6.29)$$

where P_{rv} is the probability that there are r frames in the queueing system of AC_v upon the arrival of the new frame. P_{rv} is identical to the steady state distribution of system size since the input process is Poisson.

Since the dropped frames due to multiple collisions are excluded in the calculation of the end-to-end delay and the transmission queue of AC_v is modelled as a bulk service queueing system, $(r+1)$ frames are served in r' ($r' = \lceil (r+1)/F_v \rceil$) bursts where each of $\lfloor (r+1)/F_v \rfloor$ bursts has F_v frames and the last burst has $\left[(r+1) - \lfloor (r+1)/F_v \rfloor \cdot F_v \right]$ frames. Due to the memoryless nature of the service time of individual bursts, the distribution of the service time required for r' bursts is independent of the arrival time of the frame and is the convolution of r' exponential random variables, which is a hypo-exponential distribution [21]. Specifically, let X_{iv} $(1 \le i \le r')$ represent independent exponential random variables with respective rates μ_v^i $(1 \le i \le r')$, random variable $\sum_{i=1}^{r'} X_{iv}$ follows a hypo-exponential distribution. From Eq. (6.28), $P_{Dv}(t)$ can be given by

$$P_{Dv}(t) = \sum_{r=0}^{K-1} P'_{rv} \int_0^t f_{rv}(x)dx \tag{6.30}$$

where $f_{rv}(x)$ is the probability density function (pdf) of the hypo-exponential distribution with r' ($r' = \lceil (r+1)/F_v \rceil$) states. As a special case of the phase-type distribution [21], the hypo-exponential distribution can be written in the form of a phase-type distribution with the representation $(\mathbf{S}, \boldsymbol{\alpha})$ where $\mathbf{S}$ is a $r' \times r'$ matrix with diagonal elements $-\mu_v^i$ $(1 \le i \le r')$ and super-diagonal elements μ_v^i and $\boldsymbol{\alpha}$ denotes the probability that the service process starts in each of the r' state. $\mathbf{S}$ and $\boldsymbol{\alpha}$ are given by

$$\mathbf{S} = \begin{bmatrix} -\mu_v^i & \mu_v^i & & & & & \\ & -\mu_v^i & \mu_v^i & & & & \\ & & \ddots & \ddots & & & \\ & & & -\mu_v^i & \mu_v^i & & \\ & & & & -\mu_v^i & \mu_v^i & \\ & & & & & -\mu_v^i \end{bmatrix}, \qquad \boldsymbol{\alpha} = (1, 0, \ldots, 0) \tag{6.31}$$

μ_v^i is given by

$$\mu_v^i = \begin{cases} \mu_{Fv}, & 1 \le i < r' \text{ and } r' > 1 \\ \mu_{sv}, & i = r', s = \left[(r+1) - \lfloor (r+1)/F_v \rfloor \cdot F_v \right] \end{cases} \tag{6.32}$$

where $\mu_{sv}\,(1 \le s \le F_v)$ is the service rate of the burst with s frames transmitted from AC_v.

Then the pdf, $f_{rv}(x)$, of the hypo-exponential distribution can be expressed as [21]

$$f_{rv}(x) = -\boldsymbol{\alpha} \exp(x\mathbf{S})\mathbf{S}\mathbf{e} \tag{6.33}$$

where $\mathbf{e}$ is a unit column vector of the size r' and $\exp(\mathbf{S})$ denotes the matrix exponential of $\mathbf{S}$.

Table 6.1: System Parameters for the Analysis of the EDCA protocol under Poisson Traffic

Frame payload	8000bits	PHY header	192bits
MAC header	224bits	ACK	112bits + PHY header
Date rate	11Mbit/s	Basic rate	1Mbit/s
Slot time	20μs	Buffer size	50 frames
SIFS	10μs	Retry limit	7

Table 6.2: EDCA Parameters for the Analysis of the EDCA Protocol under Poisson Traffic

Scenarios	ACs	AIFSN	CW_{min}	CW_{max}	TXOP
Scenario 1	AC0	6	32	1024	1 Frame
	AC1	2	32	512	1 Frame
	AC2	2	16	256	4 Frames
	AC3	2	8	128	2 Frames
Scenario 2	AC0	7	64	512	1 Frame
	AC1	4	32	512	1 Frame
	AC2	2	16	256	2 Frames
	AC3	2	16	256	4 Frames

With the CDF of the end-to-end delay, $P_{Dv}(t)$, its mean and standard deviation (jitter) can be given by

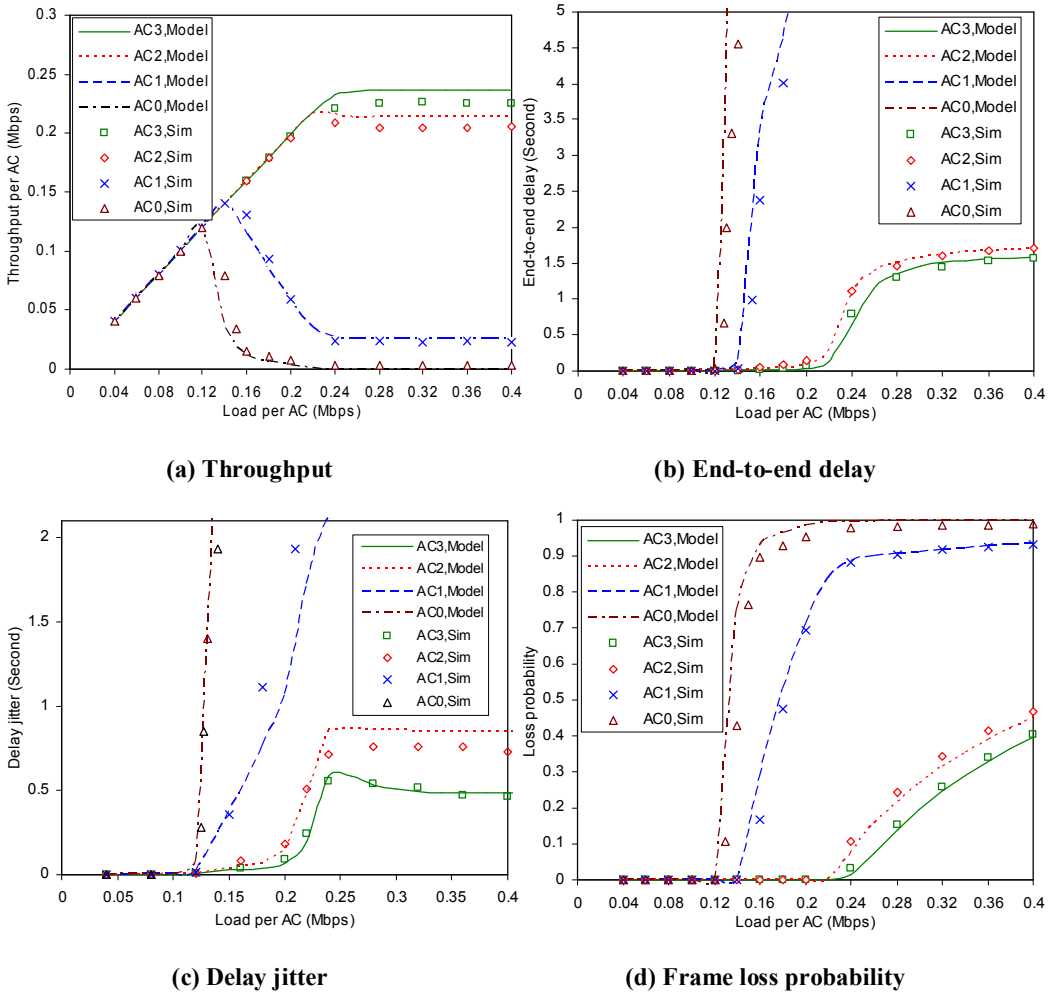
$$E[D_v] = \int_0^\infty t\, dP_{Dv}(t) \tag{6.34}$$

$$\sigma[D_v] = \sqrt{\int_0^\infty t^2\, dP_{Dv}(t) - E[D_v]^2} \tag{6.35}$$

6.3 Model Validation and Performance Evaluation

This section firstly investigates the accuracy of the proposed analytical model through extensive NS-2 simulation experiments and then uses the model to evaluate

the performance of three QoS differentiation schemes in EDCA.



(a) Throughput **(b) End-to-end delay**

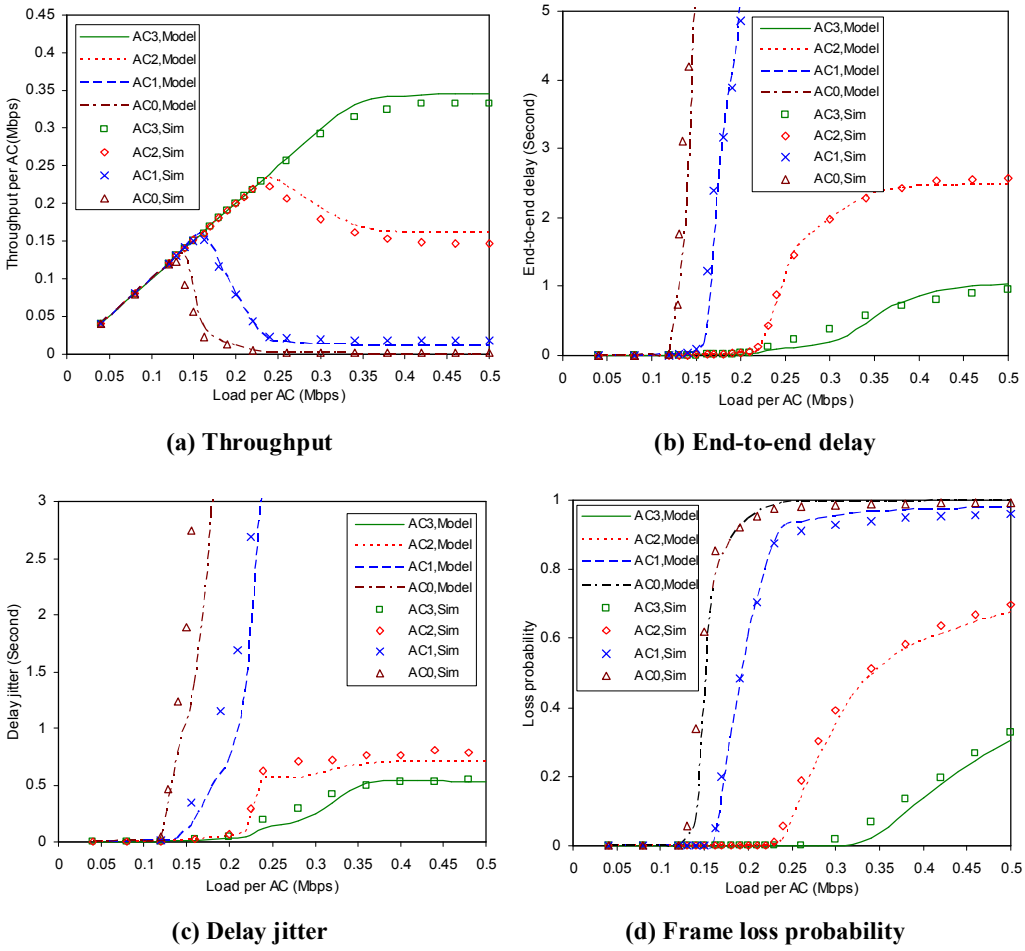
(c) Delay jitter **(d) Frame loss probability**

Fig. 6.3: Performance metrics versus the offered loads per AC in Scenario 1.

6.3.1 Model Validation

To validate the accuracy of this model, the analytical performance results are compared against those obtained from the NS-2 simulation experiments of the IEEE 802.11e EDCA [129]. A BSS with 10 stations is considered. Each station has four ACs, which are denoted by subscripts 0, 1, 2, and 3 from the lowest to highest priority. The packet arrivals at each AC are characterized by a Poisson process. The

system parameters are summarized in Table 6.1.



(a) Throughput **(b) End-to-end delay**

(c) Delay jitter **(d) Frame loss probability**

Fig 6.4: Performance metrics versus the offered loads per AC in Scenario 2.

To investigate the accuracy of the model under various working conditions, two scenarios with the different combinations of EDCA parameters are considered, as shown in Table 6.2. Figs. 6.3 and 6.4 depict the results of throughput, end-to-end delay, delay jitter, and frame loss probability versus the offered loads per AC in Scenarios 1 and 2, respectively. The close match between the analytical results and those obtained from simulation experiments demonstrates that the proposed model can produce the accurate performance results of EDCA with AIFS, CW and TXOP under arbitrary traffic loads. Moreover, as shown in Fig. 6.3, it is worth mentioning

that the maximum throughputs of AC1 and AC0 are much larger than their saturation throughput (the same phenomenon can be found in Fig. 6.4 with AC2, AC1, and AC0). This observation emphasizes the importance of analyzing the EDCA protocol under unsaturated traffic loads.

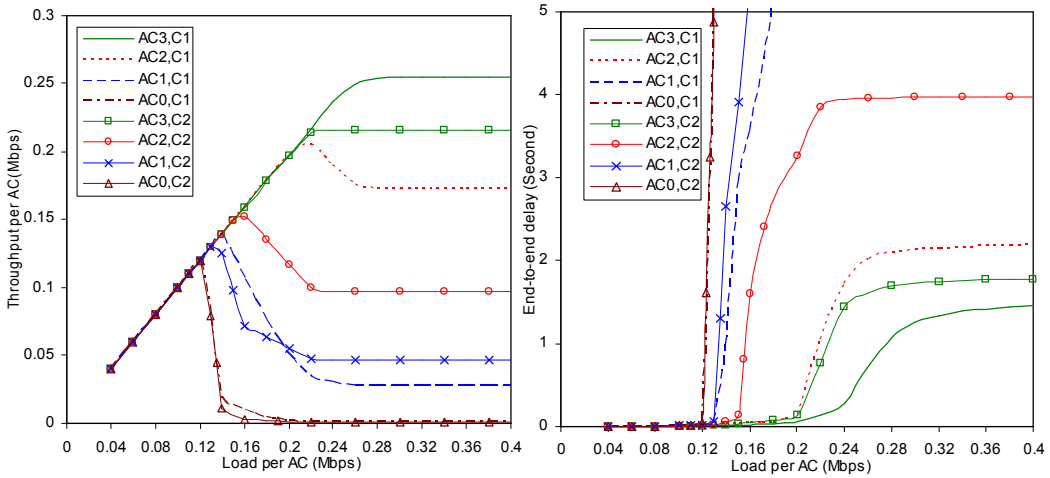
6.3.2 Performance Evaluation

This subsection uses the proposed analytical model to investigate the effects of the TXOP scheme and the buffer size on the QoS performance of EDCA.

