Markov Modulated CSMA Protocols with Backoff Scheduling Algorithms

WONG, Pui King

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in
Information Engineering

The Chinese University of Hong Kong
August 2011

Abstract of thesis entitled:
Markov Modulated CSMA Protocols with Backoff Scheduling Algorithms
Submitted by Pui King WONG
For the degree of Doctor of Philosophy
at The Chinese University of Hong Kong in August 2011
Advisor: Prof. Tony T. LEE

Medium Access Control (MAC) protocols have been continuously updated to keep up with the emerging new services and QoS requirements. Despite of the rapid changes of MAC protocols, a comprehensive performance analysis of any MAC protocol remains an open issue for over several decades.

Most of existing analysis of MAC protocols focused on the network throughput and packet access delay under the assumption that the network is saturated which is not realistic. We know very little about the stability of MAC protocol under the normal network operation for lack of a systematic model that can be adaptively applied to various MAC protocols with different service requirements and backoff scheduling algorithms.

In the light of the concern, we propose a queueing model of the general CSMA protocol with probability-based backoff scheduling algorithm. The input buffer of each node is modeled as a Geo/G/1 queue, in which the service time distribution of each individual head-of-line (HOL) packet can be described by a Markov chain. By means of this queueing model, we can obtain the characteristic equation of throughput, the packet queueing delay as well as the stable conditions with admissible input traffic. We also specify stable throughput and bounded delay regions with respect to the retransmission factor and input rate.

Furthermore, we show that geometric retransmission algorithm is intrinsically unstable for large population sizes. On the other hand, exponential backoff algorithm is more robust and scalable. Even for infinity population sizes, the stable throughput and bounded delay region still exists under certain conditions.

Other than the probability-based backoff algorithm, this thesis also includes the study of window-based backoff algorithm. It is shown that the probability-based and window-based backoff algorithms are equivalent to each other. Moreover, we find that the characteristic equation of network throughput is invariant to backoff scheduling algorithms.

Last but not least, the proposed queueing model can be systematically generalized to investigate various types of MAC protocols, such as ALOHA, CSMA protocols, IEEE 802.11 protocols. Specifically, we illustrate the methodology by full analyses of the non-persistent CSMA and 1-persistent CSMA protocols in this thesis.

摘要

隨著網絡服務的日新月異，媒體存取控制協議亦必需同步更新，才能提供服務質數的保証。儘管媒體存取控制協議的發展迅速, 但是對其性能分析的技術與理論卻沒有很大的突破。研究學者們在這幾十年來一直希望能夠創造出一個能對媒體存取控制協議進行全面分析的方法論，可惜，迄今為止還沒有得出這麼一個全面的方法論。

過往，大部份的研究集中討論飽和網絡中的吞吐量和封包輸送延遲。而現實生活中，飽和網絡的情況其實極少出現。而我們對於正常網絡系統的穩定性等均沒有深入的認知。主要原因是由於網絡系統的性能是取決於用戶的流量與及補償調度演算法，這些都不是飽和系統所能準確詮釋。

建基於以上原因，在這篇論文裏我們為媒體存取控制協議建立了一個排隊模型。這個排隊模型最大的特點便是周詳地考慮了用戶流量和補償調度演算法對系統性能的影響，這正正已往研究所欠缺的。我們將每個用戶的輸入緩衝器模擬為一個 Geo/G/1 排隊系統，當中我們用了馬爾可夫鏈來描述每一個隊頭封包的服務時間分佈。透過這個排隊模型對非飽和網絡的演釋，我們導出網絡吞吐量的代表方程式、封包排隊的延遲、系統穩定性的條件等。我們更闡述了穩定吞吐量和有界延遲的區間，發現這與重發因子和輸入速率有著密切的關係。

除此以外，我們還研究了幾何重發和指數補償兩種演算法。在一個無限用戶

的環境下，我們証明了幾何重發是不穩定的。相反地，在滿足特定條件下，指數補償卻能提供穩定的網絡流量與及可接受的延遲保証。我們得出指數補償這種演算法絕對適合龐大用戶量的網絡。

這裏所提出的模型除可應用於概率式的補償演算法之外，還可實踐在窗口式的補償演算法上。透過對它們的模型的研究，我們証明了這兩種不同的形式其實是等同的。與此同時，我們亦揭露了網絡流量的公程式是不受任何的補償演算法所影響。

最後，論文所提的方法論更可被擴展到其他的媒體存取控制協議上，例如：ALOHA、 CSMA、 IEEE 802.11 等協議。而在這論文裏，我們則利用這個方法論來對 non-persistent 和 1-persistent CSMA 協議的性能進行全面和深入分析。

Acknowledgement

I would like to express my heartfelt thanks to doctoral my advisor, Professor Tony T. Lee, for his guidance, encouragement as well as many constructive discussions in the past few years. The first time I met Professor Lee was in his class "Telecommunication Switching and Network Systems", when I was an undergraduate student. I was deeply affected by his inspiring and enthusiastic teaching style, and believe that research is interesting and challenging. This is also the main reason that I continue my academic life. What I learnt during my postgraduate study is not only the problem solving technique, but also the positive attitude towards troublesome.

It is my pleasure to be one of the members in broadband laboratory, in which I have a lot of friendly and approachable lab-mates. They gave me a wonderful and enjoyable graduate school life. I would like to thank all of them, including Dongjie Yin, Yujie Wan, Liang Huang, Hao Guan, Martin Choi and Robert Ruan. Special thanks go to Dongjie Yin for the cooperation in this project.

To be honest, without the support and encouragement of my best friends, this thesis would be not accomplished. I sincerely express my thanks to Yiu Kwong Chan, Mau Fung Wong, Yulin Deng, Lok Man Law, and Astrid Chang.

Finally, I would like to express my deepest gratitude to my family, for their unconditional love and continuous support.

This work is dedicated to my dearest parents.

Contents

List of Figures

List of Tables

Abbreviations

Notations	**Descriptions**
$\bar{\lambda}$	Packet input rate of each node
n	Number of nodes in the system
$\hat{\lambda}$	Aggregate input rate
$\hat{\lambda}_{out}$	Network throughput
$\hat{\lambda}_{\max}$	Maximum network throughput
ρ	Offered load of each node
G	Attempt rate (the rate HOL packets attempt to access the channel)
G_S	Smaller root of the characteristic equation of network throughput
G_L	Smaller root of the characteristic equation of network throughput
q	Retransmission Factor
a	Mini-slot size
γ	Collision duration of the NP-CSMA/CD and 1P-CSMA/CD
x	Collision duration $= \begin{cases} 1 \text{ for the CA scheme} \\ \gamma \text{ for the CD scheme} \end{cases}$
K	Cut-off phase of the *K*-Exponential Backoff algorithm
α	Probability of idle channel
p	Probability of successful transmission
p_1	Probability of successful Type I transmission of the 1P-CSMA
p_2	Probability of successful Type II transmission of the 1P-CSMA
X	Service time of a HOL packet
$E[X]$	Mean service time (access delay)

$E[X^2]$	Second moment of service time
$E[T]$	Mean queueing delay
n_i	Number of packets in phase *i* sensing state
R_T	Stable throughput region
R_D	Bounded delay region

Chapter 1

Introduction and Historical Background

The Medium access control (MAC) protocol defines a standard that allows multiple users to share a common transmission channel, such as the 802.11 protocol for wireless local area networks (WLANs) and the 802.15.4 protocol for wireless sensor networks (WSNs). It plays an ever increasing important role in wireless communication systems, and has been continuously updated to keep up with the emerging new services and quality of service requirements. However, despite of the rapid changes of MAC protocols, a comprehensive performance analysis of MAC protocol remains an open issue. In this thesis, we focus on the study of CSMA protocols.

In section 1.1, we use a multi-queue single-server model to generalize the methodologies in literatures; indeed, the existing methodologies can be classified into three categories considering different aspects. A survey on these three categories of works is given in section 1.2. Through the review of literatures in section 1.1, we figure out the major difficulties in analyzing the performance of MAC protocols. And it turns out that the employed backoff scheduling algorithm which is briefly discussed in section 1.3 highly affects the performance of a MAC protocol. In addition, we show our contributions of this thesis in the section 1.4. A dissertation overview is given in section 1.5.

1.1 Multi-queue Single-server System

In this section, we are going to summarize the literatures by using a simple two-stage queueing model. Besides, this two-stage queueing model also reveals key ideas in performance analysis of MAC protocols.

The property of MAC protocols can be considered as a multi-queue single-server system with n number of nodes, as shown in Fig.1.1. We assume that the arrival of packets to each node is a Bernoulli process with rate λ; hence, the aggregate input rate $\hat{\lambda}$ is equal to $n\lambda$. Suppose the input buffer of each node is a first-in-first-out (FIFO) queue, then only the head-of-line (HOL) packets have ability to access the channel/server. According to the theorem on the superposition of point processes proved in [1] and [2], the aggregate attempts of all HOL packets form a Poisson process for large population. In other words, HOL packets form a fictitious queue with attempt rate G which is the rate HOL packets try to access the server. Once collisions happen, the collapsed HOL packets are backlogged and rescheduled for retransmission in which the waiting time is governed by the backoff scheduling algorithm.

Since the multi-queue single-server system is a special case of the input-buffered packet switches with single output channel, it can be described by this two-stage queueing model as mentioned in [3] and [4]. The first-stage queue is the exactly the input buffer, while the second-stage queue is the fictitious queue consisting of all HOL packets, as shown in Fig. 1.1. The two-stage queue brings us some important information:

1) The network throughput only depends on attempt rate G of the second-stage queue.
2) The attempt rate G is an implicit function of the packet input rate λ and the backoff scheduling algorithm employed in the first-stage queue.
3) The performance of input buffer depends on the packet input rate λ and the backoff scheduling algorithm.

For item 1, the server can be accessed by HOL packets only, and hence the network throughput is determined by the activities of Head-Of-Line (HOL) packets in the entire network.

The attempt rate G can be interpreted as the rate that HOL packets try to access the channel. An empty node accesses the channel with rate equals the input rate λ, while the rate that a backlogged node sends packets is governed by the backoff scheduling algorithm. Hence, the attempt rate G is dependent on both the input rate λ and backoff scheduling algorithm, i.e., the item 2.

On the other hand, the service time distribution of the HOL packet depends on the backoff scheduling algorithm when collision occurs. It is the reason that the packet access delay, which is the time a HOL packet to complete its service, is governed by the backoff scheduling algorithm. According to the Little's law, the packet queueing delay which is defined as the time for a newly arrived packet to complete its service depends not only on the backoff scheduling algorithm but also the packet input rate. As a result, it is impossible to achieve a comprehensive analysis of the system without considering the impact of the input traffic and the backoff scheduling algorithm on the input queueing behavior.

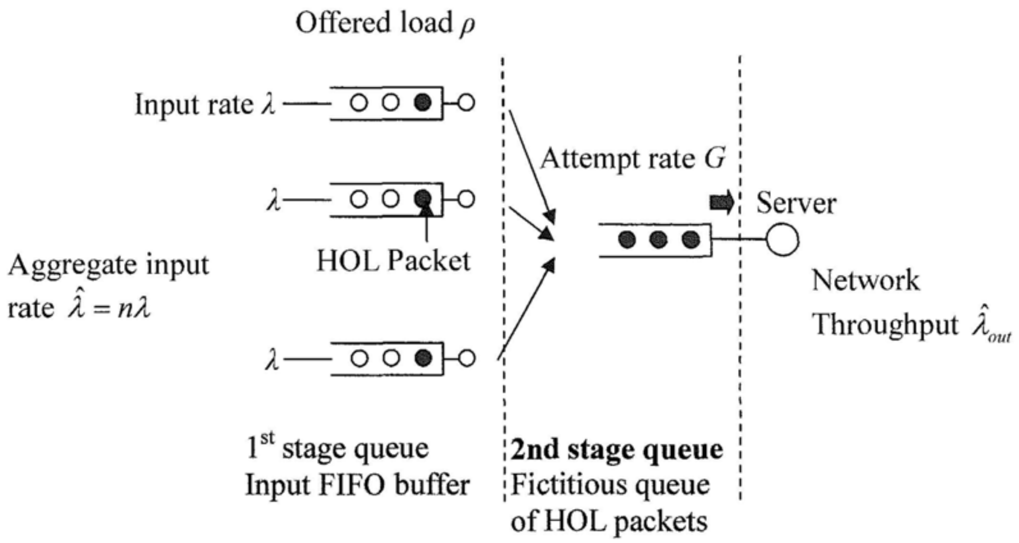


Fig. 1.1. A two-stage queueing model of multi-queue single-server systems.

Based on the two-stage queueing model, literatures can be classified into the following three categories:

Category I: consider on channel behaviors (the second stage queue) only;

Category II: consider backoff scheduling algorithms employed in the first stage queue, but ignore the input traffic pattern.

Category III: consider both backoff scheduling algorithms and input traffic.

The results obtained from these categories are summarized in the TABLE 1.

Most existing studies [5] – [17] focused on the channel behaviors, that is, the second-stage queue. These belong to the category I. Although an overall network throughput was easily obtained, the stability and bounded delay conditions of each individual input queue cannot be comprehended by those simplified models. This point was also clearly stated in [49] that focusing on aggregate attempt rate G only is not enough to study the effect of retransmission policy.

Categories	**Examples**	**Results**
Category I	[5] – [17]	Network throughput
Category II	[19] – [37]	Network throughput
		Packet access delay
Category III	[38] – [49]	Packet access delay
		Packet queueing delay

Table 1.1. A summary of the three categories of researches.

The category II greatly improved the methodology in the category I by modeling the backoff scheduling algorithm. However, the works of this category (sees [19] – [37]) do not include input traffic pattern by assuming a saturated traffic situation. Under the saturation assumption, the first-stage queue of input buffer can be ignored because each input node always has an active HOL packet. In other words, these throughput analyses can be considered as a special case of our second-stage queue, while the first stage queues are all busy. This assumption substantially simplifies the maximum throughput analysis, but the queueing behavior of each individual node is missing in the model. Since the backoff scheduling algorithm is included, the packet access delay can be derived too. In practice, these results may not be very useful because the queueing delay in each input buffer could be unbounded if the network is saturated. In a nutshell, any models of the network that ignored either the queueing behavior of the input buffer or the input traffic pattern were unable to completely tackle the stability and delay issues of the protocol.

The category III focuses on the backoff scheduling algorithm and input traffic pattern at the same time. Even though some existing studies of this category [38] – [49] obtained the queueing delay of MAC protocols by considering the input traffic, the stability issues remain open problems. More importantly, we do not have a systematic approach which gives us network throughput, queueing delay as well as stability analysis.

In this thesis, we are going to introduce a novel model for Markov modulated MAC protocols, which allow us to conduct a comprehensive queueing analysis by modeling the first stage input buffer. Before going to the details of our work, the following is a survey of related works.

1.2 Carrier Sense Multiple Access CSMA protocols

The Carrier Sense Multiple Access (CSMA) protocol is one of the most widely deployed MAC protocols that allow multiple users to share a common transmission channel. It can be implemented in various areas, such as IEEE 802.11 WLANs and Ethernets. The CSMA protocol includes two kinds of collision resolving schemes, namely the collision detection (CD) and collision avoidance (CA). For the collision detection scheme, terminals can stop transmissions if collision occurs. For the counterpart, collisions can be noticed if there is no acknowledgment message received. The collision detection scheme is suitable for wired networks, while the collision avoidance scheme is applied in wireless networks.