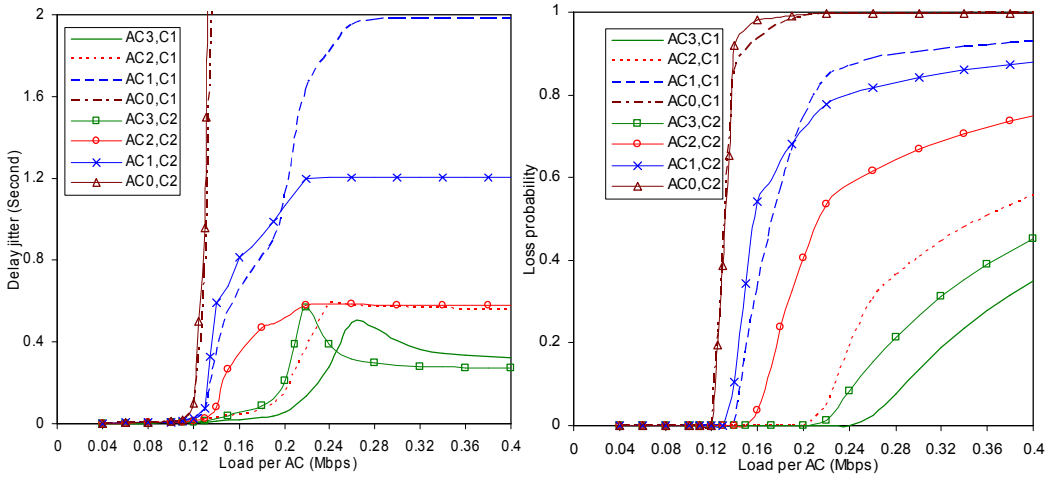
6.3.2.1 Effects of the TXOP Scheme

The existing studies on performance analysis of EDCA have been mainly focused on the AIFS, CW, and TXOP schemes, separately. Using the proposed analytical model, it is able to conduct a comprehensive evaluation of the QoS differentiation schemes in EDCA integrating the three schemes. Fig. 6.5 plots the curves of the throughput, end-to-end delay, delay jitter, and frame loss probability versus the offered loads per AC with different settings of TXOP limits. For the purpose of comparison, two distinct cases are considered: 1) the TXOP limits of AC_3, AC_2, AC_1, and AC_0 in Case 1 are set to 2, 3, 1, and 1 (frame), respectively; 2) the TXOP limits of all ACs in Case 2 are identical and equal to 1 (frame), which means that the TXOP scheme is not enabled. All the other EDCA parameters are the same as those used in Scenario 1, as shown in Table 6.2. For clarity of performance analysis, the offered loads per AC (from 0 to 0.4 Mbps) are divided into three regions where 0-0.12 Mbps is the light load region, 0.12-0.22 Mbps is the medium

load region, and 0.22-0.4 Mbps is the heavy load region.



(a) Throughput **(b) End-to-end delay**

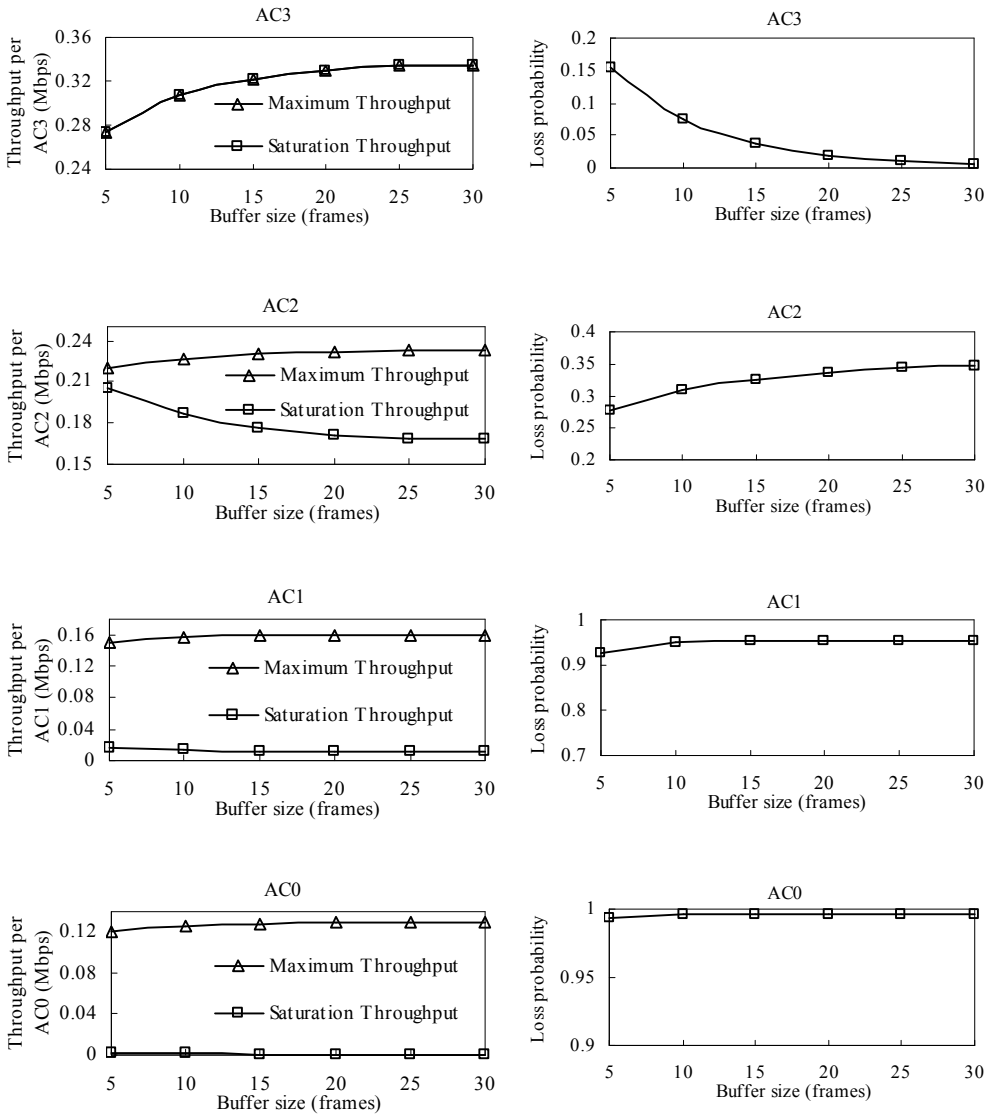


(c) Delay jitter **(d) Frame loss probability**

Fig 6.5: Performance comparison of EDCA between two distinct cases. Case 1 (C1): EDCA with the AIFS, CW, and TXOP schemes; Case 2 (C2): EDCA with the AIFS and CW schemes.

Under the light load region, the TXOP scheme has little impact on the performance of the ACs as the throughput, end-to-end delay, delay jitter, and frame loss probability of the ACs are almost the same in both cases. However, it can be observed that the TXOP scheme has a significant impact on the performance of the ACs under the medium and heavy load regions. Specifically, compared with Case 2,

the throughput, end-to-end delay, and frame loss probability of AC_3 and AC_2 ameliorate largely while those of AC_1 and AC_0 deteriorate in Case 1 when the ACs are under the heavy load region. However, different from the other performance metrics, the delay jitter of AC_3 increases while that of AC_2 is almost the same in Case 1 compared to those in Case 2. The reason is that bulk service can increase the delay jitter of a frame when the finite buffer queueing system is under heavy loads. It is interesting to note that the QoS of all the ACs in Case 1 is better than those in Case 2 under the medium load region. The QoS of AC_3 and AC_2 is improved due to the burst transmission in these ACs. However, the impact of burst transmission of AC_3 and AC_2 over the performance of AC_1 and AC_0 depends on the traffic load region. Under the medium load region, burst transmission of AC_3 and AC_2 may make their transmission queues empty frequently and thus reduces the number of contending ACs, which leads to the lower collision probabilities for AC_1 and AC_0. However, under the heavy load region, the burst transmission of AC_3 and AC_2 cannot reduce the number of contending ACs since the transmission queues of AC_3 and AC_2 are always backlogged. Moreover, the burst transmission of AC_3 and AC_2 enables them to obtain long channel occupation time. As a result, the performance of AC_1 and AC_0 is degraded. The above observations demonstrate that the TXOP scheme can not only provide service differentiation like AIFS and CW, but also improve the system performance.



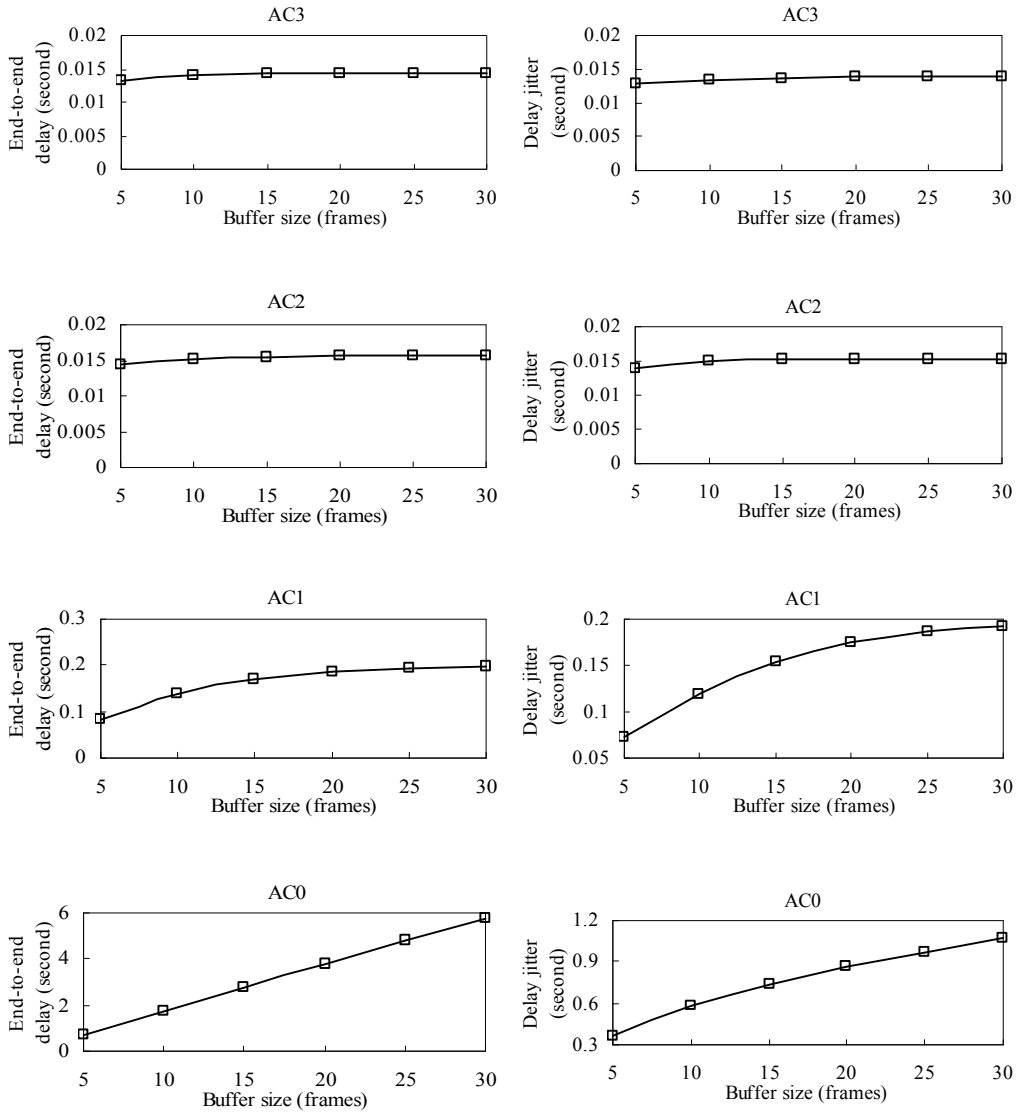
(a) Throughput **(b) Frame loss probability**

Fig. 6.6: Analytical results against the buffer size: (a) throughput; (b) loss probability.

6.3.2.2 Effects of the buffer size

Most existing models for EDCA under unsaturated traffic conditions assumed that the MAC buffer has either a very small size [67, 119] or an infinite capacity [3, 28, 36, 83, 122], because the assumption of a small buffer could avoid considering queuing dynamics while the infinite buffer assumption could reduce the difficulty of

developing the queuing model. However, these unrealistic assumptions, as to be shown later, can cause considerable inaccuracies of analytical results. Using the proposed model, the impact of the buffer size on the QoS performance of EDCA in terms of throughput, end-to-end delay, delay jitter, and frame loss probability is revealed. Fig. 6.6 depicts the performance of each AC as a function of the buffer size with the EDCA parameters of Scenario 2, as shown in Table 6.2. Fig. 6.6(a) plots the maximum throughput and saturation throughput of each AC with different buffer sizes. In Fig. 6.6(b), the loss probability performance is presented with the traffic loads per AC of 0.3 Mbps. The end-to-end delay and delay jitter are also shown in Fig. 6.6(c) and (d) respectively, when the traffic loads per AC are at 0.16 Mbps. It can be observed that the throughput (both the maximum and saturation throughput) and loss probability of AC3 and AC2 vary largely when the buffer size increases from 5 to 20 and then change little as the buffer size further augments. In contrast, for the AC1 and AC0, the throughput and loss probability are hardly affected by the buffer size. On the other hand, the end-to-end delay and delay jitter of AC1 and AC0 soar as the buffer size increases while those of AC3 and AC2 only slightly change with the buffer size. These results clearly demonstrate that the MAC buffer size has a significant impact on the QoS performance of ACs in the IEEE 802.11e WLANs.



(c) End-to-end delay **(d) Delay jitter**

Fig. 6.7: Analytical results against the buffer size: (c) End-to-end delay; (d) Delay jitter

6.4 Summary

This chapter has developed a comprehensive analytical model to accommodate the QoS differentiation schemes in terms of AIFS, CW, and TXOP specified in the IEEE 802.11e EDCA protocol with finite buffer capacity and unsaturated traffic

loads. This model employs an approach combining the Markov chain and queueing theory to analyze the backoff procedure and the burst transmission procedure in EDCA. The QoS performance measures including throughput, end-to-end delay, delay jitter, and frame loss probability have been derived and further validated through extensive NS-2 simulation experiments. The analytical model has been used for performance analysis of EDCA. Numerical results have shown that, the TXOP scheme can not only support service differentiation between various ACs but also improve the network performance, whereas the AIFS and the CW schemes provide QoS differentiation only. The effects of the MAC buffer size on the QoS performance of EDCA have been investigated. Analytical results have revealed that the buffer size has considerable impact on the QoS performance of EDCA. The comprehensiveness, efficiency, and accuracy of the proposed analytical model make it a cost-effective tool for the performance evaluation of the IEEE 802.11e EDCA protocol.

Chapter 7

Performance Analysis of the EDCA Protocol under Heterogeneous Multimedia Traffic

7.1 Introduction

WLANs are currently integrating a diverse range of traffic sources, such as video, voice and data, which significantly differ in their traffic patterns as well as QoS requirements. For instance, the VBR real-time applications, such as VoIP and video, are sensitive to delay and delay jitter and their traffic flows tend to exhibit Short-Range Dependent (SRD) bursty [114] and Long-Range Dependent (LRD) self-similar properties [15, 40], respectively. Therefore, it is critical to take the heterogeneous characteristics of multimedia applications into account in order to accurately evaluate the IEEE 802.11e EDCA protocol. As a step towards this end, this chapter proposes the first analytical model for EDCA in the presence of heterogeneous traffic.

The rest of this chapter is organized as follows. In Section 7.2, analytical model for EDCA in the presence of heterogeneous traffic is developed. Section 7.3 validates the accuracy of the model and conducts performance evaluation. Section

7.4 presents a game-theoretical admission control scheme based on the analytical model. Section 7.5 concludes the chapter.

7.2 Analytical Model

This section presents an analytical model for evaluating the QoS performance metrics of the IEEE 802.11e EDCA protocol in the presence of heterogeneous traffic. For convenience, the ACs from the lowest priority to the highest one are denoted by subscripts 0, 1, 2, $\cdots, c$. The heterogeneous SRD non-bursty Poisson, bursty ON-OFF, and LRD self-similar process are jointly adopted to model the traffic of wireless multimedia applications including background data, voice, and video traffic, respectively.

7.2.1 Modelling the Backoff Procedure

In order to derive the queueing delays and frame loss probabilities, the backoff procedure of the EDCA protocol with CW differentiation is firstly analyzed. In the following, the term time slot denotes the time interval between the starts of two consecutive decrements of the backoff counter, while the term physical time slot represents a fixed time interval (unit time) specified in the protocol [58]. The probability that an AC_i transmits in a randomly chosen time slot, τ_i', given that its transmission queue is non-empty, can be given by [130]

$$\frac{2(1-p_i)(1-p_i^{m+1})}{2W_{0i}\left(1-p_i-p_i(2p_i)^{m'}+2^{m'}p_i^{m+1}(2p_i-1)\right)+(1-p_i)(1-p_i^{m+1})} \tag{7.1}$$

where p_i is the collision probability of frames transmitted from the AC_i, m' is the maximum backoff stage and m is the retry limit.

Taking into account virtual collision, an AC will only collide with frames from other stations, or from the higher priority ACs in the same station. Therefore the collision probability, p_i, of AC_i is given by

$$p_i = 1 - \prod_{x \le i} (1-\tau_x)^{n-1} \prod_{x>i} (1-\tau_x)^n \tag{7.2}$$

where n is the number of stations.

Note that an AC can transmit only when there are pending frames in its transmission queue. Therefore the transmission probability, τ_i, that the AC_i transmits under the unsaturated traffic loads can be derived by

$$\tau_i = \tau_i'(1-P_{0i}) \tag{7.3}$$

where P_{0i} is the probability that the transmission queue of the AC_i is empty and will be given in the following subsections.

Let PT denote the probability that at least one AC transmits in a time slot. PT is given by

$$PT = 1 - \prod_{x=0}^{c} (1-\tau_x)^n \tag{7.4}$$

The probability, PS_i, that the AC_i successfully transmits can be expressed as

$$PS_i = n\tau_i \prod_{x \le i} (1-\tau_i)^{n-1} \prod_{x>i} (1-\tau_i)^n \tag{7.5}$$

Since the channel is idle with probability $(1-PT)$, a successful transmission from the AC_x $(0 \le x \le c)$ occurs with probability PS_x, a collision happens with probability $(PT - \sum_{x=0}^{c} PS_x)$, The average length of a time slot, σ_i', can be expressed as

$$\sigma_i' = (1-PT)\sigma + \sum_{x=0}^{c} PS_x T_{sx} + (PT - \sum_{x=0}^{c} PS_x) T_{ci} \tag{7.6}$$

where σ is the length of a physical time slot. The average time, T_{si}, for a successful frame transmission from the AC_i and the collision time T_{ci} can be expressed as

$$T_{si} = T_{ci} = T_L + T_H + T_{SIFS} + T_{ACK} + AIFS_i + 2\Delta \tag{7.7}$$

where Δ is the propagation delay.

The service time of the queueing system is defined as the time interval from the instant that a frame starts contending for the channel to the instant that the frame is acknowledged following successful transmission or the instant that the frame is dropped due to transmission failures. Let $E[S_i]$ denote the mean of the service time of the frame in AC_i. $E[S_i]$ is given by

$$E[S_i] = T_{ci}\varphi_i + \sigma_i' \delta_i + T_{si} \tag{7.8}$$

where φ_i and δ_i account for the average number of collisions and the average number of backoff counter decrements before a successful transmission from the AC_i, respectively. φ_i and δ_i can be computed as

$$\varphi_i = \sum_{h=0}^{m} h(1-p_i)p_i^h \tag{7.9}$$

$$\delta_i = \sum_{h=0}^{m} p_i^h W_{hi} / 2 \tag{7.10}$$

where p_i^h is the probability that the backoff counter reaches stage h, and $W_{hi}/2$ is the average value of the backoff counter generated in the *h*-th backoff stage.

7.2.2 Queueing Models for Heterogeneous ACs

To derive the performance metrics of frame delay and loss probabilities, the transmission queues of various ACs are modelled as different queueing systems. For the queueing system of AC_i, the service time is modelled by an exponential distribution function with mean $E[S_i]$. Thus, the mean service rate, μ_i, is given by $1/E[S_i]$. Many existing studies [83, 106, 140] on modelling DCF and EDCA assumed that the frame service time follows an exponential distribution and showed that this assumption is suitable. In what follows, the details of the queueing systems for the ACs with SRD non-bursty Poisson traffic, bursty ON-OFF traffic, and LRD self-similar traffic will be presented, respectively.

7.2.2.1 ACs with SRD Non-Bursty Traffic

The transmission queue at the ACs with non-bursty Poisson traffic is modelled as an M/M/1/K queueing system, where K represents the system capacity. With the arrival rate λ, the service rate μ, and the system size *K*, the probability, P_s, that there are s frames in the queueing system can be given by [117]

$$P_s = \frac{1-(\lambda/\mu)}{1-(\lambda/\mu)^{K+1}}(\lambda/\mu)^s \tag{7.11}$$

The frame loss probability is equal to the probability that the queueing system becomes full, *i.e.*, P_K.

7.2.2.2 ACs with SRD Bursty Traffic

The transmission queue at the ACs with bursty traffic is modelled as an ON-OFF/M/1/K queueing system. The probability density function (pdf), $f(t)$, of the inter-arrival time of the ON-OFF traffic can be given by

$$f(t) = \begin{cases} 0 & t < T \\ \alpha & t = T \\ (1-\alpha)v_{off} \exp(-v_{off}(t-T)) & t > T \end{cases} \tag{7.12}$$

where $T = 1/\lambda_p$ and $\alpha = 1 - v_{on}/\lambda_p$.

Then the G/M/1/K queueing model with the inter-arrival time distribution shown in Eq. (7.12) can be used to solve the ON-OFF/M/1/K queueing system. Specifically, this solution will utilize the well-developed solutions of the M/G/1/K queueing system [117] and resort the relationship between the G/M/1/K queue and the M/G/1/K+1 queue [12, 136]. Given a G/M/1/K queue which has the service rate μ, mean arrival rate λ, and pdf of the inter-arrival time distribution $f(t)$. Let P_s denote the steady-state probability of having s frames in the G/M/1/K queueing system. P_s is given by [136]

$$P_s = \frac{P'_{K+1-s}}{1 - P'_0} \tag{7.13}$$

where P'_s is the steady-state probability of having s frames in an M/G/1/K+1

queueing system with the arrival rate $\lambda' = \mu$, mean service rate $\mu' = \lambda$, and the pdf of the service time distribution $f(t)$.

Let π_s represent the steady-state probability of having s frames in the G/M/1/K queueing system as seen by an arbitrary arriving frame. Let π'_s denote the steady-state probability that there are s frames left in the M/G/1/K+1 queueing system immediately after service completion. Then the following relationship between the G/M/1/K queueing system and the M/G/1/K+1 queueing system exists [12]

$$\pi_s = \pi'_{K-s} \tag{7.14}$$

Thus, the loss probability of the G/M/1/K queueing system, π_K, is given by π'_0.

Let a_s represent the probability that there are s frames arriving during the service process of a frame in an M/G/1/K+1 queueing system. a_s is given by [117]

$$a_s = \int_0^\infty \frac{(\mu t)^s}{s!} \exp(-\mu t) f(t) dt \tag{7.15}$$

and π'_s $(0 \le s \le K)$ satisfies the following set of equations

$$\sum_{s=0}^{K} \pi'_s = 1 \tag{7.16}$$

$$\pi'_s = \pi'_0 a_s + \sum_{j=1}^{s+1} \pi'_j a_{s-j+1} \qquad 0 \le s \le K\text{-}1 \tag{7.17}$$

After solving Eqs. (7.16) and (7.17), P'_s can be obtained as [117]

$$\begin{cases} P'_s = \dfrac{\pi'_s}{\pi'_0 + \rho} & 0 \le s \le K \\ P'_s = 1 - \dfrac{1}{\pi'_0 + \rho} & s = K+1 \end{cases} \tag{7.18}$$

After obtaining the expression of P'_s, P_s can be readily computed using Eq. (7.13).

7.2.2.3 ACs with LRD Self-similar Traffic

Many models, such as chaotic maps and Fractional Brownian Motion [104], have been proposed to capture traffic self-similarity and long-range dependence. However, queueing theoretical techniques developed in the past are hardly applicable for these models. A method has been proposed in [110] to model the LRD self-similar process using a superposition of N independent streams of ON/OFF type with Truncated Power-Tail (TPT) distributed ON periods, called N-Burst traffic model. The parameters of the N-Burst traffic are determined so as to match the mean and Hurst parameter of the self-similar process. This simplified modeling method enables the existing analytical tools of Markov models to be available for deriving the desired performance measures.

The N-Burst arrival process is a superposition of N independent, identical ON/OFF sources as shown in Fig. 7.1. Each source generates frames with Poisson rate λ_p during ON-period, and the source is idle during OFF-period. The ON-period follows the Matrix Exponential (ME) distribution [110] while the OFF-period distribution is exponential. The structure of the ME distribution is

described by its entry vector **w** and rate matrix **B**. It is shown that the ON-period distribution has a great impact on the performance measures of queueing models. The N-Burst model shows asymptotically second-order self-similarity when Power-Tail (PT) distributions with infinite variance are employed for the ON-period distribution.

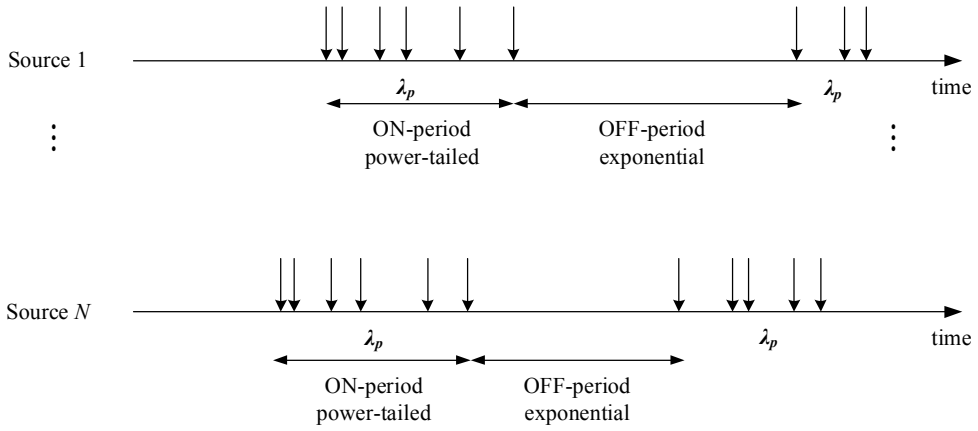


Fig. 7.1: The N-Burst process; frames from N ON/OFF sources are aggregated together [110]. The mean ON-period is Y and the mean OFF-period is Z.

Since PT distributions with infinite tails are considered to be non-realistic (there are physical limits on *e.g.*, file sizes), it would be expected that the PT behavior of the ON-period distributions is cut off at some point. Therefore, the TPT distributions are used for the ON-period distribution in the *N*-Burst model. The Cumulative Distribution Function (CDF) of the TPT distribution [110] is given by

$$F(x) = \frac{1-\theta}{1-\theta^T} \sum_{i=0}^{T-1} \theta^i \exp(\frac{-\zeta_T}{\gamma^i} x) \tag{7.19}$$

where T is the number of phases of the TPT distribution. The drop-off of PT occurs later with the larger T. The variable θ can be chosen randomly in the

range $0 < \theta < 1$. More phases are needed to obtain the same PT range as θ increases. In order to show the power-law behavior with the tail-exponent α, and to have mean Y for the TPT distribution, the other constants in Eq. (7.19) must satisfy

$$\gamma = (1/\theta)^{1/\alpha} \tag{7.20}$$

$$\zeta_T = \frac{1-\theta}{1-\theta^T}\frac{1-(\theta\gamma)^T}{1-\theta\gamma}\frac{1}{Y} \tag{7.21}$$

The Hurst parameter, H, is equal to $(3-\alpha)/2$. The entry vector and rate matrix, $< \mathbf{w}, \mathbf{B} >$, of a T-phase TPT distribution, are given by

$$\mathbf{w} = \frac{1-\theta}{1-\theta^T}[\theta^0, \cdots, \theta^{T-1}] \tag{7.22}$$

$$\mathbf{B} = \zeta_T \, \mathrm{diag}(\gamma^0, \gamma^{-1}, \cdots, \gamma^{-(T-1)}) \tag{7.23}$$

Consider the 1-Burst ON/OFF model with ME distributed ON-periods described by its entry vector $\mathbf{w}$ and rate matrix $\mathbf{B}$, and the exponential OFF-periods with mean *Z*. It has the Markov Modulated Poisson Process (MMPP) representation [110] with infinitesimal generator $\mathbf{Q}$ of the underlying Markov chain and the rate matrix $\mathbf{\Lambda}$

$$\mathbf{Q} = \begin{bmatrix} -1/Z & \mathbf{w}/Z \\ \mathbf{Be} & -\mathbf{B} \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} 0 & 0 \\ 0 & \lambda_p \mathbf{I} \end{bmatrix} \tag{7.24}$$

where $\mathbf{e}$ represents a unit column vector and $\mathbf{I}$ denotes a unit matrix. λ_p is the Poisson arrival rate during ON-period.