Basically, the CSMA protocol is a set of rules for resolving collisions when two or more nodes attempt to use a transmission channel simultaneously. The network stations governed by CSMA check the busy/idle status of the transmission channel before attempting to transmit. If either the channel is detected busy or the packets involved in the collision, the corresponding station attempt to retransmit after a random time interval which is governed by the backoff scheduling algorithm.

Category I

The complete analysis of a buffered random-access network has long been considered to be a challenging problem even for the simplest case. Since Abramson's landmark paper on the ALOHA systems [5] embarked on the throughput analysis of multiple access channels in 1970, many researchers have followed his approach in studying

CSMA protocols. Their approaches mainly based on the renewal theory, sees [6] – [14].

The analysis of slotted CSMA can be traced back to Kleinrock and Tobagi's paper [6] in 1975. Under the assumption of infinite numbers of nodes and Poisson distributed the aggregate transmission (including new and backlogged packets), they derived the throughput of various CSMA protocols by using an approach based on renewal theory. Although the characteristic equation of throughput can be easily obtained by using this elegant method, the ability to guarantee a stable throughput of the entire network and to bound the mean delay of packets are still open problems when analyzing the performance of CSMA protocols.

Almost all follow-up researches on the performance analysis of CSMA protocols [7] – [11] have been based on the network model originally proposed in [6]. For example, a finite population system with the slotted p-Persistent CSMA (pP-CSMA) protocol, which can also be extended to infinite population cases, was analyzed in [7]. Network throughputs of various CSMA-typed protocols were studied in [8], in which the transmission channel was considered as three cases: idle, collision and successful transmission. By using a state-transition diagram of the transmission channel, the characteristic equation of network throughput of the CSMA was re-devised in [8]. Authors of [9] used a similar approach to find the capacity of the 802.11 MAC protocol. The saturation throughput for Non-Persistent CSMA (NP-CSMA) protocols was deduced in [10] from the semi-Markovian property of the protocols. All these approaches [5] – [11] focused only on the aggregate transmission, the results obtained were limited to network throughput.

Researches of [11] – [12] extended the above principle to analyze the performance of CSMA-typed protocols in terms of various aspects, such as the network throughput and energy consumption. Authors of [11] proposed an alternative CSMA protocol with an adaptive array to improve the network throughput. The analytical method based on the attempt rate G only; the study was limited to network throughput. The system performance with respect to energy consumption was given in [12], based on renewal theory.

As reported in [13] that different input traffic patterns could lead to different network behaviors. Approaches with modeling input traffic patterns were provided in [13] – [17]. The references [14] and [15] evaluated the transient dynamic behavior of CSMA protocols with impulse traffic. A performance analysis of a buffer-less CSMA protocol was given in [16] by a simple Markov chain, in which the state-space was the number of idle and backlogged nodes in the system. In [17], the authors gave a throughput evaluation of asynchronous data transfer protocols in the present of hidden terminals [6].

It is arduous to study the delay issues without taking care of the service time distribution that incorporates the scheduling algorithm. In the past several decades, researchers focused on the network throughput analysis (sees [5] – [17]), but ignored the impact of backoff algorithms. We classify these works as the category I.

Category II

The works of category II were started from an analysis of the 802.11 distributed coordination function (DCF) [18], which is one of the most popular implementation of CSMA. The 802.11 protocol includes specifications for both MAC and physical layers. The DCF defines two access mechanisms, namely the Basic access mechanism and the request-to-send/clear-to-send (RTS/CTS) access mechanism. The Basic access mechanism is simply a two-way handshaking scheme, in which the destination node acknowledges the successful receipt by sending an ACK frame. On the other hand, the RTS/CTS access mechanism uses a four-way handshaking method to eliminate the hidden node problem. Before sending the packet payload, the source node broadcasts a request-to-send (RTS) short message to request channel resources. If no collision occurs, the destination node sends a clear-to-send (CTS) short message to indicate that the channel is reserved. The source node starts to transmit the packet payload after receiving the CTS. The source node also receives an ACK frame if the packet payload is successfully transmitted.

A two-dimensional Markov chain that described the binary exponential backoff algorithm [19] was proposed by Bianchi in 2000. It was a groundbreaking work in analyzing MAC protocols. This drastically different approach which belongs to the category II provided extensive throughput analysis of a saturated 802.11 DCF network in which nodes always have packets to transmit. This saturation assumption substantially simplified the analysis by ignoring packet arrival processes, but failed to comprehend the complete queueing behaviors of a single node, such as stability and delay, issues concerning regular network operation under non-saturated condition.

Abundant published works on the performance of CSMA protocols under different scenarios (see [20] – [30]) were stimulated by Bianchi's two-dimensional Markov chain model [19]. In [20], the authors used the same model to derive the network throughput with retry limit. The saturation throughput for a fading channel by considering the probability that packets collapsed due to frame error was studied in [21], which resulted in smaller network throughput as expected because of the larger probability of collapsed transmissions. Moreover, the authors in [22] and [23] applied Bianchi's two-dimensional Markov chain model to analyze the network throughput of an alteration of the 802.11 protocol.

Access delays of slotted ALOHA and CSMA with various backoff scheduling algorithms are given in [24]. A frame delay analysis on the 802.11 DCF protocol was introduced in [20], [25] and [26], which only covered the mean service time under the network saturation condition. Under the same saturated assumption, [27] – [30] studied the access delay of the 802.11 protocol in cooperation with the exponential backoff scheduling algorithm. In particular, the access delay for the 802.11 with binary exponential backoff (BEB) was studied using a bottom-up approach in [29]. Furthermore, the access delay of the 802.11 DCF protocol has been studied in [30] by considering the generating function of packet service time distribution.

Various modified backoff schemes of 802.11 protocols were considered in [31] and [32] to enhance system performance. An exponential increase and linearly decrease backoff algorithm was proposed for the 802.11 DCF in [31] to enhance the saturated throughput. An advanced version of the exponential increase and linearly decrease backoff algorithm has been integrated into a new collision resolution mechanism (called GDCF) which was proposed in [32]. Both algorithms did not resume the backoff stage to the initial zero stage after a successful transmission. The algorithm in [31] only allows the corresponding station to reduce its backoff stage after a single successful transmission, while the GDCF in [32] requires the station to reduce its backoff stage after c consecutive successful transmissions. These modified backoff algorithms can improve network throughput in a saturated condition, but their performances in a non-saturated 802.11 network are still unknown.

Other than the 802.11 protocol, the throughput analysis of the IEEE 802.15.4 protocol was given in [33] by using a modified 2D Markov chain [19]. Some cross-layer analytical approaches for CSMA/CA protocols were proposed in [34] – [37], which gave us the insight into the network throughput in terms of physical and MAC layers.

Although these works of category II [19] – [37], have revealed the connection between the maximum throughput and the underlying backoff algorithm, issues on throughput stability and bounded delay remain open under the saturation assumption.

Category III

To support real-time applications such as video streaming, queueing delay issues should be taken into account. With this aim, researchers were interested in modeling an unsaturated traffic system [38] – [49].

A buffer-less system was studied in [38], in which the unsaturated environment was described by an extra idle state. The buffer-less property can help to avoid complicated queueing analysis, but this simplified assumption is not useful in practice. An analytical model with small buffers for 802.11 DCF in the presence of unsaturated heterogeneous conditions was proposed in [39] and [40], in which the characteristic equation of network throughput was found to be the same as that under the saturated condition. The above of results [38] – [40] are consistent with the results in this thesis that network throughput equation is the same in saturated and unsaturated environments.

Furthermore, a model with finite load sources to evaluate the fairness and throughput for a multi-rate system was presented [41]. Similar as the approach in [38], authors of [42] – [45] studied the performance (in terms of throughput, access delay, and energy consumption) of an unsaturated 802.15.4 system by adding an extra idle state into the Markov chain introduced in [33]. Due to the relaxation the input traffic restriction, queueing delay of the 802.11 wireless networks was obtained in the paper series [47] – [49]. All these analyses for unsaturated networks [38] – [49], however, did not address stability issues.

In facing the difficulty of precisely modeling the system, many simulation results on the performance of CSMA were reported in [50] – [55], which obtained the simulated network throughput, packet transmission delay and etc under various system situations.

As the MAC protocol is being continuously updated to meet the requirements of different types of emerging new services, a model of the MAC protocol should be flexible enough to cope with the changes of QoS requirements. Issues on the throughput stability and bounded delay conditions remain open problems. In this thesis, we show that stability and queueing delay issues can be resolved by our model under a non-saturated condition.

1.3 Backoff Scheduling Algorithms

Collisions are unavoidable in a shared environment, and a contention resolution algorithm is an essential mechanism in any MAC protocol to resolve contention among active nodes. However, category I in [5] – [17] did not take the scheduling of collided packets into consideration.

Exponential Backoff algorithm is a widely adopted collision resolution scheme in most MAC protocols. Retransmission probability of the collapsed packet decrease exponentially with number of encountered collision times. A packet is in backoff phase i, if and only if it has encountered collision i times. When a packet is successfully transmitted, all follow-up packets are immediately able to access the channel. In other words, any follow-up packet starts in the initial phase.

It was shown in [56] that Exponential Backoff for MAC protocols was unstable for an infinite number of nodes. The authors of [57] claimed that if the arrival rate of the system is sufficiently small, the Exponential Backoff can still be stable. Despite these unfavorable observations, it was proven in [58] – [61] that a stable throughput can be achieved by the Exponential Backoff even if the number of nodes goes to infinity. Currently, a consensus on the stability issue of the Exponential Backoff scheme cannot be concluded from these controversial results.

Other researchers were also interested in comparing the characteristics of different backoff algorithms, such as linear, polynomial, binary exponential, exponential increase and linear decrease (EILD), exponential increase and exponential decrease (EIED) backoff algorithms and etc. As the name suggests, retransmission factors of the linear and polynomial backoff algorithms increase with collision times in linear and polynomial manners respectively. The binary exponential backoff is only a special case of the exponential backoff algorithm, in which the retransmission factor equals 1/2. On the other hand, EILD and EIED are same as the exponential backoff; except the starting phase of follow-up packets is not the initial phase. When a successful transmission happens, the backoff phase of the follow-up packet decreases linearly and exponentially for EILD and EIED respectively.

In [31] – [32] and [62] – [65], the authors claimed that the proposed backoff algorithms out-performed the binary exponential backoff algorithm. The basic principle of the proposed backoff algorithms was to reduce the packet traffic rate in a saturated traffic condition which is not realistic. It has been discussed in [43] that a smaller retransmission probability is suitable for the saturated network. However, an unsaturated network requires a larger retransmission probability for better performance.

Although a lot of existing paper claimed that their modified backoff algorithms were better than the binary exponential backoff. However, the evidences provided were based on a saturated traffic condition. Another reason is that complicated backoff algorithms are difficult to implement. In this thesis, we are going to consider an exponential backoff algorithm which is more general than the binary exponential one.

1.4 Contributions

We propose a generic queueing model which gives a comprehensive queueing analysis of the non-persistent and 1-persistent CSMA protocols with backoff scheduling algorithms under an unsaturated traffic environment. We assume that the input to each node is a Bernoulli process with rate λ packet/timeslot and each input buffer is modeled as a Geo/G/1 queue. The service time of HOL packets is described by a Markov chain, from which we obtain the service time distribution of HOL packets.

Based on the probability theory and the Markov chain, we obtain the characteristic equation of network throughput as a function of the attempt rate G of HOL packets. In particular, we prove that the throughput equation in saturated and unsaturated conditions remains the same.

The service time distribution of HOL packets from the proposed Markov chain gives us the moment generating function of packet service times. The packet access and queueing delays are immediately derived by the use of Little's Law and Pollaczek-Khinchin (P-K) formula of Geo/G/1 queue.

Furthermore, we also determine stable throughput and bounded delay conditions. The stable throughput condition is based on the characteristic equation of network throughput and the first moment of service time (or the offered load of each node).

While a bounded mean delay requires a bounded second moment of service time. The two conditions enable us to specify the stable regions of both throughput and bounded mean delay in terms of the retransmission factor for given input traffic. This reveals that stability of a system is highly dependent on the backoff scheduling algorithm as well as the input traffic.

Our results indicate that the stable throughput region with Exponential Backoff exists even for an infinite population, which agrees with the presentation in [58] – [61]. Hence, this scheduling scheme is robust and can scale to large number of nodes. However, the bounded delay region is only a subset of the stable throughput region, and the maximum achievable throughput of the network within this region is slightly smaller than the absolute maximum throughput. For network operating inside the stable throughput region but not within the bounded delay region, we show that a stable throughput of the entire network can be achieved with a large delay variance due to the "capture effect" described in [61] and [67]. Thus, the maximum throughput of the network within the bounded delay region is slightly smaller than that achievable in the stable throughput region.

We find that the 1-persistent CSMA protocol is more suitable for network with little input traffic; while the non-persistent CSMA protocol is more outstanding in higher input traffic situation.

1.5 Dissertation Overview

In chapter 2, we give a comprehensive performance analysis of non-persistent CSMA/CD and CSMA/CA protocols. The characteristic equation of network throughput, packet transmission delay and stability issues of the non-persistent CSMA protocol are all included in this chapter. In addition, we extend the model in chapter 2 to study the performance of 1-persistent CSMA/CD and CSMA/CA protocols, as given in chapter 3. A window-based binary exponential backoff scheduling algorithm is modeled in chapter 4, in which we show that window-based algorithm is equivalent to the probability-based one. The last chapter is the summary and future development.

Chapter 2

Non-persistent Carrier Sense Multiple Access Protocols

We introduce queueing models of input buffer to work out a comprehensive performance analysis of non-persistent carrier sense multiple access protocols (NP-CSMA). The queueing model considers employed scheduling algorithms under a non-saturated traffic environment.

2.1 Basic principle of the Non-persistent CSMA protocol

The NP-CSMA protocol is a kind of MAC protocols used in many communications networks. The major property of the carrier sensing protocol is that every node takes action after sensing the channel status. If the channel is in an idle state, the packet is sent immediately; otherwise, the packet is delayed for a random amount of time before sensing the channel again. Collisions may occur when two or more packets are simultaneously transmitted. There are several assumptions we made in this chapter:

1. The system is slotted into mini-slots in which activities can be taken only at the beginning of mini-slots,
2. The network is well synchronized.
3. Packet collapsed only due to simultaneous transmission.

The NP-CSMA protocol contains two different schemes, namely the collision detection (CD) and collision avoidance (CA). The CSMA/CA protocol is widely used in wireless communications. The collision avoidance scheme does not allow nodes to detect interference among transmissions. In other words, source nodes notice collision if there is no acknowledgement (ACK) message. For sake of simplicity, we assume the time for receiving an ACK message is negligible, that is, the time required for successful transmission is same as that for unsuccessful transmission in the case of the CA scheme. Traditionally, the collision detection scheme is applied in wired network, in which nodes can monitor the transmission progress. Nodes can then stop transmissions when a collision is detected. Hence, the collision duration of the collision detection scheme is less than the successful transmission duration.

2.2 Channel Model

In a slotted NP-CSMA network, the time axis is slotted and the network is synchronized. Packets can be sent only at the beginning of a timeslot, and each packet transmission takes one slot time. The ratio of propagation delay to packet transmission time is a; hence, the timeslot can be further divided into mini-slots of slot size a. According to the channel status, the time axis can be considered as a sequence of alternating busy periods and idle periods as shown in Fig. 2.1.