For the *N*-Burst model that is the aggregation of the traffic from N identical 1-Burst ON/OFF sources, the representation of the N-Burst model is N Kronecker sums (denoted by the symbol "⊕") of $\mathbf{Q}$ and Λ, respectively.

$$\begin{aligned} \mathbf{Q}' &= \mathbf{Q} \oplus \cdots \oplus \mathbf{Q} \\ \mathbf{\Lambda}' &= \mathbf{\Lambda} \oplus \cdots \oplus \mathbf{\Lambda} \end{aligned} \tag{7.25}$$

With $\mathbf{Q}'$, Λ', and the service rate μ, the transition rate matrix, $\mathbf{G}$, of the underlying Markov chain for the N-Burst/M/1/K queueing system can be derived.

$$\mathbf{G} = \begin{bmatrix} \mathbf{H}_1 + \mathbf{H}_2 & \mathbf{H}_0 & \mathbf{0} & \cdots & \\ \mathbf{H}_2 & \mathbf{H}_1 & \mathbf{H}_0 & \mathbf{0} & \cdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ & \mathbf{0} & \mathbf{H}_2 & \mathbf{H}_1 & \mathbf{H}_0 \\ & & & \mathbf{H}_2 & \mathbf{H}_0 + \mathbf{H}_1 \end{bmatrix} \tag{7.26}$$

where $\mathbf{H}_0 = \mathbf{\Lambda}'$, $\mathbf{H}_1 = \mathbf{Q}' - \mathbf{\Lambda}' - \mu\mathbf{I}$, $\mathbf{H}_2 = \mu\mathbf{I}$.

The steady-state probability distribution of the Markov chain, $\mathbf{P} = (P_{s,\eta}) = (\mathbf{P}_0, \mathbf{P}_1, \cdots, \mathbf{P}_K)$, where η is the dimension of the rate matrix $\mathbf{\Lambda}'$ and $\mathbf{P}_s = (P_{s,1}, P_{s,2}, ..., P_{s,\eta})^{\mathbf{T}}$, $0 \le s \le K$, satisfies the following equations

$$\mathbf{PG} = 0 \text{ and } \mathbf{Pe} = 1 \tag{7.27}$$

Solving these equations yields the steady-state probability vector as [39]

$$\mathbf{P} = \mathbf{u}(\mathbf{I} - \Re + \mathbf{eu})^{-1} \tag{7.28}$$

where $\Re = \mathbf{I} + \mathbf{G} / \min\{\mathbf{G}(\rho, \rho)\}$, $\mathbf{u}$ is an arbitrary row vector of $\Re$, and $\mathbf{e}$ is a unit column vector.

Having obtained $\mathbf{P}$, the steady-state probability that there are s frames in the queueing system, P_s, can be given by

$$P_s = \mathbf{P}_s \mathbf{e}, \text{ for } 0 \le s \le K \tag{7.29}$$

Let π_s represent the steady-state probability of having s frames in the queueing system as seen by an arbitrary arriving frame, π_s is given by [92]

$$\pi_s == \frac{\mathbf{P}_s \mathbf{\Lambda}' \mathbf{e}}{\sum_{s=0}^{K} \mathbf{P}_s \mathbf{\Lambda}' \mathbf{e}}, \quad \text{for } 0 \le s \le K \tag{7.30}$$

where $\mathbf{\Lambda}'$ is given in Eq. (7.25).

The loss probability, p_{bi}, equals the probability of having K frames in the queueing system as seen by an arbitrary arriving frame, π_K. With p_{bi}, throughput TH_i of AC_i can be computed by

$$TH_i = \lambda_i E[P](1 - p_{bi})(1 - p_i^{m+1}) \tag{7.31}$$

where $E[P]$ is the frame payload length and λ_i is the mean traffic rate of AC_i. p_i^{m+1} is the probability that the frame is discarded due to $(m+1)$ transmission failures.

The total delay is the time interval from the instant that a frame enters the transmission queue of the AC, to the instant that the frame leaves the queueing system. By virtue of Little's Law [68], the total delay, $E[D_i]$, is given by

$$E[D_i] = \frac{E[K_i]}{\lambda_i (1 - p_{bi})} \tag{7.32}$$

where $E[K_i] = \sum_{s=0}^{K} s P_{si}$ is the average number of frames in the queueing system

of AC_i. $\lambda_i(1-p_{bi})$ is the effective arrival rate to the transmission queue of AC_i since the arriving frames are discarded if the buffer becomes full.

Table 7.1: System Paramters for the Analysis of the EDCA Protocol under Heterogeneous Traffic

Frame payload	4000bits	PHY header	192bits
MAC header	224bits	ACK	112bits + PHY header
Channel data rate	11Mbit/s	Retry limit	7
Basic rate	1Mbit/s	Propagation delay	2μs
Slot time	20μs	Buffer size	30 frames
SIFS	10μs	AIFS	50μs

7.3 Validation and Performance Evaluation

The accuracy of the analytical model is verified by extensive NS-2 simulation experiments. A Basic Service Set (BSS) of WLANs is considered where all stations are located within the transmission range of each other. Each station consists of four ACs denoted by AC_3, AC_2, AC_1, AC_0, which generate voice, real-time video, non-real-time video, and background data traffic, respectively. The $[CW_{\min}, CW_{\max}]$ from AC_3 to AC_0 are set to [8, 128], [16, 256], [32, 512], [64, 512], respectively. The other system parameters adopted in this work are summarized in Table 7.1.

To demonstrate the effectiveness of the analytical model for the 802.11e WLANs with the multimedia wireless applications, the traffic parameters used in validation are obtained from the accurate measurement of the multimedia

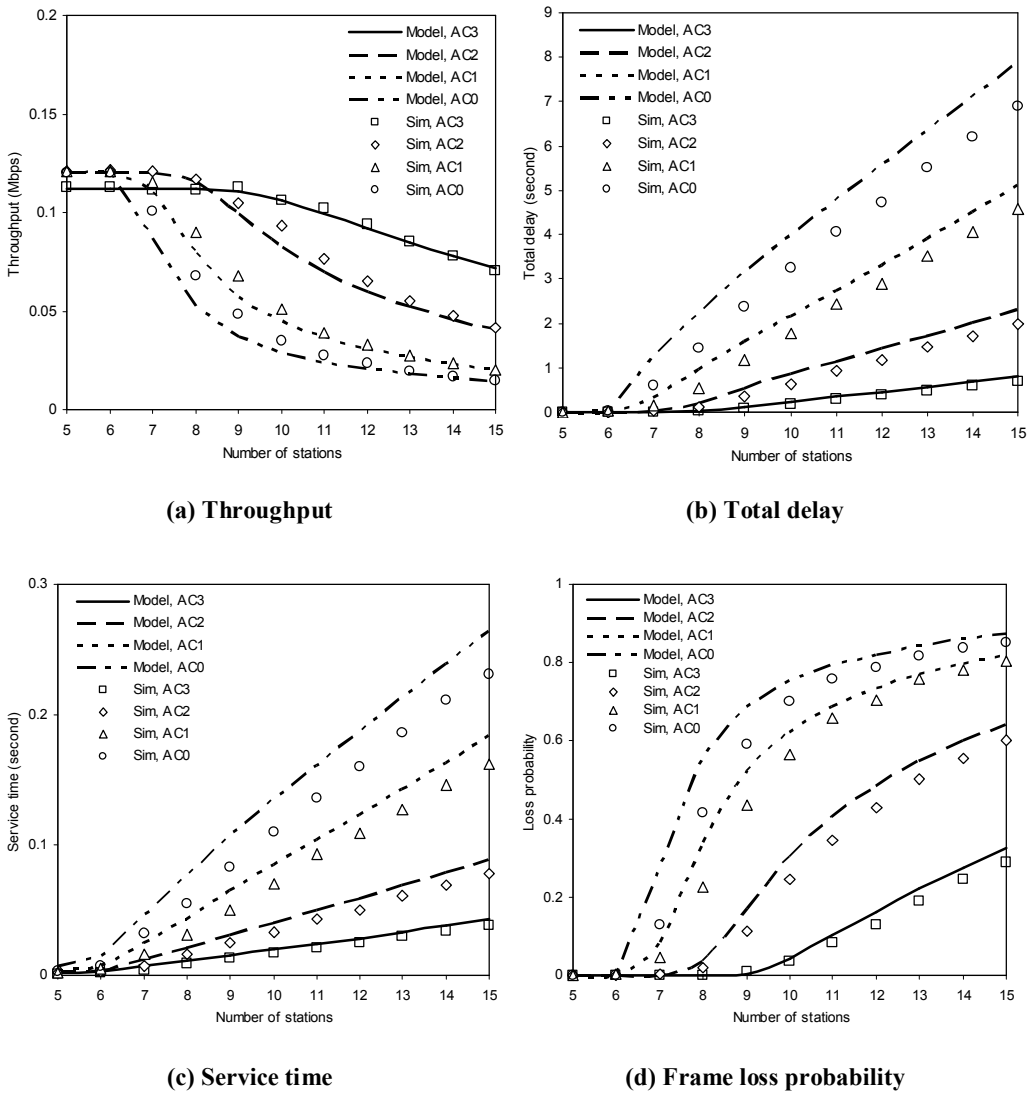
applications including the G.711 codec voice sources and the VBR encoded H.263 video streams. The details are given as follows:

Voice: The G.711 voice codec [22, 43, 77] is a popular international standard for VoIP encoding. For an ON-OFF voice source, the arrival rate of G.711 encoded frames is fixed at 64 kbps during talk spurts (in the ON state) and no frame arrives during silences (in the OFF state). The ON and OFF periods are exponentially distributed with means of 352ms and 650ms [114]. A superposition of 5 ON-OFF voice sources is considered in this study.

Video: The parameters of the video traffic are obtained from the real H.263 frame traces of the movie "Star Wars IV" [40]. H.263 is an international video coding standard which has found many applications on wireless multimedia networks. It has been shown that the VBR encoded H.263 traces exhibit a high degree of LRD self-similarity [40]. The Hurst parameter, the autocorrelation at lag1, and the mean arrival rate of the video traffic generated by the movie "Star Wars IV" are 0.88, 0.8, and 0.12 Mbps, respectively.

Fig. 7.2 depicts the throughput, total delay, frame service time, and loss probability, respectively, for each AC versus the number of stations. In order to demonstrate the QoS differentiation between various ACs, the mean arrival rates of the background data are assumed to be the same as that of the video traffic. Fig. 7.2 shows a good match between the analytical and simulation results, thus corroborating the accuracy of the proposed model. Furthermore, it can be found that

the AC with the smaller CW achieves the better performance than those with the larger one, which shows the efficiency of the CW differentiation scheme in EDCA.



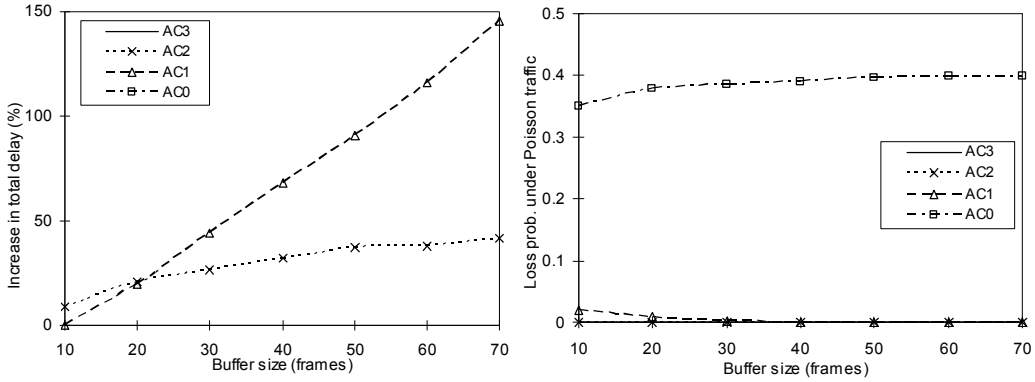
(a) Throughput **(b) Total delay**

(c) Service time **(d) Frame loss probability**

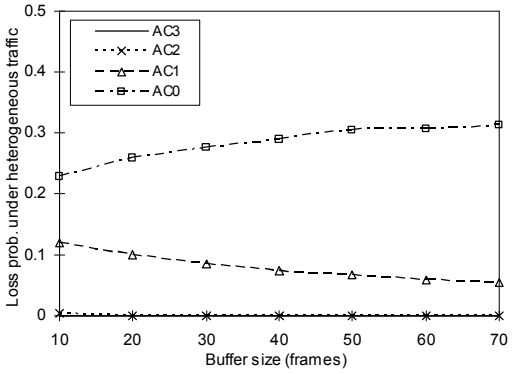
Fig. 7.2: Performance measures versus the number of stations.

To investigate the impact of heterogeneous traffic on the network performance, the throughput, total delay, frame service time, and loss probability attained from the proposed model under heterogeneous traffic are compared to those obtained from the model where the traffic of all ACs is characterized by the Poisson process. There are 7 stations with 4 ACs in each station. Fig. 7.3(a) depicts the percentage of

increase in total delay when the arrival traffic of all ACs changes from homogeneous (Poisson traffic) to heterogeneous, as a function of the buffer size. It is shown that heterogeneous traffic causes a significant increase in total delay for AC_1 and AC_2 subject to the self-similar traffic. More specifically, AC_1 experiences a dramatic increase in total delay up to 150% while AC_2 experiences an increase up to 40%.



(a) Increase in total delay (%) **(b) Loss prob. under Poisson traffic**



(c) Loss prob. under heterogeneous traffic

Fig. 7. 3: Performance comparison between the heterogeneous traffic and Poisson traffic conditions as a function of the buffer size.

Fig. 7.3(b) and 7.3(c) depict the loss probability of various ACs under Poisson traffic and heterogeneous traffic, respectively, as a function of buffer size. Fig. 7.3(b)

shows that AC_0 experiences frame losses while the loss probability at AC_1 is nearly zero when all the ACs are subject to Poisson traffic. In contrast, Fig. 7.3(c) shows that both AC_1 and AC_0 experience frame losses when the ACs are subject to heterogeneous traffic. The buffer size can affect the loss probability. As shown in Fig. 7.3(b), the loss probability of AC_0 stabilizes as the buffer size increases. However, in Fig. 7.3(c), the loss probability of AC_0 tends to increase to a certain level while that of AC_1 decreases as the buffer size increases. The above observations stress the great need to develop a comprehensive analytical model that can quantitatively and accurately capture the effects of heterogeneous traffic on the performance of EDCA.

7.4 Admission Control in IEEE 802.11e WLANs

Admission control is an important mechanism for the provisioning of the user-perceived QoS in the IEEE 802.11e WLANs. This subsection presents an efficient admission control scheme based on analytical modelling and non-cooperative game theory where the Access Point (AP) and new users are the players. The decision of admission control is made by virtue of the strategies to maximize the utilities of the players, which are determined by the QoS performance metrics in terms of the end-to-end delay and frame loss probability. These performance metrics are obtained from the analytical model incorporating the CW and TXOP differentiation schemes in EDCA.

7.4.1 Game Theoretical Approach for Admission Control

The admission control is formulated as a non-cooperative non-zero-sum $(n+1)$-player game where the AP and n new users requiring services are the players. The game is non-zero-sum [46] because the payoffs of both the AP and new user would increase if the admission of a new user does not degrade the QoS of the ongoing users. In a zero-sum game, an increase of one player's payoff implies a decrease of another player's payoff.