As shown in Fig. 2.1(a), the duration of a busy period for the collision avoidance scheme is $1 + a$ timeslots because no packets will attempt to transmit in the last mini-slot of a busy period. The successful transmission time of the collision detection scheme is the same as the counterpart. However, the collision time of CD scheme is shorter, because nodes can stop transmissions if collision is detected, as shown in Fig. 2.1(b). Let γ be the time required to detect and abort the collision (γ is less than 1), then the duration of the corresponding busy period is $\gamma + a$.

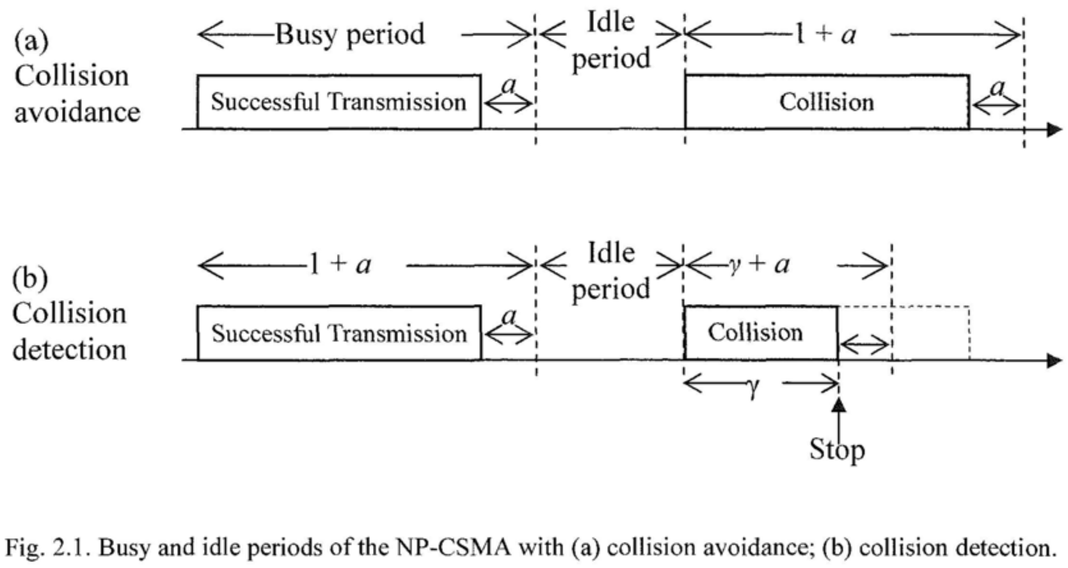


Fig. 2.1. Busy and idle periods of the NP-CSMA with (a) collision avoidance; (b) collision detection.

For both schemes, the length of an idle period is a random variable with geometric distribution. In each busy period, either a packet is successfully transmitted or a collision occurred. Therefore, the channel will be in one of three states: idle (Idle), successful transmission (Suc), and collision (Col). The state transitions of the channel can be described by the Markov chain shown in Fig. 2.2.

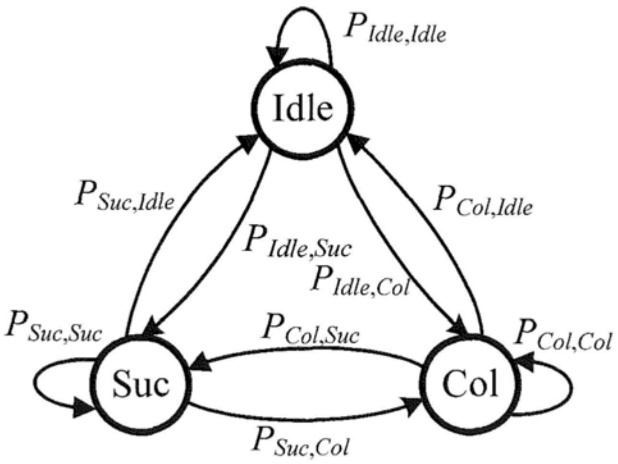


Fig. 2.2. Markov Chain of NP-CSMA channel.

The limiting probabilities π_{Idle}, π_{Suc}, and π_{Col} of the Markov chain satisfy the set of equations:

$$\begin{cases} \left(1 - P_{Idle,Idle}\right) \cdot \pi_{Idle} = P_{Suc,Idle} \cdot \pi_{Suc} + P_{Col,Idle} \cdot \pi_{Col} \\ \left(1 - P_{Suc,Suc}\right) \cdot \pi_{Suc} = P_{Idle,Suc} \cdot \pi_{Idle} + P_{Col,Suc} \cdot \pi_{Col} \\ \left(1 - P_{Col,Col}\right) \cdot \pi_{Col} = P_{Idle,Col} \cdot \pi_{Idle} + P_{Suc,Col} \cdot \pi_{Suc} \end{cases} . \tag{2.1}$$

Since packets attempt to access the channel only when the channel is detected as idle, the attempt rate in any busy period is zero. According to the theorem on the superposition of point processes proved in [1], the aggregate attempts form a Poisson process for large population. During an idle period, the aggregate attempts generated by all fresh and re-scheduled HOL packets is a Poisson stream with rate G, and the probability that no attempts is generated in a mini-slot is e^{-aG}. In the channel point of view, a transmission period is successful only if there is exactly one attempting packet in a mini-slot with the probability aGe^{-aG}. Hence, the transition probabilities of the Markov chain are given by

$$\begin{cases} P_{Idle,Idle} = P_{Suc,Idle} = P_{Col,Idle} = e^{-aG} \\ P_{Suc,Suc} = P_{Idle,Suc} = P_{Col,Suc} = aGe^{-aG} \\ P_{Col,Col} = P_{Idle,Col} = P_{Suc,Col} = 1 - e^{-aG} - aGe^{-aG} \end{cases} . \tag{2.2}$$

It is straightforward to show that the limiting probabilities are given as follows:

$$\pi_{Idle} = e^{-aG}, \quad \pi_{Suc} = aGe^{-aG} \text{ and } \pi_{Col} = 1 - e^{-aG} - aGe^{-aG} . \tag{2.3}$$

The time-average probabilities of each state can be easily obtained from sojourn times of Idle, Suc, and Col states

$$t_{Idle} = a; \; t_{Suc} = 1 + a; \; t_{Col} = \begin{cases} 1 + a & \text{for CA} \\ \gamma + a & \text{for CD} \end{cases} . \tag{2.4}$$

Corresponding to the difference of the collision duration between CA and CD schemes, we define a variable x as below,

$$x = \begin{cases} 1 & \text{for CA} \\ \gamma & \text{for CD} \end{cases} . \tag{2.5}$$

Of particular importance, the probability that the channel is in Suc state is given by

$$\tilde{\pi}_{Suc} = \frac{t_{Suc}\pi_{Suc}}{t_{Suc}\pi_{Suc} + t_{Col}\pi_{Col} + t_{Idle}\pi_{Idle}}$$
$$= \frac{(1+a)aGe^{-aG}}{aGe^{-aG} + a + x\left(1 - e^{-aG} - aGe^{-aG}\right)}. \tag{2.6}$$

Since the successful transmission of a packet only takes $\frac{1}{1+a}$ of the time in a busy period, the throughput of the network is defined by the fraction of the time that the channel is productive:

$$\hat{\lambda} = \frac{\tilde{\pi}_{Suc}}{1+a}. \tag{2.7}$$

By substituting (2.5) and (2.6) into (2.7), we have the characteristic equation of network throughput with respect to the CA and CD schemes.

$$\hat{\lambda} = \begin{cases} \dfrac{aGe^{-aG}}{1 - e^{-aG} + a} & \text{for CA} \\ \dfrac{aGe^{-aG}}{aGe^{-aG} + a + \gamma\left(1 - e^{-aG} - aGe^{-aG}\right)} & \text{for CD} \end{cases}. \tag{2.8}$$

The throughput of the CA scheme is consistent with Kleinrock and Tobagi's results [6], while that of the CD scheme agrees with Rom and Sidi's results [8].

Although the network throughput can be obtained from the above model, it is by no means a comprehensive performance analysis of the system, because this model does not consider the re-scheduling of collided HOL packets. The next section describes a more detailed queueing model of the input buffer in which the service time distribution is derived in cooperation with the backoff scheduling scheme.

2.3 Queueing Model of Input Buffer

The input buffer of each node is modeled as a Geo/G/1 with Bernoulli arrival process of rate λ packets/timeslot. In the NP-CSMA protocol, we consider a general backoff scheme, called the *K*-Exponential Backoff algorithm, for contention resolutions. The *K*-Exponential Backoff algorithm allows an HOL packet in phase i to retransmit with probability q^i, for $i = 1, \ldots, K$, where $0 < q < 1$ is the retransmission factor and K is the cut-off phase. A backlogged HOL packet is in phase i if it has encountered collisions i times. That is, the retransmission probability decreases exponentially with the number of collisions, up to K times, experienced by the HOL packet.

The activities of an HOL packet with this backoff algorithm are illustrated as a flow chart shown in Fig. 2.3. A fresh HOL packet is sent only if an idle channel is detected; otherwise, it waits for a waiting time to sense the channel activity again. A transmission is successful and returns to phase 0 if no other packet is sent at the same time. All collapsed packets are backlogged and increase its phase by one. These backlogged packets are re-scheduled to access the channel at a random later time which depends on the current backoff phase.

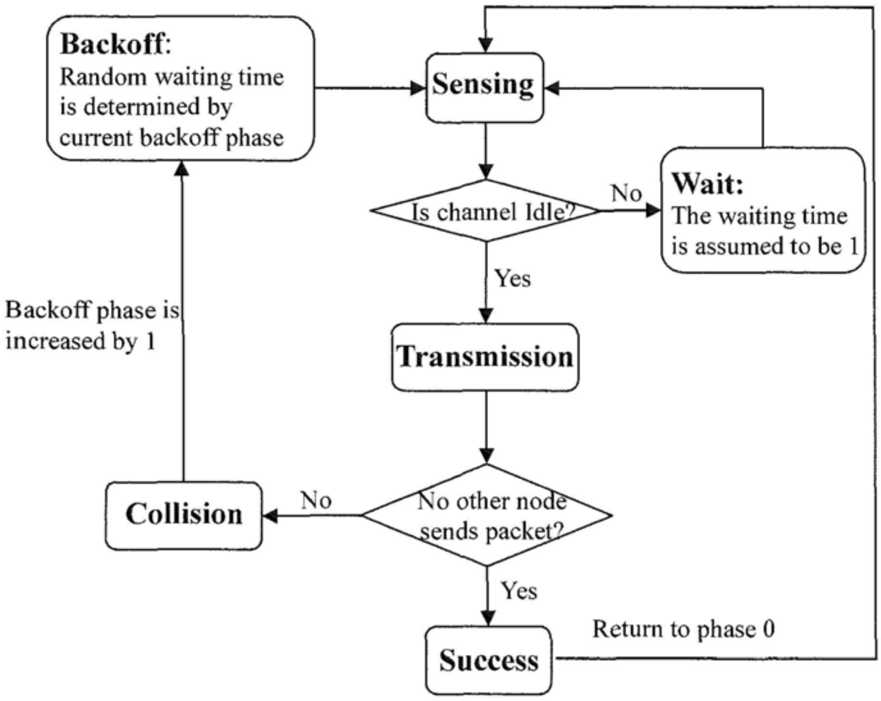


Fig. 2.3. Flow chart of access behaviors for HOL packets of the NP-CSMA.

Let α be the probability that the channel is being sensed idle. Besides idle periods, an idle channel is also detected in the last mini-slot of each busy period because of the propagation delay. Therefore, the time-average probability that the channel is sensed idle can be calculated from the limiting probabilities given by (2.3) as follows:

$$\begin{aligned}\alpha &= \frac{a \cdot \pi_{Idle} + a \cdot \pi_{Suc} + a \cdot \pi_{Col}}{a \cdot \pi_{Idle} + (1+a)\pi_{Suc} + t_{Col}\pi_{Col}} \\ &= \frac{a}{aGe^{-aG} + a + x\left(1 - e^{-aG} - aGe^{-aG}\right)}\end{aligned} \quad . \tag{2.9}$$

As shown in Fig. 2.4, the signal received by a node is shifted one mini-slot due to the propagation delay. Thus, the idle mini-slot immediately following a transmission period is still considered to be a part of the busy period by HOL packets. Also because of the propagation delay, a transmission can be successful only if all other nodes do not attempt to send packets in the first mini-slot of the busy period. Thus, the probability of a successful transmission is

$$p = e^{-aG}. \tag{2.10}$$

The equation (2.9) can be expressed in terms of p as below,

$$\alpha = \frac{a}{-p\ln p + a + x(1 - p + p\ln p)}. \tag{2.11}$$

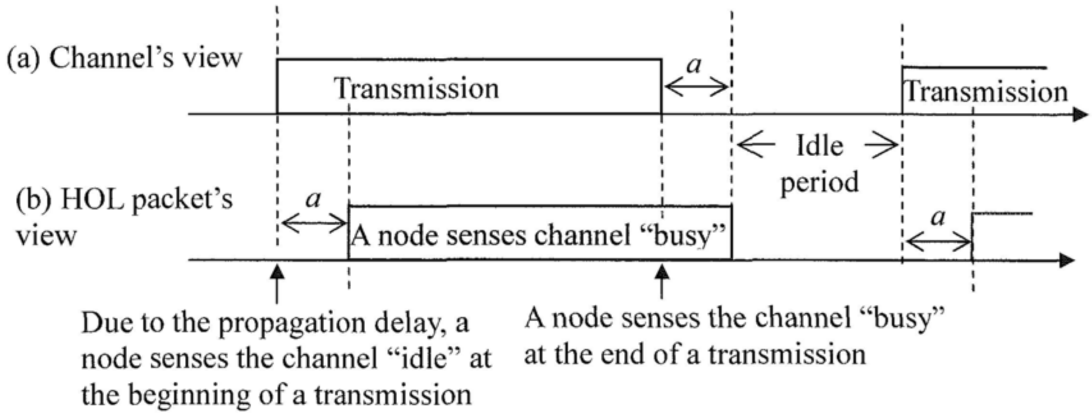


Fig. 2.4. Busy and idle periods of NP-CSMA protocol viewed by (a) the channel; (b) the HOL packet.

In the NP-CSMA protocol, the duration of successful transmission is the same for both CA and CD schemes, then a successful transmission (Suc) is used to represent this activity. Besides, an HOL packet in phase i can be in one of the three states: sensing (S_i), collision (C_i), and waiting (W_i), $i = 0, 1, \ldots, K$ at any time. The state transition diagram is shown in Fig. 2.5. If a busy channel is sensed, the packet is in waiting state W_i. If an idle channel is detected in the sensing state S_i with probability α, then the packet will be transmitted with probability q^i. With probability p, it is successfully transmitted in the state Suc. After successful transmission, a fresh HOL packet starts with the initial sensing state S_0. On the other hand, collision occurs with probability $1 - p$, the collided packet in the state C_i is moved to phase $i + 1$ and the above process repeats starting from the sensing state S_{i+1}.

We assume that a new packet arrived at an empty buffer will be delayed for one slot time in the waiting state if the channel is busy. The duration of this waiting time is not specified in non-persistent CSMA protocol. In fact, the sojourn time of the waiting state does not affect our analysis. Different waiting time may result in different delays, but the analyses of the stable throughput region and bounded delay region remain the same. Hence, the waiting time is assumed to be 1 slot time for the sake of simplicity.