Before presenting the admission control scheme, the notion of utility (*i.e.*, payoff) needs to be introduced. Utility is a measure of the user's satisfaction level and can be modelled as a function of the QoS perceived by the user. The modified sigmoid function [80] is employed to characterize the user's utility. The value of the utility function is between zero and one, which can facilitate the representation of the user's satisfaction from the lowest to highest levels. Moreover, both the slope and the centre of the curve of this function can be adjusted to customize the utility of a specific class of users. Different traffic classes have diverse QoS preferences. For the user with real-time applications such as voice and video, there is a stringent constraint on the end-to-end delay. Therefore, the utility for the real-time user depends on the end-to-end delay, $E[d]$, and can be expressed as

$$U_{re} = 1 - \frac{1}{1+\exp(-\alpha_{re}(E[d]-E[d_{tol}]))} \tag{7.33}$$

where the value of α_{re} indicates the user's sensitivity to the increase in delay and $E[d_{tol}]$ is the largest delay that the real-time user could tolerate (*i.e.*, delay

constraint). With the increase of delay, α_{re} determines the rate of the decrease and $E[d_{tol}]$ determines when the utility decreases until below 0.5.

For the user with non-real-time applications, there is not a strict delay requirement and its QoS satisfaction depends on the frame loss probability, LO. Thus, the utility function for the non-real-time user can be given by

$$U_{nr} = 1 - \frac{1}{1 + \exp(-\alpha_{nr}(LO - LO_{tol}))} \tag{7.34}$$

where the value of α_{nr} indicates the user's sensitivity to the increase of frame loss probability and LO_{tol} is the largest frame loss probability that the non-real-time flow could tolerate (*i.e.*, loss probability constraint). With the increase of frame loss probability, α_{nr} determines the rate of the decrease and LO_{tol} decides when the utility decreases below 0.5.

The admission control scheme is invoked when the AP receives the requests from n new users. The AP then makes a decision which requests to be accepted or rejected. The new users also need decide whether to accept the service or deny it. The process of making the decisions is modelled as an admission control game that is solved to obtain the best strategies of the players. Let the *k*-th request belong to the user in Class C_k $(1 \le C_k \le C)$. Based on either admitting or rejecting a request, the AP has 2^n strategies, each of which can be expressed by an *n*-element vector. The *i*-th strategy for the AP is given by

$$A_i = [a_{i1}, a_{i2}, \cdots, a_{in}] \tag{7.35}$$

where $a_{ik} \in \{0,1\}$ for $1 \le i \le 2^n$ and $1 \le k \le n$. $a_{ik} = 0$ means to accept the k-th request while $a_{ik} = 1$ represents to reject it. Similarly, each new user has two strategies as well. Let B_k be the k-th new user's strategy $(1 \le k \le n)$, then $B_k = 0$ denotes to accept the service provided by the AP and $B_k = 1$ means to deny the service.

The payoff matrices of the AP and the k-th new user, $\mathbf{S}$ and $\mathbf{R}_k$, are denoted as [80]

$$\mathbf{S} = [s_{A_i B_1 B_2 \cdots B_n}]_{2^n \times 2 \times \cdots \times 2}, \qquad \mathbf{R}_k = [r^k_{A_i B_1 B_2 \cdots B_n}]_{2^n \times 2 \times \cdots \times 2} \tag{7.36}$$

where $s_{A_i B_1 B_2 \cdots B_n}$ and $r^k_{A_i B_1 B_2 \cdots B_n}$ denote the payoff of the AP and k-th new user, respectively, if the AP chooses strategy A_i and the j-th new user $(j = 1, 2, \cdots, n)$ chooses strategy B_j. The values of each element of payoff matrices $\mathbf{S}$ and $\mathbf{R}_k$ are given by

$$s_{A_i B_1 B_2 \cdots B_n} = U + \sum_{1 \le k \le n \,\&\, a_{ik} = B_k = 0} (U_k - L_k) \tag{7.37}$$

$$r^k_{A_i B_1 B_2 \cdots B_n} = (1 - a_{ik})\big((1 - B_k)U_k + B_k(1 - U_k)\big) \tag{7.38}$$

where U represents the total payoff of the ongoing users. L_k is the decrease in payoff of the ongoing users after admitting the k-th new user. U_k is the payoff of the k-th new user after it accepts the service and $(1 - U_k)$ is the payoff of the new user when it denies the service due to the undesired QoS performance.

To have the best strategies that maximize the payoffs of the AP and new users, the Nash equilibrium [46] needs to be obtained, which is a strategy profile where

none of the players can increase its own utility by unilaterally changing its strategy. Alternatively, a strategy profile $(A_{i'}, B_1', B_2', \cdots, B_n')$ is said to constitute a Nash equilibrium solution of the game if the following inequalities are satisfied for any other strategies A_i and B_j $(j = 1, 2, \cdots, n)$.

$$\begin{cases} \mathbf{S}(A_{i'}, B_1', B_2', \cdots, B_n') \geq \mathbf{S}(A_i, B_1, B_2, \cdots, B_n) \\ \mathbf{R}_k(A_{i'}, B_1', B_2', \cdots, B_n') \geq \mathbf{R}_k(A_i, B_1, B_2, \cdots, B_n), \ 1 \leq k \leq n \end{cases} \tag{7.39}$$

Before discussing the solution of Nash equilibrium, the concept of dominance in game theory [46] is firstly introduced.

Definition 7.1: In a non-cooperative *n*-player game, strategy A_τ' of player P_τ $(\tau = 1, 2, \cdots, n)$ dominates another strategy A_τ if it obtains the better payoff with strategy A_τ' than with strategy A_τ, for any strategy A_γ of any other player P_γ $(1 \leq \gamma \leq n \ \& \ \gamma \neq \tau)$.

Using the above definition, it can be noticed that there exists a dominant strategy, A_d, for the AP satisfying

$$s_{A_d B_1 B_2 \cdots B_n} = \max\{s_{A_i B_1 B_2 \cdots B_n}\} \ \text{ for } \ (1 \leq i \leq 2^n) \tag{7.40}$$

which, with Eq. (7.37), can be expressed as

Table 7.2: System Paramters for Admission Control

Frame payload	8000bits	PHY header	192bits
MAC header	224bits	ACK	112bits + PHY header
Channel data rate	11Mbit/s	Maximum backoff stages	5
Basic rate	1Mbit/s	Buffer size	40 frames
Propagation	2μs	AIFS	50μs
Slot time	20μs	SIFS	10μs

$$\sum_{1\le k\le n \,\&\, a_{dk}=B_k=0}(U_k - L_k) = \max\left\{\sum_{1\le k\le n \,\&\, a_{ik}=B_k=0}(U_k - L_k)\right\} \text{ for } (1 \le i \le 2^n) \tag{7.41}$$

Similarly, there exists a dominant strategy, B_k, for the *k*-th new user satisfying

$$r^k_{A_i B_1 B_2 \cdots B_n} = \max\{r^k_{A_i B_1 B_2 \cdots B_n}\} \quad \text{for } B_k \in \{0,1\} \tag{7.42}$$

which, with Eq. (7.38), leads to

$$\begin{aligned} B_k &= 0, \quad \text{if } U_k \ge 0.5 \\ B_k &= 1, \quad \text{if } U_k < 0.5 \end{aligned} \tag{7.43}$$

Therefore, the Nash equilibrium strategy profile can be expressed as $A_d B_1 B_2 \cdots B_n$, if A_d and B_k $(1 \le k \le n)$ satisfy Eqs. (7.41) and (7.43).

7.4.2 Numerical Results

The effectiveness and the superior performance of the proposed admission control scheme are verified through NS-2 simulation tests. Users in WLANs are classified into the real-time and non-real-time ones with various QoS constraints. The traffic arrival rate of the real-time and non-real-time users is set to 0.2 and 0.1 Mbps, respectively. The sensitivity of the utility function for the real-time and non-real-time users is set to 10 and 5, respectively, *i.e.*, $\alpha_r = 10$ and $\alpha_{nr} = 5$. All the users are located in a Basic Service Set (BSS) of 150m×150m rectangular grid. In what follows, without any specification, the system parameters are the same as those given in Table 7.2. The $CW_{\min}$ for the real-time and non-real-time user is set to 16 and 32, respectively, and the TXOP limit of the real-time and non-real-time

users is set to 2 and 1 (frame), respectively.

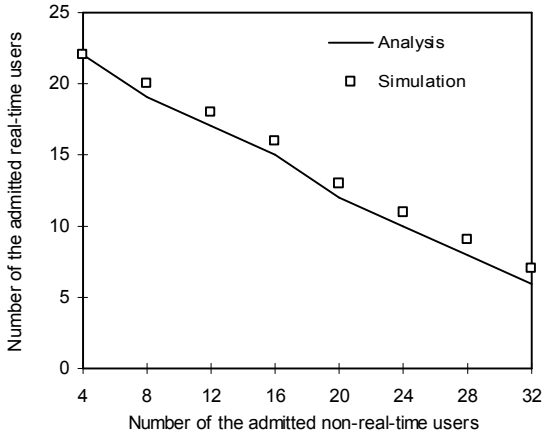


Fig. 7.4: Maximum number of the admitted real-time users versus that of the admitted non-real-time users.

Fig. 7.4 plots the admission region of the real-time and non-real-time users. The delay and frame loss probability constraints are assumed to be 30ms and 0.1, respectively. It is shown that the analytical predictions of the maximum number of admitted stations are in a good agreement with the simulation results. Fig. 7.5 depicts the performance results of the proposed admission control scheme used to regulate the maximum number of admitted users. It can be seen that the simulation results with the proposed admission control scheme in terms of end-to-end delay and loss probability clearly satisfy the QoS constraints. The above observations demonstrate the effectiveness of the proposed admission control scheme.

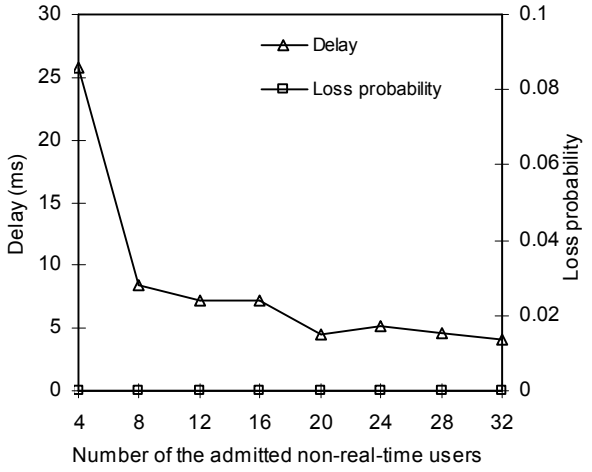


Fig. 7.5: Simulation results with the proposed admission control scheme.

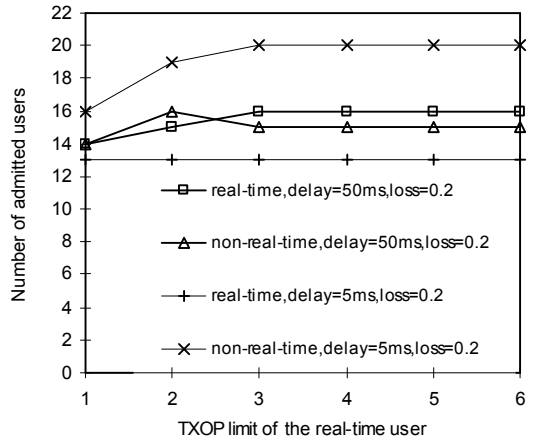


Fig. 7.6: Maximum number of admitted users versus the TXOP limit of the real-time user.

Utilizing the admission control scheme, the capacity of the 802.11e based WLAN with different QoS requirements and MAC configurations is investigated. Fig. 7.6 depicts the maximum number of admitted users in each class versus the TXOP limit of the real-time user with the different delay constraints. The TXOP limit of the non-real-time user is fixed at 1 (frame). In the case that the delay constraint is 50ms, it is observed that the maximum number of both real-time and non-real-time users increases as the TXOP limit increases from 1 to 3. More

specifically, the maximum number of the real-time users keeps increasing as the TXOP limit increases from 1 to 3, while the maximum number of the non-real-time users firstly increases as the TXOP limit increases from 1 to 2, and then decreases as the TXOP limit increases from 2 to 3. However, the further increase in the TXOP limit does not have any impact on the system capacity and thus the maximum number of the real-time and non-real-time users that can be admitted stabilizes at 16 and 15, respectively.

Next, the impact of the QoS constraints on the system capacity is investigated. When the delay constraint of the real-time user becomes more stringent (*e.g.*, $E[d_{tol}] = 5\text{ms}$), the AP admits less real-time users, while it can accept more non-real-time ones. The reason is that the incoming real-time users reject the service since their delay constraint cannot be satisfied while the incoming non-real-time users are admitted as their loss probability requirements can be fulfilled.

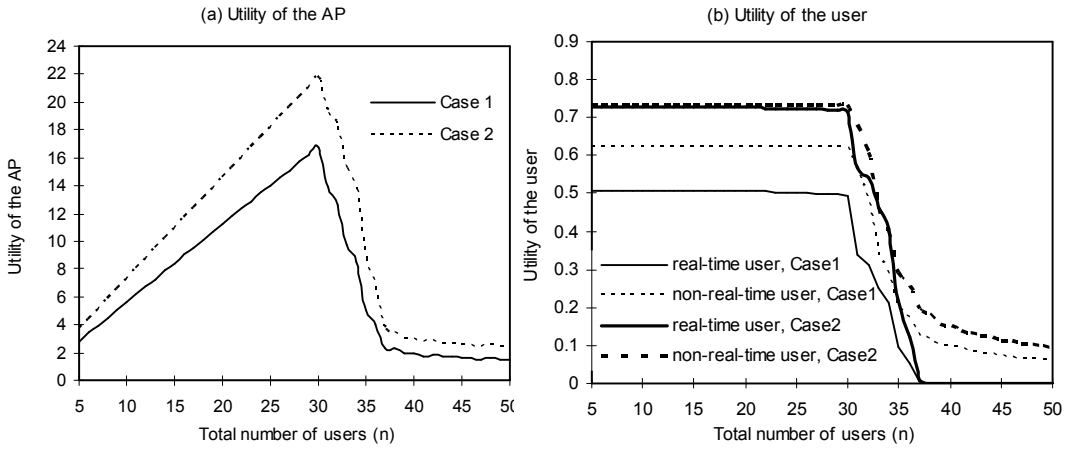


Fig. 7.7: Utility versus the total number of users. Case 1: $E[d_{tol}]$ = 5 ms, LO_{tol} = 0.1; Case 2: $E[d_{tol}]$ = 100 ms, LO_{tol} = 0.2.