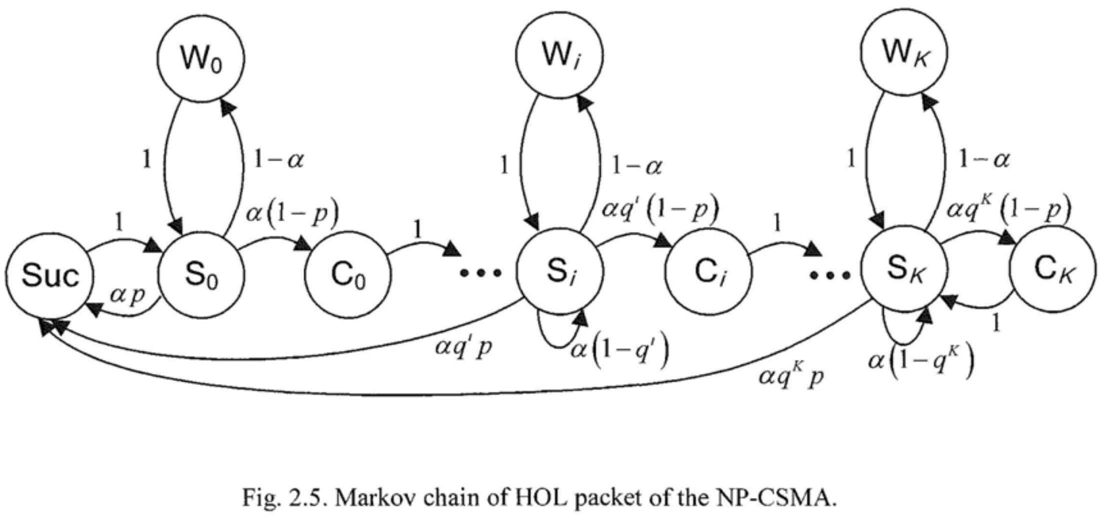


Fig. 2.5. Markov chain of HOL packet of the NP-CSMA.

Let b_{Suc}, b_{S_i}, b_{C_i}, and b_{W_i} be the limiting probabilities of states Suc, S_i, C_i, and W_i, respectively. From the Markov chain described in Fig. 2.5, we obtain the following set of state equations:

$$\begin{cases} b_{Suc} = \alpha p\left(b_{S_0} + q b_{S_1} + \cdots + q^K b_{S_K}\right) \\ b_{S_0} = b_{W_0} + b_{Suc} \\ b_{S_K} = \alpha\left(1-q^K\right) b_{S_K} + b_{W_K} + b_{C_{K-1}} + b_{C_K} \\ b_{S_i} = \alpha\left(1-q^i\right) b_{S_i} + b_{W_i} + b_{C_{i-1}}, & \text{for } i = 1, \ldots, K-1 \\ b_{W_i} = \left(1-\alpha\right) b_{S_i}, & \text{for } i = 0, \ldots, K \\ b_{C_i} = \alpha q^i \left(1-p\right) b_{S_i}, & \text{for } i = 0, \ldots, K \end{cases}. \tag{2.12}$$

It can be observed from (2.12) that if $p + q > 1$ then all states of the Markov chain are positive recurrent and aperiodic. Thus, the time-average probabilities of those states can be determined from (2.12) with the sojourn time $t_{Suc} = 1$, $t_{S_i} = a$, $t_{C_i} = x = \begin{cases} 1 & \text{for CA} \\ \gamma & \text{for CD} \end{cases}$, and $t_{W_i} = 1$ of states Suc, S_i, C_i, and W_i, respectively, for $i = 0, \ldots, K$ as follows:

$$\tilde{b}_{Suc} = \frac{\alpha p}{D}$$

$$\tilde{b}_{S_i} = \begin{cases} \frac{ap}{q^i D}(1-p)^i & \text{for } i = 0,\ldots K-1 \\ \frac{a}{q^K D}(1-p)^K & \text{for } i = K \end{cases}$$

$$\tilde{b}_{W_i} = \begin{cases} \frac{p(1-\alpha)}{q^i D}(1-p)^i & \text{for } i = 0,\ldots K-1 \\ \frac{(1-\alpha)}{q^K D}(1-p)^K & \text{for } i = K \end{cases}$$

$$\tilde{b}_{C_i} = \begin{cases} \frac{\alpha p x}{D}(1-p)^{i+1} & \text{for } i = 0,\ldots K-1 \\ \frac{\alpha x}{D}(1-p)^{K+1} & \text{for } i = K \end{cases} \quad , \tag{2.13}$$

where

$$D = q\left(\frac{1+a-\alpha}{p+q-1}\right)\left[p-(1-q)\left(\frac{1-p}{q}\right)^{K+1}\right]+\alpha\left[p+x(1-p)\right] \quad . \tag{2.14}$$

The offered load ρ of each input queue is the probability that the queue is non-empty, in which $\rho < 1$ is a basic requirement for a stable input buffer. The input rate λ of the Bernoulli arrival process can be interpreted as the probability of finding a packet arrived at input in any time slot. Each input packet will eventually become a fresh HOL packet, and visit the successful transmission state (Suc) for one slot time. Therefore, the input rate λ should equal $\rho\tilde{b}_{suc}$ the probability of finding an HOL packet in the successful transmission state in any time slot. With the time-average probability of state Suc given in (2.13), the expression of the offered load can be obtained as follows:

$$\rho = \frac{\lambda}{\tilde{b}_{Suc}} = \begin{cases} \dfrac{\lambda q(1+a-\alpha)}{\alpha p(p+q-1)}\left[p-(1-q)\left(\dfrac{1-p}{q}\right)^{K+1}\right]+\dfrac{\lambda}{p} & \text{for CA} \\ \dfrac{\lambda q(1+a-\alpha)}{\alpha p(p+q-1)}\left[p-(1-q)\left(\dfrac{1-p}{q}\right)^{K+1}\right]+\lambda+\lambda\gamma\dfrac{1-p}{p} & \text{for CD} \end{cases}. \quad (2.15)$$

From (2.15), we can observe that the mean service time of each HOL packet $E[X]$ is equal to $\tilde{b}_{suc}{}^{-1}$. We show in the next theorem that the network throughput derived from the queueing model of input buffer is the same as (2.7), which was obtained from the channel model. Although the service time is obviously dependent on the retransmission factor q of the backoff scheduling algorithm, the fact that the throughput is invariant with respect to q implies that the stability of the system cannot be determined by the characteristic equation of throughput alone; it is mainly related to the queueing behavior of each input buffer.

Theorem 1. *For buffered non-persistent CSMA/CA with K-Exponential Backoff, the throughput in equilibrium is given by*

$$\hat{\lambda}_{out} = \frac{-p\ln p}{1+a-p} = \frac{aGe^{-aG}}{1+a-e^{-aG}}. \quad (2.16)$$

Proof: For NP-CSMA protocol, a node is ready to send an HOL packet only if an idle channel has been detected. The probability of successful transmission from a desired node is the conditional probability that none of other nodes access the channel given that all nodes sense the channel idle,

$$\begin{aligned} p &= \Pr\{\text{none of other } n-1 \text{ nodes access the channel} \mid \text{channel is sensed idle}\} \\ &= \frac{\Pr\{\text{none of other } n-1 \text{ nodes access the channel}\}}{\Pr\{\text{channel is sensed idle}\}}. \end{aligned} \tag{2.17}$$

If no one access the channel, it means that all other $n-1$ nodes are either empty, or in sensing states but not accessing the channel. Thus, we have

$$\begin{aligned} &\Pr\{\text{none of other } n-1 \text{ nodes access the channel}\} \\ &= \left[\Pr\{\text{node is empty}\} + \Pr\{\text{node is in sensing but not scheduled to send packet}\}\right]^{n-1} \\ &= \left[(1-\rho) + \rho \sum_{i=0}^{K} \tilde{b}_{S_i}\left(1-q^{i}\right)\right]^{n-1} \\ &\overset{\text{for large n}}{=} \exp\left\{-n\rho\left[1 - \sum_{i=0}^{K} \tilde{b}_{S_i}\left(1-q^{i}\right)\right]\right\}. \end{aligned} \tag{2.18}$$

The probability that the node senses an idle channel is given by

$$\begin{aligned} \Pr\{\text{channel is sensed idle}\} &= \left[\Pr\{\text{node is empty}\} + \Pr\{\text{node is in sensing state}\}\right]^{n-1} \\ &= \left[(1-\rho) + \rho \sum_{i=0}^{K} \tilde{b}_{S_i}\right]^{n-1} \\ &\overset{\text{for large n}}{=} \exp\left\{-n\rho\left[1 - \sum_{i=0}^{K} \tilde{b}_{S_i}\right]\right\}. \end{aligned} \tag{2.19}$$

Substituting (2.15) into (2.18) and (2.19), then the probability of successful transmission p defined by (2.17) can be expressed as

$$\begin{aligned} p &= \exp\left\{-n\rho\sum_{i=0}^{K}\tilde{b}_{S_i}q^i\right\} \\ &= \exp\left\{-\frac{a\hat{\lambda}}{\alpha p}\left[\sum_{i=0}^{K-1}p(1-p)^i+(1-p)^K\right]\right\}. \\ &= \exp\left(-\frac{a\hat{\lambda}}{\alpha p}\right) \end{aligned} \tag{2.20}$$

After substituting the expression of α for the CA scheme from (2.9) into (2.20), we have

$$p = \exp\left\{-\frac{\hat{\lambda}(1+a-p)}{p}\right\}. \tag{2.21}$$

At equilibrium, the aggregate input rate $\hat{\lambda}$ should equals the network throughput $\hat{\lambda}_{out}$. Hence, it is easy to show from (2.21) that the network throughput in equilibrium is given by (2.16). □

Corollary 1. *For buffered non-persistent CSMA/CD with K-Exponential Backoff, the throughput in equilibrium is given by*

$$\hat{\lambda}_{out} = \frac{aGe^{-aG}}{aGe^{-aG}+a+\gamma\left(1-e^{-aG}-aGe^{-aG}\right)}. \tag{2.22}$$

Proof: Substituting the corresponding expression of α from (2.11) into (2.20), we have

$$p = \exp\left\{-\frac{\hat{\lambda}}{p}\left[-p\ln p+a+\gamma\left(1-p+p\ln p\right)\right]\right\}. \tag{2.23}$$

The network throughput of the CSMA/CD in (2.22) is immediately obtained. □

A generalized characteristic equation of throughput displayed below comes from (2.16) and (2.22), in which $x = 1$ or $= \gamma$ for the CA and CD schemes respectively.

$$\hat{\lambda}_{out} = \frac{aGe^{-aG}}{x + a - xe^{-aG} + (1-x)aGe^{-aG}} . \tag{2.24}$$

It should be noted that the throughput given by (2.16) also agrees with that obtained by Kleinrock and Tobagi in [6] and our channel model in section II. The network throughput can be derived either from the first-stage (input buffer) or second-stage (channel activities) queues of the two-stage queueing model mentioned in chapter 1. Whatever the value K is, the same network throughput can be derived from the queueing model of input buffer. The consistency between these approaches indicates that it is appropriate to adopt the independence assumption among input buffers. This is because the correlation among n input queues becomes weak in a multi-queue system when n is large [3].

Compare (2.20) and (2.23), it should be noted that the network throughput of the CD scheme is larger than that of the CA scheme due to the shorter collision time. Examples are shown in Fig. 2.6 which plots the network throughput $\hat{\lambda}_{out}$ against the probability of successful transmission p for the CA and CD (with $\gamma = 0.5$) schemes under different values of mini-slot a. The only difference between the CA and CD scheme is the distinct collision duration. This is the reason that the curves of the two schemes approach each other for the values of p near 1 because collisions seldom occur.

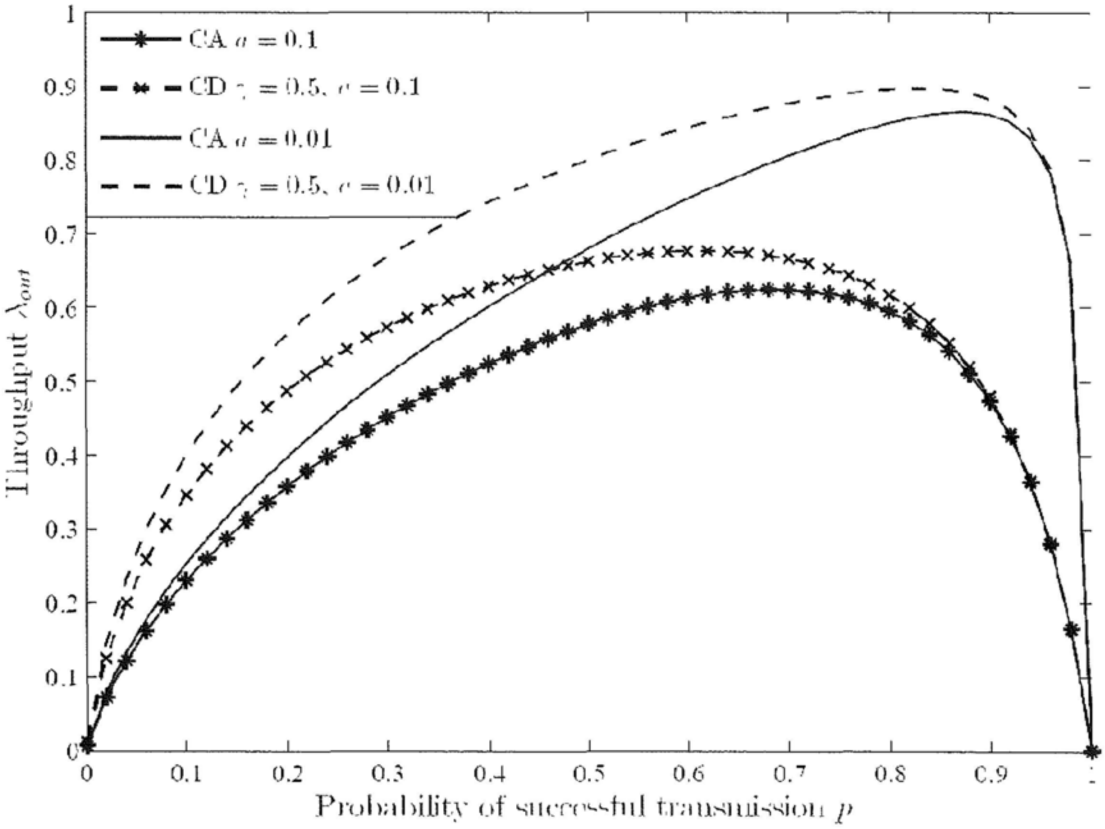


Fig. 2.6. Network throughput verses probability of successful transmission.

Furthermore, let random variables Suc^*, S_i^*, C_i^* and W_i^* be the service completion time of an HOL packet, starting from the states Suc, S_i, C_i, and W_i, respectively, until it is successfully transmitted. We assume, without loss of generality, that $M = a^{-1}$ and $X = \begin{cases} M & \text{for CA} \\ \gamma a^{-1} & \text{for CD} \end{cases}$ are integers. It is easy to show from the Markov chain of Fig. 2.5 that the generating functions $Suc(z)$, $S_i(z)$, $C_i(z)$, and $W_i(z)$ of these service completion times can be found by solving the following set of equations:

$$\begin{aligned}
&S_i(z) = E\left[z^{S_i^*}\right] \\
&= \alpha pq' zSuc(z) + \alpha\left(1-q'\right) zS_i(z) + \left(1-\alpha\right) zW_i(z) + \alpha q'\left(1-p\right) zC_i(z) \text{ for } i = 0,\ldots,K \\
&W_i(z) = E\left[z^{W_i^*}\right] = z^M S_i(z) \text{ for } i = 0,\ldots,K \\
&Suc(z) = E\left[z^{Suc^*}\right] = z^M \\
&C_i(z) = E\left[z^{C_i^*}\right] = z^X S_{i+1}(z) \\
&C_K(z) = E\left[z^{C_K^*}\right] = C_{K-1}(z)
\end{aligned} \qquad .(2.25)$$

Let X denote the service time of a HOL packet. Since the service of each HOL packet starts from state S_0, therefore the first and second moments of service time, $E[X]$ and $E[X^2]$, can be derived from the generating function $S_0(z)$ and the expressions of them are given in Appendix A. Note that the mean service time $E[X]$ derived from the generating functions $S_0(z)$ is same as that observed from the expression of offered load (2.15), that is the inverse of the time-average probability $\tilde{b}_{suc}$.

Based on the queueing model of the input buffer, we will investigate various stability issues of NP-CSMA concerning throughput and delay in the following sections.

2.4 Stable Regions

The aim of this section is to specify the stable throughput region as well as the bounded delay region with respect to the retransmission factor. Firstly, through the study of the relationship between the attempt rate G, input rate λ, and retransmission factor q, we can obtain an expression of the retransmission factor q. After that, we state the stable conditions in the coming sub-sections. With the expression of retransmission factor q and the stable conditions, we can work out the stable regions.

2.4.1 The expression of attempt rate G

Considering the NP-CSMA network with n nodes, a desired HOL packet can be transmitted successfully with probability e^{-aG}, when all other nodes are inactive in the mini-slot prior to its transmission. Suppose that there are a total of $n_b = \sum_{i=1}^{K} n_i$ HOL packets in a mini-slot, in which n_i packets are in the sensing state of phase i, for $i = 0,\ldots, K$. The following HOL packets may desire to transmit during the mini-slot of a sensing state:

1) An empty node may send a newly arrived packet with probability $a\lambda$;
2) An HOL packet in phase 0 will be transmitted immediately;
3) A backlogged HOL packet in phase i will be transmitted with probability q^i, for $i = 1,\ldots, K$.

Hence, the attempts rate per mini-slot, aG, can be expressed as

$$aG = (n - n_b)a\lambda + \sum_{i=0}^{K} n_i q^i \quad . \tag{2.26}$$

Let $\phi_i = \tilde{b}_{s_i} / \sum_{j=0}^{K} \tilde{b}_{s_j}$ be the probability that an HOL packet is in the sensing state of phase i, for $i = 0,1,\ldots,K$. The following attempt rate per mini-slot can be obtained from (2.13) and (2.26):

$$
\begin{aligned}
aG &= n(1-\rho)a\lambda + n\rho\sum_{l=0}^{K}\phi_l q^l \\
&= a\hat{\lambda}(1-\rho) + \frac{n\rho\sum_{l=0}^{K}\tilde{b}_{s_l} q^l}{\sum_{l=0}^{K}\tilde{b}_{s_l}} \\
&= a\hat{\lambda}(1-\rho) + n\rho\frac{(p+q-1)\left[1-(1-p)^{K+1}\right]}{pq\left[1-\left(\frac{1-p}{q}\right)^{K+1}\right]}.
\end{aligned}
\tag{2.27}
$$

For Exponential Backoff ($K \to \infty$), the expression of G can be further simplified into

$$aG = a\hat{\lambda}(1-\rho) + \frac{n\rho\left(e^{-aG}+q-1\right)}{qe^{-aG}}. \tag{2.28}$$

Equivalently, the retransmission factor q can be formulated as a function of attempt rate G as follows:

$$
\begin{aligned}
q &= h(G) \\
&= (1-p)\frac{A(p)+2\alpha\hat{\lambda}p\left(1-x+\frac{x}{p}\right)+\sqrt{A(p)^2+4\frac{a\alpha\hat{\lambda}^3p^3(1+a-\alpha)}{n}\left(1-x+\frac{x}{p}\right)}}{2A(p)+2\hat{\lambda}\alpha p\left(1-x+\frac{x}{p}\right)-2\frac{a\hat{\lambda}^2p^2}{n}(1+a-\alpha)}
\end{aligned}
\tag{2.29}
$$

where

$$A(p) = a\alpha p^2\left(\hat{\lambda}-G\right)+\hat{\lambda}p(1+a-\alpha)-a\alpha\hat{\lambda}^2p^2\left(1-x+xp^{-1}\right)n^{-1}. \tag{2.30}$$

In addition, the expressions of the retransmission factor q for both CA and CD schemes are easily obtained by substituting (2.5) into (2.29).

2.4.2 Stable throughput region

Intuitively, if the attempt rate G is too high, the probability of successful transmission $p = e^{-aG}$ will be too small; if it is too low, then the channel may not be fully utilized. Similar to the Aloha system discussed in [4] that the maximum throughput $\hat{\lambda}_{\max}$ can be expressed in terms of the Lambert W function $W_0(z)$, which was first considered by J. Lambert in 1758, and later studied by L. Euler [66]. Taking the differentiation of the throughput equation (2.16), the maximum throughput $\hat{\lambda}_{\max}$ for the CA scheme can be derived from the corresponding attempt rate G^* and expressed as below,

$$\hat{\lambda}_{\max} = \frac{\left[W_0\left(-\frac{1}{(1+a)e}\right)+1\right]\exp\left[-W_0\left(-\frac{1}{(1+a)e}\right)-1\right]}{1+a-\exp\left[-W_0\left(-\frac{1}{(1+a)e}\right)-1\right]}. \tag{2.31}$$

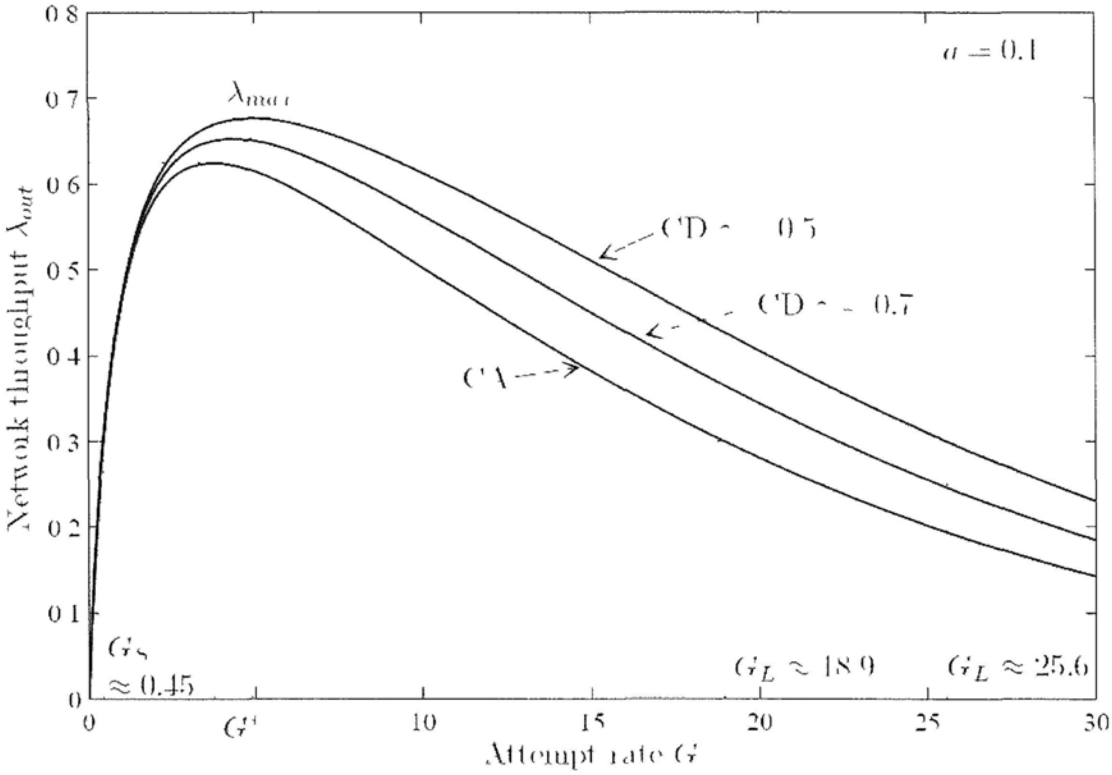


Fig 2 7 Network throughput verse attempt rate with $a = 0\ 1$

As plotted in Fig. 2.7 and Fig. 2.8, for any attempt rate $G < G^*$, the network throughput increases with attempt rate G; otherwise, the throughput decreases with G. Therefore, for any throughput smaller than the maximum throughput, the characteristic equation (2.24) has two roots: the smaller $G_S(\hat{\lambda})$ and larger $G_L(\hat{\lambda})$ roots as given below,

$$G_S(\hat{\lambda}) = -\frac{1}{a}\left[W_0\left(-L(a+x)e^{-xl}\right) + xL\right] \tag{2.32}$$

and

$$G_l(\hat{\lambda}) = -\frac{1}{a}\left[W_{-1}\left(-L(a+x)e^{-xl}\right) + xL\right], \tag{2.33}$$

where $L = \dfrac{\hat{\lambda}}{1-\hat{\lambda}+x\hat{\lambda}}$; $W_0(z)$ and $W_{-1}(z)$ are the Lambert W function defined in [66].

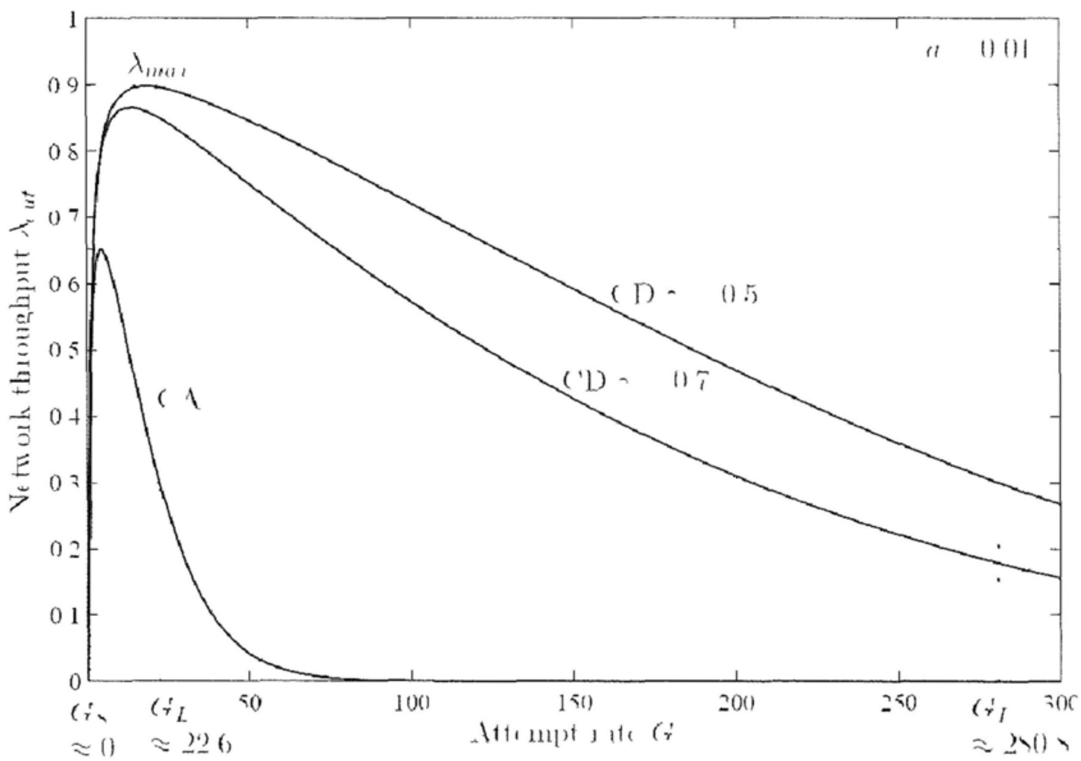


Fig. 2.8 Network throughput verse attempt rate with $a = 0.01$

Considering the tradeoff between G and p, we know the attempt rate G should be bounded in the range between $G_S(\hat{\lambda})$ and $G_l(\hat{\lambda})$ to ensure a stable network throughput $\hat{\lambda}_{out} = \hat{\lambda}$. For example shown in Fig. 2.7, the network has a stable throughput of $\hat{\lambda}_{out} = \hat{\lambda} = 0.3$ when G is within the range [$G_S \approx 0.45$, $G_L \approx 18.9$] and [$G_S \approx 0.45$, $G_L \approx 25.6$] for CA and CD with $\gamma = 0.5$ schemes. Specifically, a necessary condition of stable throughput of the entire system can be stated as follows:

Stable Throughput Condition:

STC. For any aggregate input rate $\hat{\lambda} < \hat{\lambda}_{\max}$, the attempt rate G should satisfy

$$G_S(\hat{\lambda}) \le G \le G_L(\hat{\lambda}) . \tag{2.34}$$

In general, the attempt rate G is an implicit function of the retransmission factor q associated with the underlying scheduling algorithm as shown in (2.27). For retransmission factor q in the range $0 < q < 1$, this function is monotonically increasing with respect to the attempt rate G. This functional relationship, together with the inequality (2.34), determines a stable region of q. For the NP-CSMA protocol with exponential backoff scheduling algorithm, it is easy to show that the function $q = h(G)$ monotonically increases with respect to G. Thus, a region of the retransmission factor q, called the stable throughput region R_T, can be specified by the combination of (2.34) and (2.27) as follows:

$$q \in R_T = \left[h(G_S), h(G_L)\right]. \tag{2.35}$$

The network throughput is defined by $\hat{\lambda}_{out} = \min\{n\tilde{b}_{suc}, \hat{\lambda}\}$ and the stable throughput condition **STC** ensures that $\hat{\lambda}_{out} = \hat{\lambda}$. It follows that the **STC** implies $n\tilde{b}_{suc} \ge \hat{\lambda} = n\hat{\lambda}$, that is, the offered load $\rho \le 1$. On the other hand, it can be shown from that the offered load ρ is a monotonically increasing function of the retransmission factor q if the attempt rate G is bounded in the range $[G_S, G_L]$, or equivalently $q \in R_T = \left[h(G_S), h(G_L)\right]$. In particular, the attempt rate G will reach G_L when the offered load $\rho = 1$. Therefore, if the retransmission factor q is chosen from the region $[h(G_S), h(G_L))$, the offered load of Geo/G/1 queue for each input buffer should be strictly less than 1, which is the stable condition of any queueing system that the arrival rate is less than the service rate.