Fig. 7.7 depicts the utility for the AP and new user, respectively, when the total number of users increases from 5 to 50 in two cases where Case 1 has the more stringent QoS requirements than Case 2. It shows that the point for maximizing the utility of the AP is $n = 30$ in both cases. For the new user, the utility decreases as the number of total users increases. The maximum number of users to keep the utility above 0.5 (*i.e.*, the QoS constraints are satisfied) is $n = 26$ and $n = 31$ in Case 1 for the real-time and non-real-time users respectively, and $n = 32$ in Case 2 for both classes of users. Therefore, the AP could admit 30 users while the new user will not accept the service if the number of total users is larger than 26 in Case 1, which means that the WLAN will accommodate at most 26 users that are decided by the new user in Case 1. On the other hand, in Case 2, the AP can admit 30 users and the new user would only accept the service if the number of total users is not larger than 32. Thus, the WLAN will accept a maximum number of 30 users that are decided by the AP in Case 2. The results demonstrate that the proposed admission control scheme can assure that the system operates under an optimal point where the utility of the AP is maximized with the QoS constraints of both the real-time and non-real-time users.

7.5 Summary

This chapter has proposed a new analytical model for EDCA in the presence of heterogeneous traffic captured by the SRD non-bursty Poisson, SRD bursty ON-OFF, and LRD self-similar traffic. QoS performance metrics in terms of

throughput, total delay, frame service time, and loss probability have been derived. The accuracy of the model has been validated by extensive NS-2 simulation experiments. The traffic parameters used in validation are obtained from the accurate measurement of the multimedia applications including the G.711 codec voice sources and the VBR encoded H.263 video streams. It is shown that the heterogeneous traffic has a significant impact on the performance of EDCA with different buffer sizes. The performance results have highlighted the importance of considering the heterogeneous traffic for the design and performance evaluation of EDCA with wireless multimedia applications.

An admission control scheme has also been presented for the IEEE 802.11e WLANs based on analytical modelling and game theory. The decision of admission control is made on the base of the strategies maximizing the utilities of the players, which are calculated through the QoS performance measures including the end-to-end delay and frame loss probability derived from an analytical model for the IEEE 802.11e EDCA protocol. The admission control scheme has been validated through NS-2 simulation experiments and has been applied to investigate the network capacity of WLANs under different network configurations and QoS constraints. Specifically, the impact of the TXOP limit and the delay constraint over the network capacity has been evaluated. Moreover, it is demonstrated that the proposed game-theoretical admission control scheme can maintain the system operation at an optimal point where the utility of the AP is maximized with the QoS constraints of various users.

Chapter 8

Conclusions and Future Work

With the explosive growth in emerging multimedia applications, such as VoIP, video on demand, and video conferencing, one of the main thrust of current research on wireless networks is towards the provisioning of differentiated QoS. The MAC protocol, which coordinates the data transmission of wireless stations, plays a pivotal role in wireless networks. The primary focus of this thesis is to investigate the QoS performance of MAC protocols in WLANs with multimedia applications. In the following, a summary of the work in this thesis is provided and some directions of future research are indicated.

8.1 Conclusions

The thesis has presented new analytical tools for performance analysis and enhancement of wireless MAC protocols in the presence of heterogeneous multimedia traffic. The accuracy of the proposed models has been validated through comparing the analytical results against those obtained from NS-2 simulation experiments. The proposed analytical models have been used to investigate the QoS performance attributes of the 802.11 and 802.11e MAC protocols and develop

efficient admission control schemes for multimedia WLANs. The major achievements in this research are summarized as follows:

- A new analytical model is developed for evaluating the burst transmission schemes specified in the IEEE 802.11e protocol in terms of CFB and BACK under unsaturated traffic loads and finite buffer capacity. The QoS performance metrics including throughput, end-to-end delay, and frame loss probability have been derived in the analytical model. A thorough investigation into the impact of the traffic load, TXOP limit, buffer size, channel data rate, minimum contention window, and number of stations on the QoS performance of the burst transmission schemes has been conducted.

- A novel analytical model for the TXOP scheme in WLANs consisting of unbalanced stations with different traffic loads has been proposed. The model derives the expressions of the important QoS performance metrics including throughput, end-to-end delay, frame loss probability, and energy consumption of the TXOP scheme. The developed model has been used to investigate the efficiency of the TXOP scheme for QoS differentiation and evaluate the effects of the TXOP limit on the network performance in the presence of unbalanced stations. In addition, it is shown that the desirable throughput differentiation for various stations can be achieved by setting the appropriate TXOP limits.

- A new analytical model for the TXOP scheme in unsaturated WLANs with bursty error channels has been developed. The transmission queue of each

station is modelled by a CTMC that captures the bursty characteristics of the wireless channel errors and the burst transmission mechanism of the TXOP scheme. The impacts of traffic load, TXOP limit, and the number of stations on the performance of the TXOP scheme under different channel conditions have been investigated.

- A dynamic TXOP scheme which can adjust the TXOP limits according to the current status of the transmission queue and the pre-setting threshold has been presented. An original analytical performance model for this dynamic TXOP scheme subject to self-similar traffic has been developed. The analytical performance results have demonstrated that the dynamic TXOP scheme can achieve the better QoS performance than the original scheme under self-similar traffic. The analytical model has been adopted as a cost-efficient tool to investigate the effects of several important parameters of the dynamic TXOP scheme on network performance with the aim of obtaining the optimal parameter settings of this scheme.

- The Poisson, MMPP and self-similar process have been adopted to jointly model the heterogeneous traffic of wireless multimedia applications. A versatile model has been developed to address the difficulties of queueing analysis arising from the bulk service of the TXOP scheme in the presence of heterogeneous traffic. The accuracy of the developed model has been corroborated through extensive NS-2 simulation experiments subject to the traffic parameters obtained from the accurate measurements of the real-world

multimedia voice and video sources. The analytical model has been used to investigate the effects of the TXOP scheme on the QoS of heterogeneous stations. The performance results have highlighted the importance of taking into account the heterogeneous traffic for the accurate evaluation of the TXOP scheme in wireless multimedia networks.

- A new 3D Markov chain has been proposed to model the backoff procedure of EDCA. The proposed model can handle the large MAC buffer size without heavily increasing the complexity of the solution. Moreover, this model takes into account the virtual collisions between ACs in a station and the effects of the frame retry limit in order to capture the behavior of EDCA accurately. This comprehensive analytical model incorporates the combination of the QoS differentiation schemes in terms of AIFS, CW, and TXOP simultaneously. The performance metrics including throughput, end-to-end delay, delay jitter, and frame loss probability have been derived. Performance results obtained from the developed analytical model have shown that the TXOP scheme can not only support service differentiation but also improve the network performance, whereas the AIFS and CW schemes provide QoS differentiation only. Moreover, the results have demonstrated that the MAC buffer size has considerable impact on the QoS performance of ACs in IEEE 802.11e WLANs. The performance results have highlighted the importance of using the combination of these three schemes in WLANs and demonstrate the value of this comprehensive analytical model as a cost-efficient tool for performance evaluation of EDCA.

- An analytical model has been developed to investigate the effects of heterogeneous traffic on the QoS of EDCA. The heterogeneous Poisson, bursty ON-OFF, and self-similar processes have been jointly adopted to model the heterogeneous traffic of wireless multimedia applications including background, voice, and video traffic, respectively. The traffic parameters used in the validations are obtained from the accurate measurements of the multimedia applications including the G.711 codec voice sources and the VBR encoded H.263 video streams. The results have revealed the significant effects of heterogeneous traffic on the total delay and frame losses of EDCA with different buffer sizes. Moreover, the results have demonstrated the importance of taking the traffic heterogeneity into account for the design and performance evaluation of EDCA.

- An admission control scheme has been presented for the IEEE 802.11e WLANs based on analytical modelling and a game theoretical approach. The admission control is formulated as a non-cooperative (n+1)- player game where the AP and n new users are the players. The utility for users in different classes is expressed as a function of the end-to-end delay or frame loss probability based on their traffic characteristics. The proposed admission control scheme has been used to carry out the investigation on the capacity of WLANs under different network configurations and QoS constraints. The numerical results have demonstrated that our admission control scheme can maintain the system

operation at an optimal point where the utility of the AP is maximized with the QoS constraints of various users.

8.2 Future Work

The thesis mainly investigates the QoS performance of MAC protocols in WLANs with multimedia applications. Although this thesis work has realized the main research objective, the following future work can be suggested to extend this research to accommodate emerging wireless networks and more general working scenarios.

Cognitive MAC: Sparked by recent advances in cognitive radios [38, 90, 111, 112, 141], a new communication paradigm presents a possible solution to the spectrum inefficiency problem, which allows secondary users equipped with cognitive radios to opportunistically access unoccupied bands of licensed spectrum while limiting the interference perceived by the primary users. This networking paradigm is referred to as Cognitive Radio Networks (CRNs). The cognitive MAC protocol [31, 37, 62, 115, 124, 143], which makes efficient sensing decisions to explore spectrum opportunities for secondary users and then utilize such opportunity to conduct data transmission, plays a crucial role in CRNs. There exist many open research challenges that need to be investigated for the development and design of the cognitive MAC schemes in CRNs. For instance, designing an efficient QoS-aware MAC protocol in cognitive radio networks is highly challenging due to

the dynamic nature of spectrum opportunities and the particular characteristics of heterogeneous multimedia traffic. A significant amount of work is needed to explore that how to select channel switch patterns in the multi-channel environment with heterogeneous and dynamic channel conditions and how to guarantee the QoS of secondary users in the presence of channel sensing errors and unpredictable primary user activities.

Multi-Hop Wireless Mesh Networks: The multi-hop wireless mesh networks [51, 128] present more challenges to the MAC protocol design than the single-hop WLANs. The collisions caused by hidden nodes can seriously degrade the channel utilization. The spatial locations of the contending nodes can greatly affect the channel access opportunity of each flow, resulting in starvation of some flows [50]. The MAC scheme should be scalable and its complexity and overhead do not increase dramatically with the scale of the mesh networks. Wireless mesh networks are expected to support heterogeneous multimedia traffic including voice, video, and data with various QoS requirements. A cost-efficient analytical tool that can accurately evaluate the performance of the proposed MAC scheme for wireless Mesh networks under heterogeneous multimedia traffic is crucial to the MAC design.

Cross-layer Bandwidth Allocation: There are a few open research issues related to bandwidth allocation in wireless ad hoc networks [47, 65, 116, 135]. Bandwidth allocation is needed in wireless ad hoc networks to ensure each user receives its desired QoS. Achieving this requires cross-layer cooperation, *i.e.*, the

network and MAC layers needs to cooperate. The network layer calculates the rate each flow is granted. This calculation depends on the number of active flows and the interference model of wireless networks. Based on the calculation, the MAC layer coordinates data transmissions of users so that each flow obtains its desired bandwidth allocation. A protocol is required for scheduling the transmissions at each node to guarantee the allocated rates. The widely used CSMA/CA MAC protocols cannot perform distributed scheduling as they are based on random access. Hence new MAC protocols need to be developed for bandwidth allocation in wireless ad hoc networks.

Heterogeneous Wireless Access Networks: The integration of different wireless access technologies such as third-generation (3G) cellular networks, Wireless Metropolitan Area Networks (WMANs), and WLANs is emerging as an effective approach to provide ubiquitous availability of multimedia services and applications [4, 84, 97, 138]. In Next-Generation Heterogeneous Wireless Networks (NGHWNs), vertical handoff (user roaming among different technologies) will significantly affect various aspects of network design and planning since these access technologies possess great characteristic diversity in terms of coverage, resources, and cost. Therefore, the development of new mobility models that can accurately capture vertical mobility is crucial for addressing different design challenges in NGHWNs.

Bibliography

[1] A. Abdrabou and W. Zhuang, "Service Time Approximation in IEEE 802.11 Single-Hop Ad Hoc Networks," *IEEE Transactions on Wireless Communications,* vol. 7, no. 1, pp. 305-313, 2008.

[2] A. Abdrabou and W. Zhuang, "Stochastic Delay Guarantees and Statistical Call Admission Control for IEEE 802.11 Single-Hop Ad Hoc Networks," *IEEE Transactions on Wireless Communications,* vol. 7, no. 10, pp. 3972-3981, 2008.

[3] O. Abu-Sharkh and A. Tewfik, "Toward Accurate Modeling of the IEEE 802.11e EDCA under Finite Load and Error-prone Channel," *IEEE Transactions on Wireless Communications,* vol. 7, no. 7, pp. 2560-2570, 2008.

[4] J. Akhtman and L. Hanzo, "Heterogeneous Networking: An Enabling Paradigm for Ubiquitous Wireless Communications," *Proceedings of the IEEE,* vol. 98, no. 2, pp. 135-138, 2010.

[5] Hamed M. K. Alazemi, A. Margolis, J. Choi, R. Vijaykumar, and S. Roy, "Stochastic Modelling and Analysis of 802.11 DCF with Heterogeneous Non-Saturated Nodes," *Computer Communications,* vol. 30, no. 18, pp. 3652-3661, 2007.

[6] A. T. Andersen and B. F. Nielson, "A Markovian Approach for Modeling Packet Traffic with Long-Range Dependence," *IEEE Journal on Selected Areas in Communications,* vol. 16, no. 5, pp. 719-732, 1998.

[7] C. M. Assi, A. Agarwal, and Y. Liu, "Enhanced Per-Flow Admission Control and QoS Provisioning in IEEE 802.11e Wireless LANs," *IEEE Transactions on Vehicular Technology,* vol. 57, no. 2, pp. 1077-1088, 2008.

[8] G. Van der Auwera, P. T. David, and M. Reisslein, "Traffic and Quality Characterization of Single-Layer Video Streams Encoded with

H.264/MPEG-4 Advanced Video Coding Standard and Scalable Video Coding Extension," *IEEE Transactions on Broadcasting,* vol. 54, no. 3, pp. 698-718, 2008.

[9] G. Van der Auwera, P. T. David, and M. Reisslein, "Traffic Characteristics of H.264/AVC Variable Bit Rate Video," *IEEE Communications Magazine,* vol. 46, no. 11, pp. 164-174, 2008.

[10] F. Babich and M. Comisso, "Throughput and Delay Analysis of 802.11-Based Wireless Networks Using Smart and Directional Antennas," *IEEE Transactions on Communications,* vol. 57, no. 5, pp. 1413-1423, 2009.

[11] A. V. Babu and L. Jacob, "Fairness Analysis of IEEE 802.11 Multirate Wireless LANs," *IEEE Transactions on Vehicular Technology,* vol. 56, no. 5, pp. 3073 - 3088, 2007.

[12] A. Baiocchi, "Asymptotic Behaviour of the Loss Probability of the M/G/1/K and G/M/1/K queues," *Queueing Systems,* vol. 10, no. 3, pp. 235-248, 1992.

[13] A. Banchs, P. Serrano, and A. Azcorra, "End-to-end Delay Analysis and Admission Control in 802.11 DCF WLANs," *Computer Communications,* vol. 29, no. 7, pp. 842-854, 2006.

[14] Y. Barowski, S. Biaz, and D. P. Agrawal, "Towards the Performance Analysis of IEEE 802.11 in Multi-hop Ad-Hoc Networks," *Proc. IEEE WCNC'05,* vol. 1, pp. 100-106, 2005.