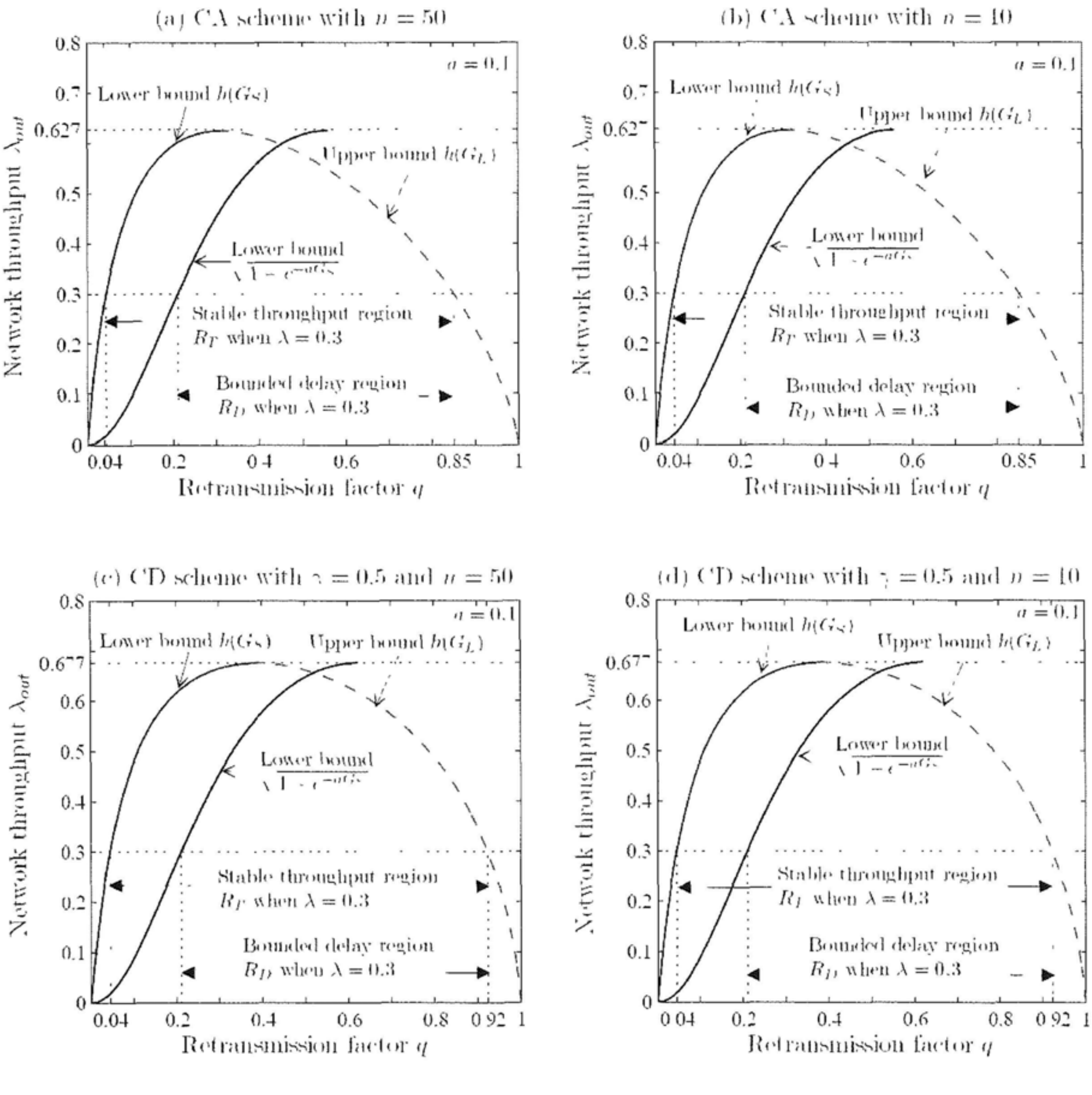


Fig. 2.9. Stable throughput and bounded delay regions of Exponential Backoff.

Fig. 2.9 shows stable regions of Exponential Backoff with $a = 0.1$ for the CSMA/CA and CSMA/CD ($\gamma = 0.5$) when $n = 10$ and 50. Areas under the lower bound $h(G_S)$ and upper bound $h(G_I)$ are the stable throughput regions. In Fig. 2.9 (a) and (b) with given aggregate input rate $\hat{\lambda} = 0.3$, the network throughput can be stabilized when the retransmission factor q is selected within the range [0.04, 0.85]. On the other hand, [0.04, 0.92] is the stable region for the CSMA/CD when $\hat{\lambda} = 0.3$. It can be observed from this figure that the stable throughput region of the CSMA protocol less depends on the population size n, but more depends on the types of collision schemes and the value of γ. This can be explained by the expression (2.29) that values of terms containing variable n are very small.

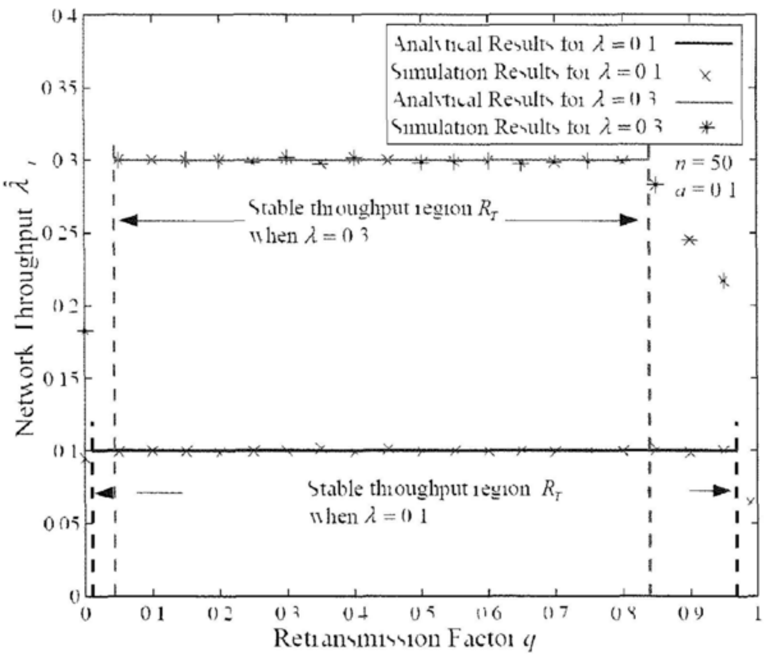


Fig. 2.10 Simulation results of stable throughput region when $n = 50$

Our analysis agrees with simulations result as shown is shown in Fig. 2.10 with n = 50. For aggregate input rates $\hat{\lambda}=0.1$ and $\hat{\lambda}=0.3$, simulation results match perfectly with the analytical one when the retransmission factor q is within the stable throughput region. Outside the stable regions, network throughputs drop gradually.

Furthermore, the dependence of collision duration γ is illustrated in Fig. 2.11 with simulation results for aggregate input rate $\hat{\lambda}=0.3$ and n = 10. The smaller the value γ we choose, the larger the stable throughput region we have. Our stability analysis is confirmed by the simulation results which depict that a stable throughput can always be achieved if the retransmission factor q is properly chosen from the stable throughput region R_T. Fig. 2.10 and Fig. 2.11 demonstrate that the network performance is predictable by our analysis if it is operated within the stable region.

For an extreme case that the population size n goes to infinity, the expression of retransmission factor q in (2.29) can be further reduced to

$$q=1-p. \tag{2.36}$$

Hence, from (2.35), we immediately obtain the following non-empty stable throughput region for infinity population:

$$R_I=[1-e^{-aG_S},1-e^{-aG_L}], \tag{2.37}$$

which coincides with the existing results on Exponential backoff reported in [58] –[60] that the network throughput can be non-zero even when the number of nodes goes to infinity. The stable throughput region in (2.37) is true for both CA and CD schemes, but with different expressions of G_S and G_L.

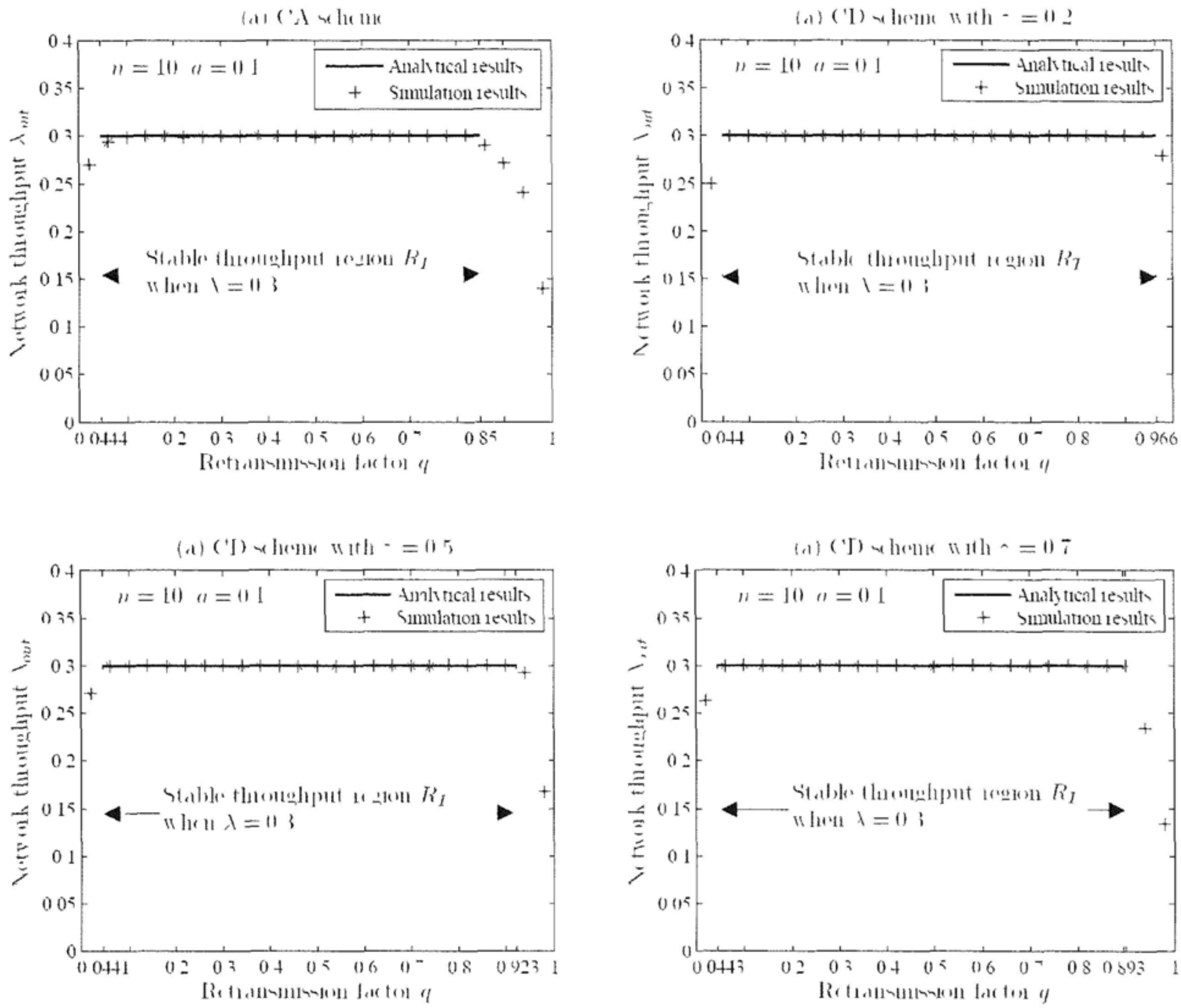


Fig. 2.11. Simulation results of stable throughput region when $n = 10$.

2.4.3 Bounded delay region

As mentioned before, the stable throughput condition (2.34) is not sufficient to guarantee a bounded mean delay of packets queued in each input buffer. To complete the analysis of stable regions, we deduce the following additional constraint from our queueing model of the input buffer.

Bounded Delay Condition:

BDC. The Pollaczek-Khinchin formula for mean delay $E[T]$ of Geo/G/1 queue [68]

$$E[T] = E[X] + \lambda \frac{E[X^2] - E[X]}{2(1 - \lambda E[X])} \tag{2.38}$$

requires bounded second moment of service time $0 < E[X^2] < \infty$.

The condition **BDC** is more restrictive than the condition **STC.** A region of the factor q, denoted as R_D, that guarantees bounded mean delay can be determined by the second moment of service time $E[X^2]$. In fact, the bounded delay region is a subset of the stable throughput region.

The second moment of service time of the Exponential Backoff scheme is given by (A2) in Appendix A as follows:

$$E[X^2] = B(p,q) \lim_{K \to \infty} \left(\frac{1-p}{q^2} \right)^{K-1} + C(p,q), \tag{2.39}$$

where $B(p,q)$ and $C(p,q)$ are two polynomials given by (A3) and(A4), respectively, in Appendix A. It can be observed from the second moment of service time that the retransmission factor q should satisfy the following condition to ensure a bounded second moment of service time $E[X^2]$:

$$q^2 > 1 - p, \tag{2.40}$$

or equivalently,

$$q > \sqrt{1 - e^{-aG}}\,. \tag{2.41}$$

Since bounded delay implies stable throughput, thus the bounded delay region of Exponential Backoff can be specified by the combination of condition (2.41) and the stable throughput region R_T excluding the point of upper bound, and given as follows:

$$R_D = \left[\sqrt{1 - e^{-aG_S}}, h(G_L)\right). \tag{2.42}$$

Suppose the bounded delay condition is satisfied, the mean access delay $E[X]$ and queueing delay $E[T]$ are given:

$$E[X] = \frac{aq(M + 1 - M\alpha)}{\alpha(p + q - 1)} + \frac{aX + ap(M - X)}{p} \tag{2.43}$$

and

$$\begin{aligned} E[T] = E[X] - \frac{\lambda E[X]}{2(1 - \lambda E[X])} + \frac{a^2\lambda}{2(1 - \lambda E[X])}[\frac{2q(1-p)X(M+1-\alpha M)}{\alpha(p+q-1)^2} - \frac{\alpha q M}{\alpha(p+q-1)} \\ + \frac{p(M+1)(M-2) + (1-p)(X+1)(X-2) + 2}{p} + \frac{2(1-p)X(pM + X - pX)}{p^2} \\ + \frac{2q^3(M+1-\alpha M)^2}{\alpha^2(p+q^2-1)(p+q-1)} + \frac{q(M+1-\alpha M)(3pM + 2X - 2pX - 2p)}{\alpha p(p+q-1)}] \end{aligned} \tag{2.44}$$

As shown in Fig. 2.9, the shaded bounded delay region R_D is a subset of the stable throughput region R_T. Outside this bounded delay region when $q \in R_I \setminus R_D$, the system may still have a stable throughput but the mean delay will quickly jump up to an unacceptable level. This phenomenon is as demonstrated in Fig. 2.12 and Fig. 2.13, which exhibit the queueing delay of the packet versus retransmission factor q for a fixed aggregate input rate $\hat{\lambda} = 0.3$ when $n = 50$ and 10 respectively.

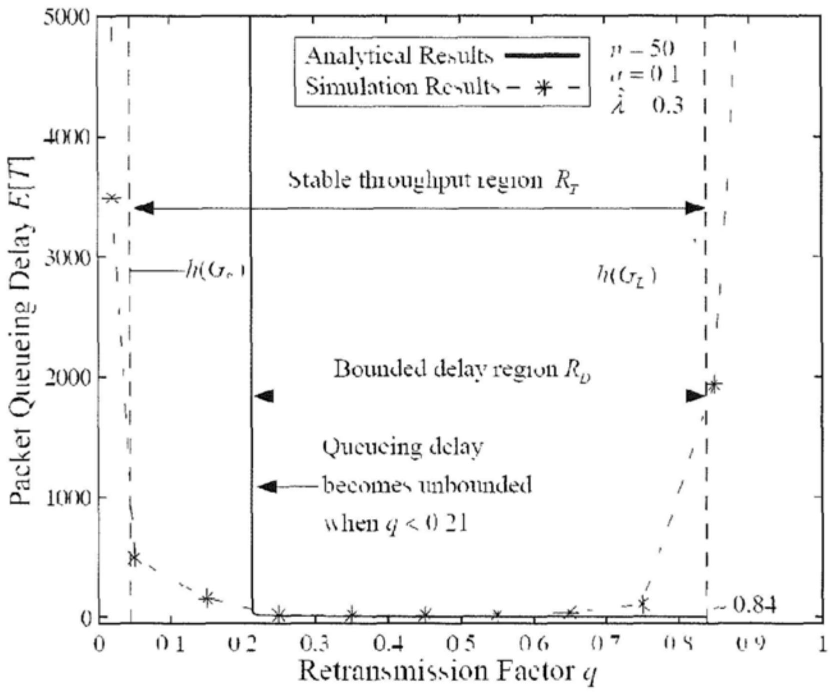


Fig. 2.12. Mean queueing delay verses retransmission factor q when $n = 50$.

From Fig. 2.12, it can be observed that if $0.21 < q < 0.84$, within the region of bounded delay, the mean queueing delay of the packet in the network can always be kept small. However, if the retransmission factor q is outside the bounded delay region, the mean delay of the packet would become considerably larger; especially when q is greater than the upper bound of the delay region, the queueing delay increases rapidly. Moreover, the simulation results also confirm that the analytical lower bound of the region R_D given in (2.42) is much more conservative than the upper bound.

Other simulation results with smaller population size $n = 10$ are shown in Fig. 2.13. All these results show that the queueing delay $E[T]$ keeps small when the retransmission factor q is selected from the bounded delay region. The CD scheme with smaller value of γ has larger bounded delay region.

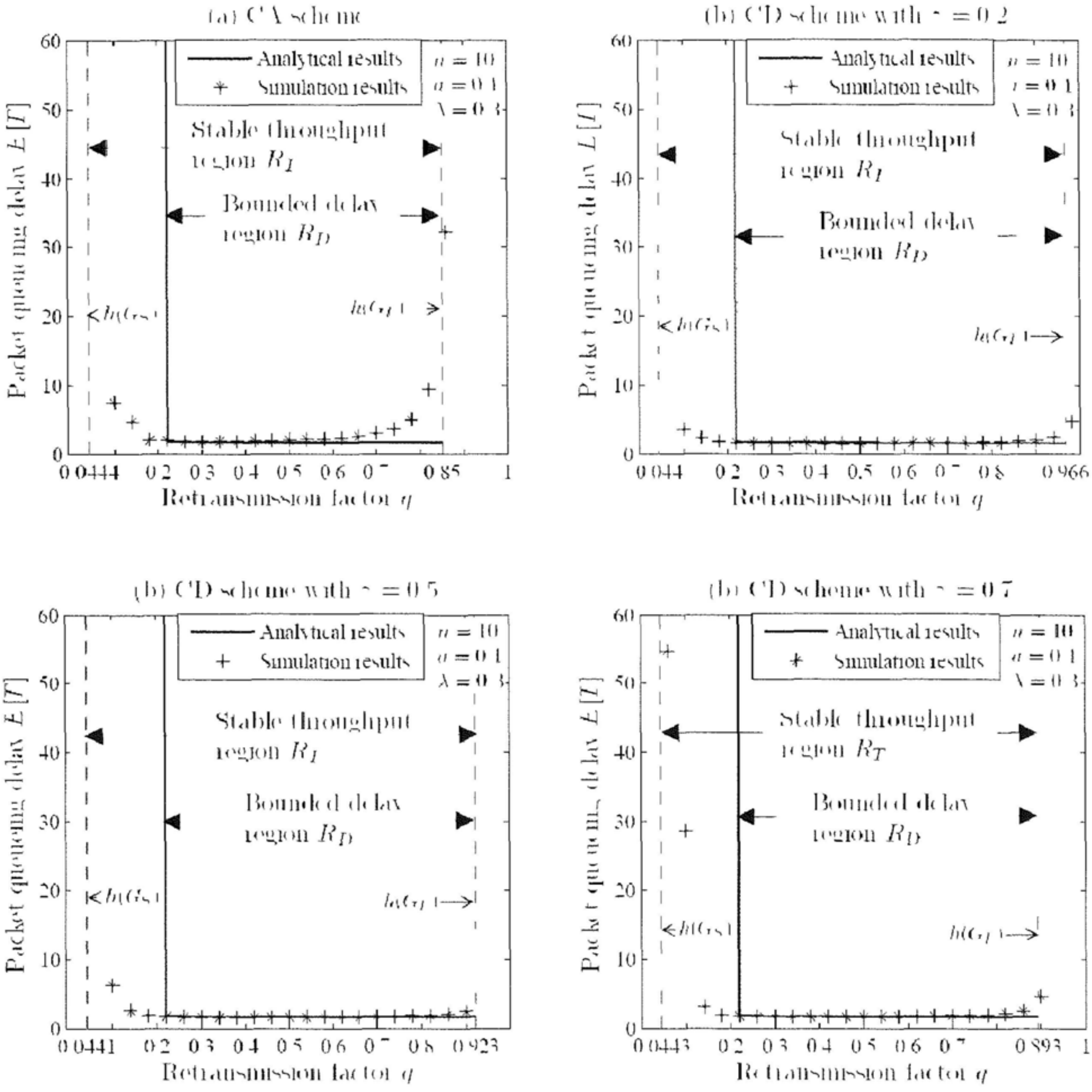


Fig 2 13 Mean queueing delay verses retransmission factor q when $n = 10$

Since the bounded delay region R_D specified by (2.42) is a subset of the stable throughput region R_T, the maximum throughput $\hat{\lambda}_{\max}$ may not be achievable if $q \in R_D$. As shown in Fig. 2.9, the network throughput increases with the lower bound of the bounded delay region ($\sqrt{1-e^{-aG_S}}$), and decreases with the upper bound $h(G_L)$. Therefore, the maximum throughput within the bounded delay region R_D can be achieved when the lower bound and upper bound of the retransmission factor q given by (30) are equal. That is,

$$\sqrt{1-e^{-aG_S}} = h(G_L). \tag{2.45}$$

In the case of infinite number of nodes, the above equation becomes

$$\sqrt{1-e^{-aG_S}} = 1-e^{-aG_I}, \tag{2.46}$$

or equivalently,

$$G_S = -\frac{1}{a}\ln\left(e^{-aG_I}\left(2-e^{-aG_I}\right)\right). \tag{2.47}$$

In Fig. 2.14, we plot the maximum throughput within the stable throughput region and bounded delay region for the NP-CSMA protocols. For $n \to \infty$ and $0 < a < 1$, the maximum throughput $\hat{\lambda}_{\max}$ is always larger than that obtained from the bounded delay region. In other words, the absolute maximum throughput $\hat{\lambda}_{\max}$ in the stable throughput region cannot be achieved with bounded mean delay guarantee.

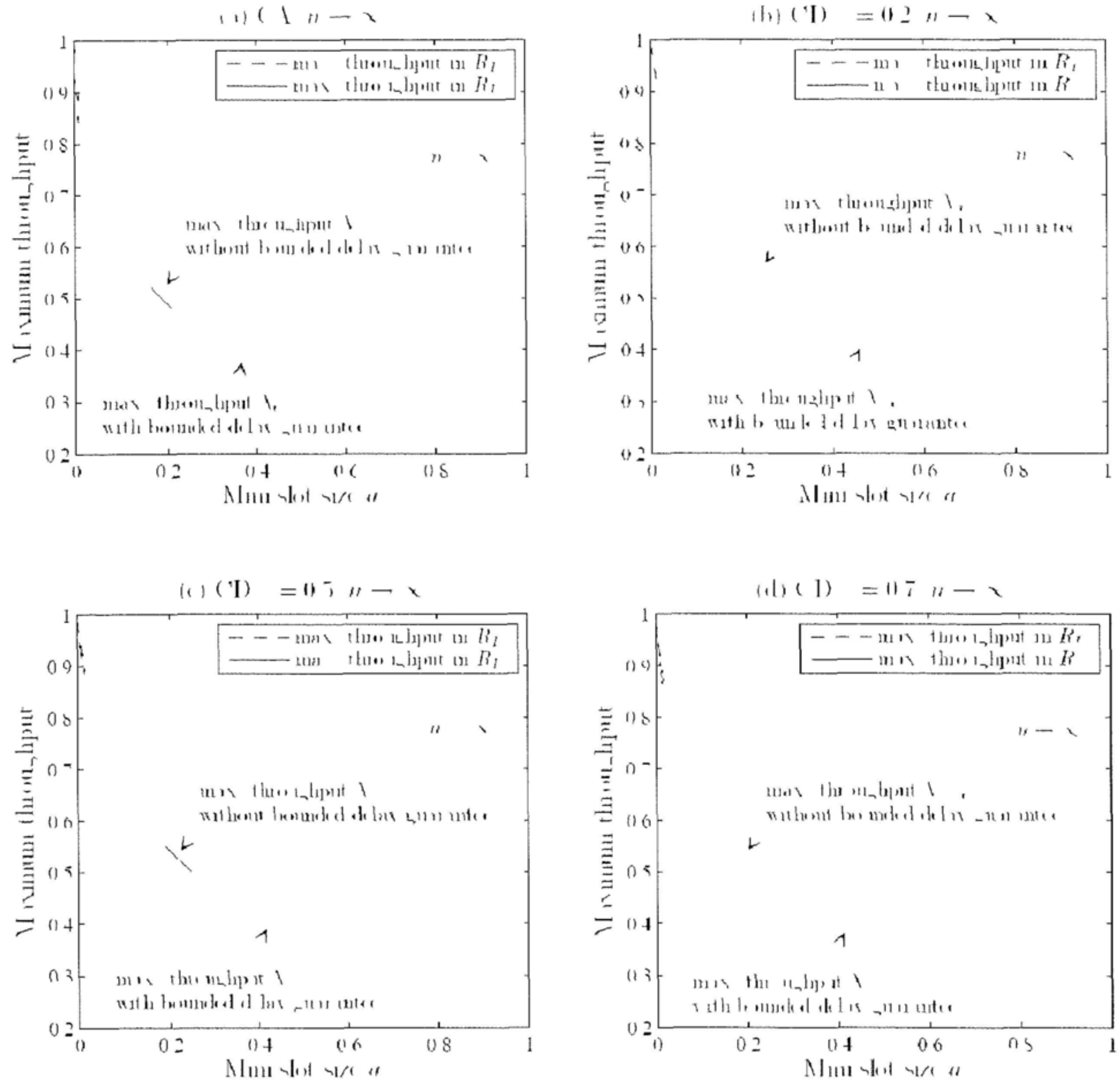


Fig 2 14 Maximum throughputs obtained from the stable throughput and bounded delay regions

In a nutshell, the Exponential Backoff has two types of stable regions as shown in Fig. 2.9. The first type is the bounded delay region R_D, in which the network throughput is stable and the mean delay is bounded. The second type is the region $R_T \setminus R_D$ that only guarantees stable throughput but with unacceptable large mean delay. In the Exponential Backoff, it is possible that predominating backlogged packets are pushed to deep phases with very low retransmission probabilities when the network becomes congested. If a node tries to send its HOL packet, the successful probability will be very high. Once the backlogged HOL packet is cleared, then the channel may be "captured" by subsequent packets in the input buffer, which are all in phase 0, until the queue is cleared. During this time, it appears that the network throughput is still stable, but the variance of the service time of each individual packet can be infinitely large due to this unfairness of services caused by the capture effect mentioned in [67]. Nevertheless, the Exponential Backoff is robust and can scale to a larger population.

2.5 Conclusion

A multi-queue-single-server model is proposed in this paper to explore the stability and delay issues of multiple access networks with NP-CSMA/CA and NP-CSMA/CD protocols. We studied the throughput of the entire system and the performance of each individual input buffer. The Exponential Backoff scheduling algorithm is considered in the analysis of the service time of HOL packets. Based on this model, we formulated conditions on stable throughput and bounded mean delay according to the basic principle of queueing theory. Stable throughput region and bounded delay region of the retransmission factor are established for the Exponential Backoff scheme. Our method is mainly based on the construction of Markov chains of the transmission channel and the service time of HOL packets; they can be easily generalized to investigate other MAC protocols, such as IEEE 802.11, in the future.

Chapter 3

1-persistent Carrier Sense Multiple Access Protocols

Chapter 2 gives a complete queueing analysis of non-persistent carrier sense multiple access protocols (NP-CSMA). In this chapter, we generalize the queueing model to study the performance of the 1-persistent CSMA protocol (1P-CSMA). The queueing model in this chapter also considers the backoff scheduling algorithm and admissible traffic characteristics.

3.1 Basic principle of the 1-persistent CSMA protocol

The 1P-CSMA protocol is a special case of the p-persistent CSMA protocol with p = 1. The p-persistent CSMA protocols try to maximize the channel utilization by continuously monitoring the transmission medium. An active node with a packet ready to transmit constantly checks the channel status. If the channel is sensed idle, the node transmits the packet with a probability *p*. Otherwise, the node persistently monitors the channel until it becomes idle. At this time, it transmits the packet with the same probability *p*. This process repeats until the packet is dispatched. Collisions may still occur when two or more packets are simultaneously transmitted. In other words, the node sends packet immediately if the channel is idle. There are several assumptions we made in this chapter:

1. The system is slotted into mini-slots in which activities can be taken only at the beginning of mini-slots,
2. The network is well synchronized.
3. Packet collapsed only due to simultaneous transmission.

The 1P-CSMA protocol also contains two different schemes, namely the collision detection (CD) and collision avoidance (CA). The CSMA/CA protocol is widely used in wireless communications. Source nodes notice collision if there is no acknowledgement (ACK) message. For sake of simplicity, we assume the time for receiving an ACK message is negligible, that is, the time required for successful transmission is the same as that for unsuccessful transmission. For the collision detection scheme, nodes can detect collision during packets transmission and then stop the transmission. In this way, the collision time is less than the successful transmission time. The collision detection scheme can be applied in wired network which allow users to keep monitoring their transmission, such as the Ethernet.

3.2 Channel Model

In a slotted 1P-CSMA network, suppose the ratio of propagation delay to packet transmission time is a; the time axis consists of mini-slots with size a. Packets can be sent only at the beginning of a timeslot, and each packet is identical and takes one slot time for transmission.

For the CA scheme, the duration of a transmission period is $1 + a$ timeslots because no packets will attempt to transmit in the mini-slot immediately after a transmission. The successful transmission time of the CD scheme is the same as the counterpart; while the collision time of the CD scheme is shorter, because transmissions can be stopped during the collision, as shown in Fig. 3.1. Let γ be the time required to detect and abort the collision (γ is less than 1), then the duration of the corresponding transmission period is $\gamma + a$. For convenience, we define a variable x with the following expression:

$$x = \begin{cases} 1 & \text{for CA} \\ \gamma & \text{for CD} \end{cases}. \tag{3.1}$$

A busy period is defined as a series of transmission periods, and a pure idle period is the time in which the channel is idle and no packet presents awaiting transmission. Hence, the time axis can be considered as a sequence of alternating busy periods and pure idle periods, as displayed in Fig. 3.2.