[15] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic," *IEEE Transactions on Communications,* vol. 43, no. 2/3/4, pp. 1566-1579, 1995.

[16] G. Bianchi, "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.

[17] G. Bianchi and I. Tinnirello, "Remarks on IEEE 802.11 DCF Performance Evaluation," *IEEE Communications Letters,* vol. 9, no. 8, pp. 765-767, 2005.

[18] G. Bianchi, I. Tinnirello, and L. Scalia, "Understanding 802.11e Contention-Based Prioritization Mechanisms and Their Coexistence with Legacy 802.11 Stations," *IEEE Network,* vol. 19, no. 4, pp. 28-34, 2005.

[19] G. Bolch, S. Greiner, H. Meer, and K.S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*: John Wiley & Sons, 1998.

[20] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho, "Seven Years and One Day: Sketching the Evolution of Internet Traffic," *Proc. IEEE INFOCOM'09*, pp. 711-719, 2009.

[21] L. Breuer and D. Baum, *An Introduction to Queueing Theory and Matrix-Analytic Methods*: Springer, 2005.

[22] L. Cai, X. Shen, J. W. Mark, L. Cai, and Y. Xiao, "Voice Capacity Analysis of WLAN With Unbalanced Traffic," *IEEE Transactions on Vehicular Technology,* vol. 55, no. 3, pp. 752-761, 2006.

[23] F. Cali, M. Conti, and E. Gregori, "Dynamic Tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Transactions on Networking,* vol. 18, no. 6, pp. 785-799, 2000.

[24] G. Cantieni, Q. Ni, C. Barakat, and T. Turletti, "Performance Analysis under Finite Load and Improvements for Multirate 802.11," *Computer Communications,* vol. 28, no. 10, pp. 1095-1109, 2005.

[25] M. Carvalho and J. J Garcia-Luna-Aceves, "Delay Analysis of IEEE 802.11 in Single-hop Networks," *Proc. IEEE ICNP'03*, pp. 146-155, 2003.

[26] P. Chatzimisios, A. C. Boucouvalas, and V. Vitsas, "IEEE 802.11 Packet Delay: A Finite Retry Limit Analysis," *Proc. IEEE GLOBECOM'03*, pp. 950-954, 2003.

[27] C. W. Chen, C. C. Weng, and C. J. Ku, "Design of a Low Power and Low Latency MAC Protocol with Node Grouping and Transmission Pipelining in Wireless Sensor Networks," *Computer Communications,* vol. 31, no. 15, pp. 3725-3738, 2008.

[28] X. Chen, H. Zhai, X. Tian, and Y. Fang, "Supporting QoS in IEEE 802.11e wireless LANs," *IEEE Transactions on Wireless Communications,* vol. 5, no. 8, pp. 2217-2227, 2006.

[29] Y. Cheng, X. Ling, W. Song, W. Zhuang, and X. Shen, "A Cross-Layer Approach for WLAN Voice Capacity Planning," *IEEE Journal on Selected Areas in Communications,* vol. 25, no. 4, pp. 678-688, 2007.

[30] J. Choi, J. Yoo, and C. K. Kim, "A Distributed Fair Scheduling Scheme With a New Analysis Model in IEEE 802.11 Wireless LANs," *IEEE Transactions on Vehicular Technology,* vol. 57, no. 5, pp. 3083-3093, 2008.

[31] C. Cordeiro and K. Challapali, "C-MAC: A Cognitive MAC Protocol for Multi-Channel Wireless Networks," *Proc. IEEE DYSPAN'07*, pp. 147-157, 2007.

[32] M. E. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *IEEE/ACM Transactions on Networking,* vol. 5, no. 6, pp. 835-846, 1997.

[33] F. Daneshgaran, M. Laddomada, F. Mesiti, and M. Mondin, "Modelling and Analysis of the Distributed Coordination Function of IEEE 802.11 with Multirate Capability " *Proc. IEEE WCNC'08*, pp. 1344 -1349, 2008.

[34] F. Daneshgaran, M. Laddomada, F. Mesiti, and M. Mondin, "Unsaturated Throughput Analysis of IEEE 802.11 in Presence of Non Ideal Transmission Channel and Capture Effects," *IEEE Transactions on Wireless Communications,* vol. 7, no. 4, pp. 1276-1286, 2008.

[35] K. Duffy and A. J. Ganesh, "Modeling the Impact of Buffering on 802.11," *IEEE Communications Letters,* vol. 11, no. 2, pp. 219-221, 2007.

[36] P. E. Engelstad and O. N. Osterbo, "Analysis of the Total Delay of IEEE 802.11e EDCA and 802.11 DCF," *Proc. IEEE ICC'06,* vol. 2, pp. 552-559, 2006.

[37] M. Felegyhazi, M. Cagalj, and J. P. Hubaux, "Efficient MAC in Cognitive Radio Systems: A Game-Theoretic Approach," *IEEE Transactions on Wireless Communications,* vol. 8, no. 4, pp. 1984 - 1995, 2009.

[38] S. Feng, Z. Liang, and D. Zhao, "Providing Telemedicine Services in an Infrastructure-Based Cognitive Radio Network," *IEEE Wireless Communications,* vol. 17, no. 1, pp. 96-103, 2010.

[39] W. Fischer and K. Meier-Hellstern, "The Markov-Modulated Poisson Process (MMPP) Cookbook," *Performance Evaluation,* vol. 18, no. 2, pp. 149-171, 1993.

[40] F. Fitzek and M. Reisslein, "MPEG-4 and H.263 Traces for Network Performance Evaluation," *IEEE Network,* vol. 15, no. 6, pp. 40-54, 2001.

[41] C. H. Foh, M. Zukerman, and J. W. Tantra, "A Markovian Framework for Performance Evaluation of IEEE 802.11," *IEEE Transactions on Wireless Communications,* vol. 6, no. 4, pp. 1265-1276, 2007.

[42] J. R. Gallardo, S. C. Cruz, D. Makrakis, and A. Shami, "Analysis of the EDCA Access Mechanism for An IEEE 802.11e-Compatible Wireless LAN," *Proc. IEEE ISCC'08*, pp. 891-898, 2008.

[43] D. Gao, J. Cai, C. H. Foh, C. T. Lau, and K. N. Ngan, "Improving WLAN VoIP Capacity Through Service Differentiation," *IEEE Transactions on Vehicular Technology,* vol. 57, no. 1, pp. 465-474, 2008.

[44] R. G. Garroppo, S. Giordano, S. Lucetti, and L. Tavanti, "A Model-based Admission Control for IEEE 802.11e Networks," *Proc. IEEE ICC'07*, pp. 398-402, 2007.

[45] K. Ghaboosi, M. Latva-aho, Y. Xiao, and B. H. Khalaj, "Modeling Nonsaturated Contention-Based IEEE 802.11 Multihop Ad Hoc Networks " *IEEE Transactions on Vehicular Technology,* vol. 58, no. 7, pp. 3518 - 3532, 2009.

[46] R. Gibbons, *A Primer in Game Theory*: Prentice Hall, 1992.

[47] P. Goudarzi, F. Ayatollahi, and M. R. Nezami Ranjbar, "Enabling Differentiated QoS Based on Cross-Layer Optimization in Wireless Ad Hoc Networks," *Proc. IEEE ICC'09*, 2009.

[48] M. Grossglauser and J. C. Bolot, "On the Relevance of Long-Range Dependence in Network Traffic," *IEEE/ACM Transactions on Networking,* vol. 7, no. 5, pp. 629-640, 1999.

[49] N. Gupta and P. R. Kumar, "A Performance Analysis of the IEEE 802.11 Wireless LAN Medium Access Control," *Communications in Information and Systems,* vol. 3, no. 4, pp. 279-304, 2003.

[50] O. Gurewitz, V. Mancuso, J. Shi, and E. W. Knightly, "Measurement and Modeling of the Origins of Starvation of Congestion-Controlled Flows in Wireless Mesh Networks," *IEEE/ACM Transactions on Networking,* vol. 17, no. 6, pp. 1832-1845, 2009.

[51] G. R. Hiertz, D. Denteneer, S. Max, R. Taori, J. Cardona, L. Berlemann, and B. Walke, "IEEE 802.11s: The WLAN Mesh Standard," *IEEE Wireless Communications,* vol. 17, no. 1, pp. 104 -111, 2010.

[52] J. Hu, G. Min, and M. E. Woodward, "Analysis and Comparison of Burst Transmission Schemes in Unsaturated 802.11e WLANs," *Proc. IEEE GLOBECOM'07*, pp. 5133-5137, 2007.

[53] J. Hu, G. Min, M. E. Woodward, and W. Jia, "A Comprehensive Analytical Model for IEEE 802.11e QoS Differentiation Schemes under Unsaturated Traffic Loads," *Proc. IEEE ICC'08*, pp. 241-245, 2008.

[54] C. Hua and R. Zheng, "Starvation Modeling and Identification in Dense 802.11 Wireless Community Networks," *Proc. IEEE INFOCOM'08*, pp. 1022 -1030, 2008.

[55] C. L. Huang and W. Liao, "Throughput and Delay Performance of IEEE 802.11e Enhanced Distributed Channel Access (EDCA) under Saturation Condition," *IEEE Transactions on Wireless Communications,* vol. 6, no. 1, pp. 136-145, 2007.

[56] J. Hui and M. Devetsikiotis, “A Unified Model for the Performance Analysis of IEEE 802.11e EDCA,” *IEEE Transactions on Communications,* vol. 53, no. 9, pp. 1498-1510, 2005.

[57] H. Y. Hwang, S. J. Kim, D. K. Sung, and N. O. Song, “Performance Analysis of IEEE 802.11e EDCA With a Virtual Collision Handler,” *IEEE Transactions on Vehicular Technology,* vol. 57, no. 2, pp. 1293-1297, 2008.

[58] IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," IEEE Standard 802.11, 1999.

[59] IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Quality of Service (QoS) Enhancements," IEEE Standard 802.11e, 2005.

[60] M. U. Ilyas and H. Radha, “Long Range Dependence of IEEE 802.15.4 Wireless Channels,” *Proc. IEEE ICC'08*, pp. 4261-4265, 2008.

[61] I. Inan, F. Keceli, and E. Ayanoglu, “Analysis of the 802.11e Enhanced Distributed Channel Access Function,” *IEEE Transactions on Communications,* vol. 57, no. 6, pp. 1753-1764, 2009.

[62] J. Jia, Q. Zhang, and X. Shen, “HC-MAC: A Hardware-Constrained Cognitive MAC for Efficient Spectrum Management ” *IEEE Journal on Selected Areas in Communications,* vol. 26, no. 1, pp. 106 -117, 2008.

[63] X. Jin and G. Min, “Modelling and Analysis of Priority Queueing Systems with Multi-Class Self-Similar Network Traffic: A Novel and Efficient Queue-Decomposition Approach,” *IEEE Transactions on Communications,* vol. 57, no. 5, pp. 1444-1452, 2009.

[64] E-S. Jung and N. H. Vaidya, “An Energy Efficient MAC Protocol for Wireless LANs,” *Proc. IEEE INFOCOM’02,* vol. 3, pp. 1756-1764, 2002.

[65] Y. F. Kao and J. H. Huang, “Price-Based Resource Allocation for Wireless Ad Hoc Networks with Multi-Rate Capability and Energy Constraints,” *Computer Communications,* vol. 31, no. 15, pp. 3613-3624, 2008.

[66] E. Karamad and F. Ashtiani, "Performance Analysis of IEEE 802.11 DCF and 802.11e EDCA Based on Queueing Networks," *IET Communications,* vol. 3, no. 5, pp. 871 - 881, 2009.

[67] S. Kim, R. Huang, and Y. Fang, "Deterministic Priority Channel Access Scheme for QoS Support in IEEE 802.11e Wireless LANs," *IEEE Transactions on Vehicular Technology,* vol. 58, no. 2, pp. 855 - 864, 2009.

[68] L. Kleinrock, *Queueing Systems: Theory*: John Wiley & Sons, 1975.

[69] L. Kleinrock and F.A. Tobagi, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple Access and their Throughput-Delay Characteristics," *IEEE Transactions on Communications,* vol. 23, no. 13, pp. 1417-1433, 1975.

[70] Z. Kong, D. Tsang, B. Bensaou, and D. Gao, "Performance Analysis of IEEE 802.11e Contention-based Channel Access," *IEEE Journal on Selected Areas in Communications,* vol. 22, no. 10, pp. 2095-2106, 2004.

[71] K. Kosek, "Problems with Providing QoS in EDCA Ad-Hoc Networks with Hidden and Exposed Nodes " *Proc. IEEE INFOCOM Workshops'09*, 2009.

[72] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New Insights From a Fixed-point Analysis of Single Cell IEEE 802.11 WLANs," *IEEE/ACM Transactions on Networking,* vol. 15, no. 3, pp. 588 - 601, 2007.

[73] W. K. Kuo, "Energy Efficiency Modeling for IEEE 802.11 DCF System Without Retry Limits," *Computer Communications,* vol. 30, no. 4, pp. 856-862, 2007.

[74] Y. L. Kuo, H. K. Wu, and E. G. Chen, "Noncooperative admission control for differentiated services in IEEE 802.11 WLANs," *Proc. IEEE GLOBECOM'04,* vol. 5, pp. 2981-2986, 2004.

[75] J. Y. Lee, H. S. Lee, and J. S. Ma, "Model-based QoS Parameter Control for IEEE 802.11e EDCA," *IEEE Transactions on Communications,* vol. 57, no. 7, pp. 1914-1918, 2009.

[76] W. E. Leland, M. S. Taqqu, W. Willinger, and D.V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Transactions on Networking,* vol. 2, no. 1, pp. 1-15, 1994.

[77] F. H. Li, Y. Xiao, and J. Zhang, "Variable Bit Rate VoIP in IEEE 802.11e Wireless LANs," *IEEE Wireless Communications,* vol. 15, no. 1, pp. 56-62, 2008.

[78] T. Li, Q. Ni, and Y. Xiao, "Investigation of the Block ACK Scheme in Wireless Ad-Hoc Networks," *Wireless Communications and Mobile Computing,* vol. 6, no. 6, pp. 877-888, 2006.

[79] Q. Liang, "Ad Hoc Wireless Network Traffic - Self-similarity and Forecasting," *IEEE Communications Letters,* vol. 6, no. 7, pp. 297-299, 2002.

[80] H. Lin, M. Chatterjee, S. K. Das, and K. Basu, "ARC: an integrated admission and rate control framework for competitive wireless CDMA data networks using noncooperative games," *IEEE Transactions on Mobile Computing,* vol. 4, no. 3, pp. 243-258, 2005.

[81] L. Lin, H. Fu, and W. Jia, "An Efficient Admission Control for IEEE 802.11 Networks Based on Throughput Analysis of (Un)saturated Channel," *Proc. IEEE GLOBECOM'05,* vol. 5, pp. 3017-3021, 2005.

[82] C. Liu, Y. Shu, W. Yang, and O. W. W. Yang, "Throughput Modeling and Analysis of IEEE 802.11 DCF with Selfish Node " *Proc. IEEE GLOBECOM'08*, 2008.

[83] J. Liu and Z. Niu, "Delay Analysis of IEEE 802.11e EDCA Under Unsaturated Conditions," *Proc. IEEE WCNC'07*, pp. 430-434, 2007.

[84] M. Liu, Z. Li, X. Guo, and E. Dutkiewicz, "Performance Analysis and Optimization of Handoff Algorithms in Heterogeneous Wireless Networks " *IEEE Transactions on Mobile Computing,* vol. 7, no. 7, pp. 846-857, 2008.

[85] K. Lu, D. Wu, Y. Qian, Y. Fang, and R. C. Qiu, "Performance of an Aggregation-based MAC Protocol for High Data Rate Ultra-Wideband Ad

Hoc Networks," *IEEE Transactions on Vehicular Technology,* vol. 56, no. 1, pp. 312-321, 2007.

[86] X. Ma and X. Chen, "Performance Analysis of IEEE 802.11 Broadcast Scheme in Ad Hoc Wireless LANs " *IEEE Transactions on Vehicular Technology,* vol. 57, no. 6, pp. 3757 - 3768, 2008.

[87] D. Malone, K. Duffy, and D. J. Leith, "Modeling the 802.11 Distributed Coordination Function in Non-saturated Heterogeneous Conditions," *IEEE/ACM Transactions on Networking,* vol. 15, no. 1, pp. 159-172, 2007.

[88] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor, "IEEE 802.11e Wireless LAN for Quality of Service," *Proc. European Wireless Conf.,* vol. 1, pp. 31-39, 2002.

[89] D. Manjunath and B. Sikdar, "Input Queued Switches for Variable Length Packets: Analysis for Poisson and Self-Similar Traffic," *Computer Communications,* vol. 25, no. 6, pp. 590-610, 2002.

[90] M. J. Marcus, "Europe Contemplates Cognitive Radio Policies," *IEEE Wireless Communications,* vol. 17, no. 1, pp. 7-7, 2010.

[91] K. Medepalli and F. A. Tobagi, "Towards Performance Modelling of IEEE 802.11 based Wireless Networks: A Unified Framework and Its Applications," *Proc. IEEE INFOCOM'06*, 2006.

[92] K. S. Meier-Hellstern, "The Analysis of A Queue Arising in Overflow Models," *IEEE Transactions on Communications,* vol. 37, no. 4, pp. 367-372, 1989.

[93] G. Min, J. Hu, and M. E. Woodward, "An Analytical Model of the TXOP Scheme with Heterogeneous Classes of Stations," *Proc. IEEE GLOBECOM'08*, 2008.

[94] G. Min and M. Ould-Khaoua, "A Performance Model for Wormhole-Switched Interconnection Networks under Self-Similar Traffic," *IEEE Transactions on Computers,* vol. 53, no. 5, pp. 601-613, 2004.

[95] D. Miorandi, A. A. Kherani, and E. Altman, "A Queueing Model for HTTP Traffic over IEEE 802.11 WLANs," *Computer Networks,* vol. 50, no. 1, pp. 63-79 2006.

[96] A. Nafaa and A. Ksentini, "On Sustained QoS Guarantees in Operated IEEE 802.11 Wireless LANs," *IEEE Transactions on Parallel and Distributed Systems,* vol. 19, no. 8, pp. 1020-1033, 2008.

[97] D. Niyato and E. Hossain, "A Noncooperative Game-Theoretic Framework for Radio Resource Management in 4G Heterogeneous Wireless Access Networks " *IEEE Transactions on Mobile Computing,* vol. 7, no. 3, pp. 332-345, 2008.

[98] Dusit Niyato and Ekram Hossain, "QoS-aware bandwidth allocation and admission control in IEEE 802.16 broadband wireless access networks: A non-cooperative game theoretic approach," *Computer Networks,* vol. 51, no. 11, pp. 3305-3321, 2007.

[99] NS-2. "network simulator," http://www.isi.edu/nanam/ns/.

[100] M. Ozdemir and A. B. McDonald, "On the Performance of Ad Hoc Wireless LANs: A Practical Queuing Theoretic Model," *Performance Evaluation,* vol. 63, no. 11, pp. 1127-1156, 2006.

[101] P. Li, X. Geng, and Y. Fang, "An Adaptive Power Controlled MAC Protocol for Wireless Ad Hoc Networks," *IEEE Transactions on Wireless Communications,* vol. 8, no. 1, pp. 226-233, 2009.

[102] S. Pack, X. Shen, J. W. Mark, and L. Cai, "A Two-Phase Loss Differentiation Algorithm for Improving TFRC Performance in IEEE 802.11 WLANs," *IEEE Transactions on Wireless Communications,* vol. 6, no. 11, pp. 4164-4175, 2007.

[103] L. Pan, H. Wu, and X. Cao, "Design and Analysis of A Distributed and Fair Access (DFA) MAC Protocol for Multihop Wireless Networks," *IEEE Transactions on Wireless Communications,* vol. 8, no. 5, pp. 2434-2442, 2009.

[104] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*: John Wiley & Sons, 2000.

[105] F. Peng, H. M. Alnuweiri, and V. C. M. Leung, "Analysis of Burst Transmission in IEEE 802.11e Wireless LANs," *Proc. IEEE ICC'06,* vol. 2, pp. 535-539, 2006.

[106] P. P. Pham, S. Perreau, and A. Jayasuriya, "New Cross-Layer Design Approach to Ad Hoc Networks under Rayleigh Fading " *IEEE Journal on Selected Areas in Communications,* vol. 23, no. 1, pp. 28-39, 2005.

[107] D. Pong and T. Moors, "Call Admission Control for IEEE 802.11 Contention Access Mechanism," *Proc. IEEE GLOBECOM'03,* vol. 1, pp. 174-178, 2003.

[108] V. Ramaiyan, A. Kumar, and E. Altman, "Fixed Point Analysis of Single Cell IEEE 802.11e WLANs: Uniqueness and Multistability," *IEEE/ACM Transactions on Networking,* vol. 16, no. 5, pp. 1080-1093, 2008.

[109] J. W. Robinson and T. S. Randhawa, "Saturation throughput analysis of IEEE 802.11e Enhanced Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications,* vol. 22, no. 5, pp. 917-928, 2004.

[110] H. P. Schwefel and L. Lipsky, "Performance Analysis of Intermediate Systems Serving Aggregated ON/OFF Traffic with Long-Range Dependent properties," Dissertation, Technische Universitat Munchen, 2000.

[111] G. Scutari and D. P. Palomar, "MIMO Cognitive Radio: A Game Theoretical Approach," *IEEE Transactions on Signal Processing,* vol. 58, no. 2, pp. 761-780, 2010.

[112] M. Sherman, A. N. Mody, R. Martinez, C. Rodriguez, and R. Reddy, "IEEE Standards Supporting Cognitive Radio and Networks, Dynamic Spectrum Access, and Coexistence," *IEEE Communications Magazine,* vol. 46, no. 7, pp. 72 - 79 2008.

[113] S. Shin and H. Schulzrinne, "Call Admission Control in IEEE 802.11 WLANs using QP-CAT," *Proc. IEEE INFOCOM'08*, pp. 726-734, 2008.

[114] J. W. So, "Performance Analysis of VoIP Services in the IEEE 802.16e OFDMA System With Inband Signaling," *IEEE Transactions on Vehicular Technology,* vol. 57, no. 3, pp. 1876-1885, 2008.

[115] H. Su and X. Zhang, "Cross-Layer Based Opportunistic MAC protocols for QoS Provisionings Over Cognitive Radio Wireless Networks," *IEEE Journal on Selected Areas in Communications,* vol. 26, no. 1, pp. 118-129, 2008.

[116] X. Su, S. Chan, and J. H. Manton, "Bandwidth Allocation in Wireless Ad Hoc Networks: Challenges and Prospects," *IEEE Communications Magazine,* vol. 48, no. 1, pp. 80-85, 2010.

[117] H. Takagi, *Queueing Analysis: Finite Systems*: vol. 2, North-Holland, 1993.

[118] J. W. Tantra, C. H. Foh, and A. B. Mnaouer, "Throughput and Delay Analysis of the IEEE 802.11e EDCA Saturation," *Proc. IEEE ICC'05,* vol. 5, pp. 3450-3454, 2005.

[119] J. W. Tantra, C. H. Foh, I. Tinnirello, and G. Bianchi, "Analysis of the IEEE 802.11e EDCA Under Statistical Traffic," *Proc. IEEE ICC'06,* vol. 2, pp. 546-551, 2006.

[120] Z. Tao and S. Panwar, "Throughput and Delay Analysis for the IEEE 802.11e Enhanced Distributed Channel Access," *IEEE Transactions on Communications,* vol. 54, no. 4, pp. 596-603, 2006.

[121] Y. Tay and K. Chua, "A Capacity Analysis for the IEEE 802.11 MAC Protocol," *Wireless Networks,* vol. 7, no. 2, pp. 159-171, 2001.

[122] O. Tickoo and B. Sikdar, "Modeling Queueing and Channel Access Delay in Unsaturated IEEE 802.11 Random Access MAC Based Wireless Networks," *IEEE/ACM Transactions on Networking,* vol. 16, no. 4, pp. 878-891, 2008.

[123] O. Tickoo and B. Sikdar, "On the impact of IEEE 802.11 MAC on traffic characteristics," *IEEE Journal on Selected Areas in Communications,* vol. 21, no. 2, pp. 189-203, 2003.

[124] M. Timmers, S. Pollin, A. Dejonghe, L. Van der Perre, and F. Catthoor, "A Distributed Multichannel MAC Protocol for Multihop Cognitive Radio

Networks," *IEEE Transactions on Vehicular Technology,* vol. 59, no. 1, pp. 446 - 459, 2010.

[125] I. Tinnirello and S. Choi, "Efficiency Analysis of Burst Transmission with Block ACK in Contention-based 802.11e WLANs," *Proc. IEEE ICC'05,* vol. 5, pp. 3455-3460, 2005.

[126] I. Tinnirello and S. Choi, "Temporal Fairness Provisioning in Multi-Rate Contention-Based 802.11e WLANs," *Proc. IEEE WoWMoM'05*, pp. 220-230, 2005.

[127] F. A. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part II - the Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution," *IEEE Transactions on Communications,* vol. 23, no. 12, pp. 1417-1433, 1975.

[128] P. Wang and W. Zhuang, "A Collision-Free MAC Scheme for Multimedia Wireless Mesh Backbone," *IEEE Transactions on Wireless Communications,* vol. 8, no. 7, pp. 3577-3589, 2009.

[129] Sven Wietholter, M. Emmelmann, Christian Hoene, and Adam Wolisz, *TKN EDCA Model for ns-2,* Technical Report TKN-06-003, Technical University of Berlin, 2006.

[130] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of Reliable Transport Protocol over IEEE 802.11 Wireless LAN: Analysis and Enhancement," *Proc. IEEE INFOCOM'02,* vol. 2, pp. 599-607, 2002.

[131] Y. Xiao, "Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs," *IEEE Transactions on Wireless Communications,* vol. 4, no. 4, pp. 1506-1515, 2005.

[132] Y. Xiao and Y. Pan, "Differentiation, QoS guarantee, and Optimization for Real-Time Traffic over One-Hop Ad Hoc Networks," *IEEE Transactions on Parallel and Distributed Systems,* vol. 16, no. 6, pp. 538-549, 2005.

[133] C. Xu, K. Liu, G. Liu, and J. He, "Accurate Queuing Analysis of IEEE 802.11 MAC Layer " *Proc. IEEE GLOBECOM'08*, 2008.

[134] D. Xu, T. Sakurai, and H. L. Vu, "An Access Delay Model for IEEE 802.11e EDCA," *IEEE Transactions on Mobile Computing,* vol. 8, no. 2, pp. 261-275, 2009.

[135] Y. Xue, B. Li, and K. Nahrstedt, "Optimal Resource Allocation in Wireless Ad Hoc Networks: A Price-Based Approach," *IEEE Transactions on Mobile Computing,* vol. 5, no. 4, pp. 347-364, 2006.

[136] P. Yang, "A Unified Algorithm for Computing the Stationary Queue Length Distributions in M(k)/G/1/N and GI/M(k)/1/N Queues," *Queueing Systems,* vol. 17, no. 3-4, pp. 383-401, 1994.

[137] J. Yin, X. Wang, and D. P. Agrawal, "Impact of Bursty Error Rates on the Performance of Wireless Local Area Network (WLAN)," *Ad Hoc Networks,* vol. 4, no. 5, pp. 651-668, 2006.

[138] A. H. Zahran and B. Liang, "A Generic Framework for Mobility Modeling and Performance Analysis in Next-Generation Heterogeneous Wireless Networks," *IEEE Communications Magazine,* vol. 45, no. 9, pp. 92-99, 2007.

[139] A. Zenella and F. Pellegrini, "Statistical Characterization of the Service Time in Saturated IEEE 802.11 Networks," *IEEE Communications Letters,* vol. 9, no. 3, pp. 225-227, 2005.

[140] H. Zhai, Y. Kwon, and Y. Fang, "Performance Analysis of IEEE 802.11 MAC Protocols in Wireless LANs," *Wireless Communications and Mobile Computing,* vol. 4, no. 8, pp. 917-931, 2004.

[141] R. Zhang and Y. C. Liang, "Investigation on Multiuser Diversity in Spectrum Sharing based Cognitive Radio Networks," *IEEE Communications Letters,* vol. 14, no. 2, pp. 133-135, 2010.

[142] L. Zhao, L. Guo, Z. Jiang, and H. Zhang, "Game-Theoretic Medium Access Control Protocol for Wireless Sensor Networks," *IET Communications,* vol. 3, no. 8, pp. 1274-1283, 2009.

[143] Q. Zhao, L.Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework " *IEEE Journal on Selected Areas in Communications,* vol. 25, no. 3, pp. 589 - 600, 2007.

[144] Q. Zhao, D. H. K. Tsang, and T. Sakurai, "A Simple and Approximate Model for Nonsaturated IEEE 802.11 DCF," *IEEE Transactions on Mobile Computing,* vol. 8, no. 11, pp. 1539-1553, 2009.

[145] Y. Zheng, K. Lu, D. Wu, and Y. Fang, "Performance Analysis of IEEE 802.11 DCF in imperfect Channels," *IEEE Transactions on Vehicular Technology,* vol. 55, no. 5, pp. 1648-1656, 2006.

[146] H. Zhu and I. Chlamtac, "Performance Analysis for IEEE 802.11e EDCF Service Differentiation," *IEEE Transactions on Communications,* vol. 4, no. 4, pp. 1779-1788, 2005.

[147] E. Ziouva and T. Antonakopoulos, "CSMA/CA Performance under High Traffic Conditions: Throughput and Delay Analysis," *Computer Communications,* vol. 25, no. 3, pp. 313-321, 2002.