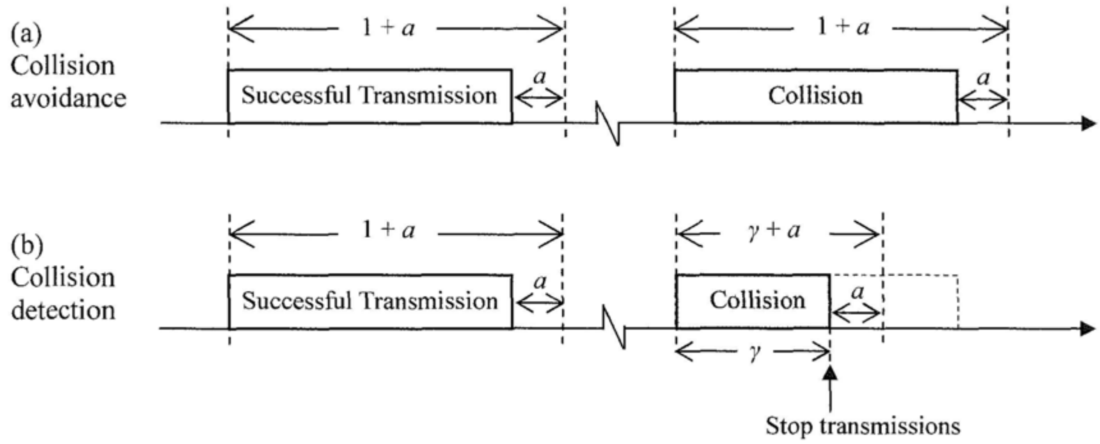


Fig. 3.1. Transmission periods of the 1P-CSMA with (a) collision avoidance; (b) collision detection.

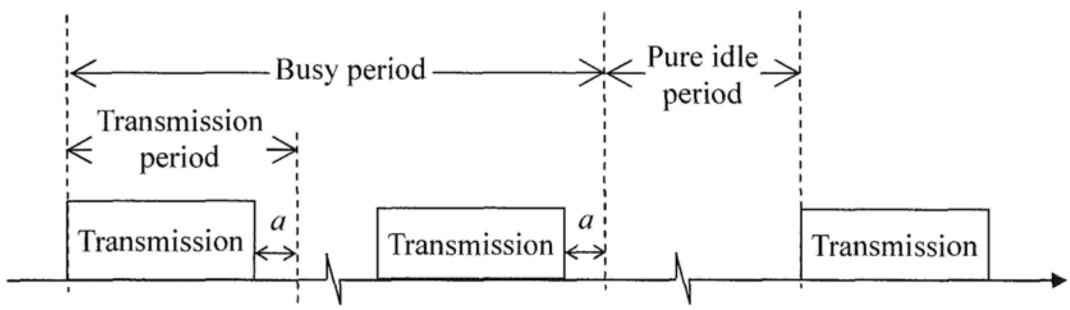
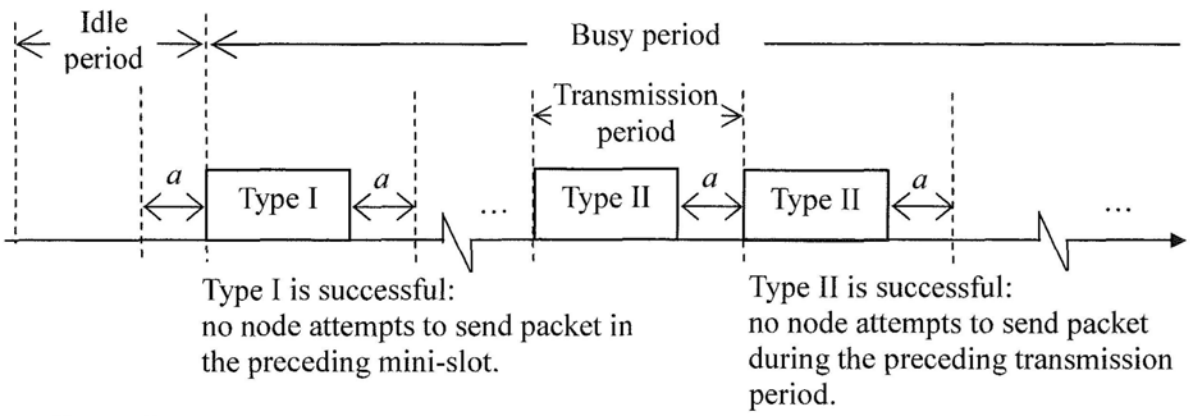


Fig. 3.2. Alternative renewal process of the 1P-CSMA protocol.

The 1P-CSMA protocol contains two types of transmissions with different probabilities of success. As shown in Fig. 3.3, the leading transmission period in each busy period is called Type I transmission period, in which the packet is sent successfully if and only if no one is scheduled to access the channel in the previous mini-slot. All subsequent transmission periods are defined as Type II transmission. Due to the persistent property of 1P-CSMA, packets scheduled to access the channel in any transmission period accumulate at the end of the transmission period; hence, the second type of transmission is successful if and only if no one attempts to access the channel during the previous transmission period.



Remarks:

CA: duration of transmission period $= 1 + a$

CD: duration of transmission period $= x + 1 = \begin{cases} 1 + a & \text{for successful transmission} \\ \gamma + a & \text{for collision} \end{cases}$

Fig. 3.3. Properties of the Type I and Type II transmissions.

The length of a pure idle period is a random variable with geometric distribution. In each transmission period, a packet is either successfully transmitted or collided. Therefore, the channel has three fundamental states: pure idle (Idle), successful transmission (Suc), and collision (Col). The successful transmission state should be divided into two sub-states Suc1 and Suc2, corresponding to the success of Type I and Type II transmissions respectively. The state transitions of the channel can be described by the Markov chain shown in Fig. 3.4.

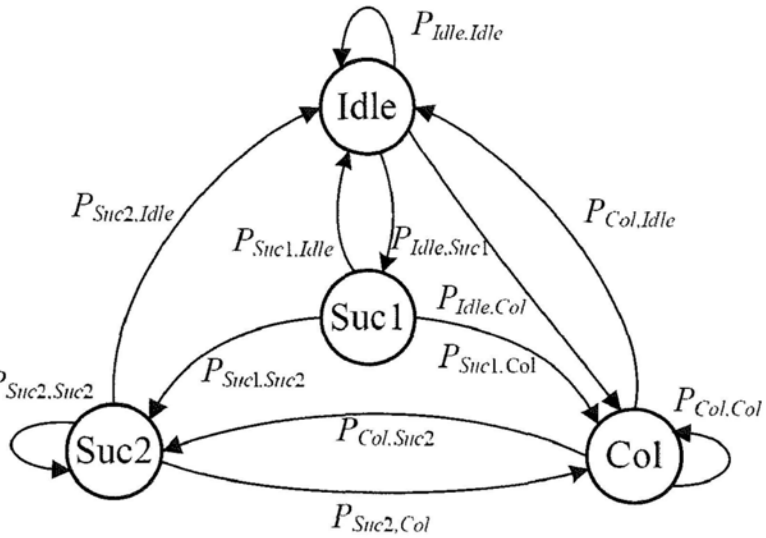


Fig. 3.4. Markov Chain of 1P-CSMA channel.

The limiting probabilities π_{Idle}, π_{Suc}, and π_{Col} of the Markov chain satisfy the following set of equations:

$$\begin{aligned}
\left(1-P_{Idle,Idle}\right)\cdot\pi_{Idle} &= P_{Suc1,Idle}\cdot\pi_{Suc1}+P_{Suc2,Idle}\cdot\pi_{Suc2}+P_{Col,Idle}\cdot\pi_{Col} \\
\pi_{Suc1} &= P_{Idle,Suc1}\cdot\pi_{Idle} \\
\left(1-P_{Suc2,Suc2}\right)\cdot\pi_{Suc2} &= P_{Suc1,Suc2}\cdot\pi_{Suc1}+P_{Col,Suc2}\cdot\pi_{Col} \\
\left(1-P_{Col,Col}\right)\cdot\pi_{Col} &= P_{Idle,Col}\cdot\pi_{Idle}+P_{Suc1,Col}\cdot\pi_{Suc1}+P_{Suc2,Col}\cdot\pi_{Suc2}
\end{aligned}\quad ; \qquad (3.2)$$

while sojourn times of Idle, Suc1, Suc2, and Col states are given below,

$$\begin{aligned}
&t_{Idle} = a \\
&t_{Suc1} = 1 + a \\
&t_{Suc2} = 1 + a \\
&t_{Col} = x + a = \begin{cases} 1 + a & \text{for CD} \\ \gamma + a & \text{for CD} \end{cases}
\end{aligned} \quad . \tag{3.3}$$

The aggregate attempt, which is defined as the collection of all HOL packets trying to access the channel, is generated by all fresh and re-scheduled HOL packets. The corresponding attempt rate G is Poisson distributed, therefore the probability that no attempts is generated in a mini-slot is e^{-aG}. Similarly, the probabilities that no node attempts to access the channel in $(1+a)$ and $(\gamma+a)$ slots are $e^{-(1+a)G}$ and $e^{-(\gamma+a)G}$ respectively. Therefore, the transition probability from the four states of the Markov chain in to the Idle state is given as below,

$$\begin{aligned}
&P_{Idle\ Idle} = e^{-aG} \\
&P_{Suc1\ Idle} = P_{Suc2\ Idle} = e^{-(1+a)G} \, . \\
&P_{Col\ Idle} = e^{-(\gamma+a)G}
\end{aligned} \tag{3.4}$$

In the channel's point of view, the Type I transmission is successful only if there is exactly one attempting packet in a mini-slot with the probability aGe^{-aG}. Hence,

$$P_{Idle\ Suc1} = aGe^{-aG} \, . \tag{3.5}$$

All packets scheduled to access the channel in any transmission period are accumulated at the end of previous period and then attempt to access the channel at the same time. For the CA scheme, the transition probability to the Suc2 state is equivalent to the case that there is exactly one scheduled packet within $(1+a)$ slots, i.e., $(1+a)e^{-(1+a)G}$. For the CD scheme, the transition probabilities from the Col to Suc2 states are different due to the shorter duration of the state Col, i.e., $(\gamma+a)e^{-(\gamma+a)G}$. The transition probabilities to the Suc2 state can be generalized as follows:

$$\begin{aligned} &P_{Suc1,Suc2} = P_{Suc2,Suc2} = (1+a)Ge^{-(1+a)G} \\ &P_{Col,Suc2} = (x+a)Ge^{-(x+a)G} \end{aligned} . \quad (3.6)$$

With (3.4), (3.5), and (3.6), the remaining transition probabilities of the Markov chain are obvious as below,

$$\begin{aligned} &P_{Idle,Col} = 1 - e^{-aG} - aGe^{-aG} \\ &P_{Suc1,Col} = P_{Suc2,Col} = 1 - e^{-(1+a)G} - (1+a)Ge^{-(1+a)G} \\ &P_{Col,Col} = 1 - e^{-(x+a)G} - (x+a)Ge^{-(x+a)G} \end{aligned} . \quad (3.7)$$

We obtain the following limiting probabilities from (3.2), (3.4), (3.5), and (3.6) in a straightforward manner:

$$\pi_{Idle} = \frac{e^{-(x+a)G}\left(1 - Ge^{-(1+a)G} + Gxe^{-(1+a)G}\right)}{\left(1 - e^{-aG} + e^{-(x+a)G}\right)\left(1 - Ge^{-(1+a)G}\right) + \left(a + x - xe^{-aG} + xe^{-(1+a)G}\right)Ge^{-(x+a)G} - aGe^{-(1+a)G}}$$

$$\pi_{Suc1} = \frac{aGe^{-aG}e^{-(x+a)G}\left(1 - Ge^{-(1+a)G} + Gxe^{-(1+a)G}\right)}{\left(1 - e^{-aG} + e^{-(x+a)G}\right)\left(1 - Ge^{-(1+a)G}\right) + \left(a + x - xe^{-aG} + xe^{-(1+a)G}\right)Ge^{-(x+a)G} - aGe^{-(1+a)G}}$$

$$\pi_{Suc2} = \frac{Ge^{-(x+a)G}\left[(a+x)\left(1 - e^{-aG}\right) + (1-x)aGe^{-aG}e^{-(1+a)G}\right]}{\left(1 - e^{-aG} + e^{-(x+a)G}\right)\left(1 - Ge^{-(1+a)G}\right) + \left(a + x - xe^{-aG} + xe^{-(1+a)G}\right)Ge^{-(x+a)G} - aGe^{-(1+a)G}}$$

$$\pi_{Col} = \frac{1 - e^{-aG} - \left(1 + a - e^{-aG}\right)Ge^{-(1+a)G}}{\left(1 - e^{-aG} + e^{-(x+a)G}\right)\left(1 - Ge^{-(1+a)G}\right) + \left(a + x - xe^{-aG} + xe^{-(1+a)G}\right)Ge^{-(x+a)G} - aGe^{-(1+a)G}}.$$

$$(3.8)$$

With the sojourn time given in (3.3), the time-average probabilities of each state are:

$$\tilde{\pi}_{Idle} = \frac{ae^{-(x+a)G}\left(1 - Ge^{-(1+a)G} + Gxe^{-(1+a)G}\right)}{D_{CH}}$$

$$\tilde{\pi}_{Suc1} = \frac{(1+a)aGe^{-aG}e^{-(x+a)G}\left(1 - Ge^{-(1+a)G} + Gxe^{-(1+a)G}\right)}{D_{CH}}$$

$$\tilde{\pi}_{Suc2} = \frac{(1+a)Ge^{-(x+a)G}\left[(a+x)\left(1 - e^{-aG}\right) + (1-x)aGe^{-aG}e^{-(1+a)G}\right]}{D_{CH}}$$

$$\tilde{\pi}_{Col} = \frac{(1+x)\left[1 - e^{-aG} - \left(1 + a - e^{-aG}\right)Ge^{-(1+a)G}\right]}{D_{CH}} \qquad (3.9)$$

where

$$D_{CH} = (1+a)Ge^{-(x+a)G}\left(x + a - xe^{-aG}\right) + (x+a)\left[1 - e^{-aG} - \left(1 + a - e^{-aG}\right)Ge^{-(1+a)G}\right]$$
$$+ae^{-(x+a)G}\left(1 - Ge^{-(1+a)G} + Gxe^{-(1+a)G}\right).$$

Since the successful transmission of a packet only takes $\frac{1}{1+a}$ of the time in a transmission period, the following network throughput is defined by the fraction of the time that the channel is productive:

$$\hat{\lambda}_{out} = \frac{\tilde{\pi}_{Suc1} + \tilde{\pi}_{Suc2}}{(1+a)}. \tag{3.10}$$

From (3.9) and (3.10), we have the expression of network throughput

$$\hat{\lambda}_{out} = \frac{\tilde{\pi}_{Suc1} + \tilde{\pi}_{Suc2}}{(1+a)} = \frac{Ge^{-(1+a)G}\left(x + a - xe^{-aG}\right)}{D_{CH}}, \tag{3.11}$$

which is consistent with the results in [6] and [8]. The familiar characteristic equation of throughput is immediately given by putting $x = 1$, that is,

$$\hat{\lambda}_{out} = \frac{Ge^{-(1+a)G}\left(1 + a - e^{-aG}\right)}{(1+a)\left(1-e^{-aG}\right) + ae^{-(1+a)G}}, \tag{3.12}$$

The characteristic equation of network throughput can be derived from the channel model, but it is not sophisticated enough to give a comprehensive analysis of the 1P-CSMA protocol due to the ignorance of the re-scheduling of collided HOL packets. In the next section, we describe a more detailed queueing model of the input buffer which considers the backoff scheduling algorithm as well as the input traffic characteristic.

3.3 Queueing Model of Input Buffer

The behavior of each HOL packet in the 1P-CSMA protocol with the *K*-Exponential Backoff algorithm is described by a simple flow chart shown in Fig. 3.5. A fresh HOL packet is initially in phase 0 and is sent only when an idle channel is detected. If the channel is busy, it waits until the channel becomes idle and then transmits the packet immediately. If the transmission is successful, a new fresh HOL packet starts from phase 0. Otherwise, the packet is in backoff mode in which the random waiting time is determined by the current backoff phase.

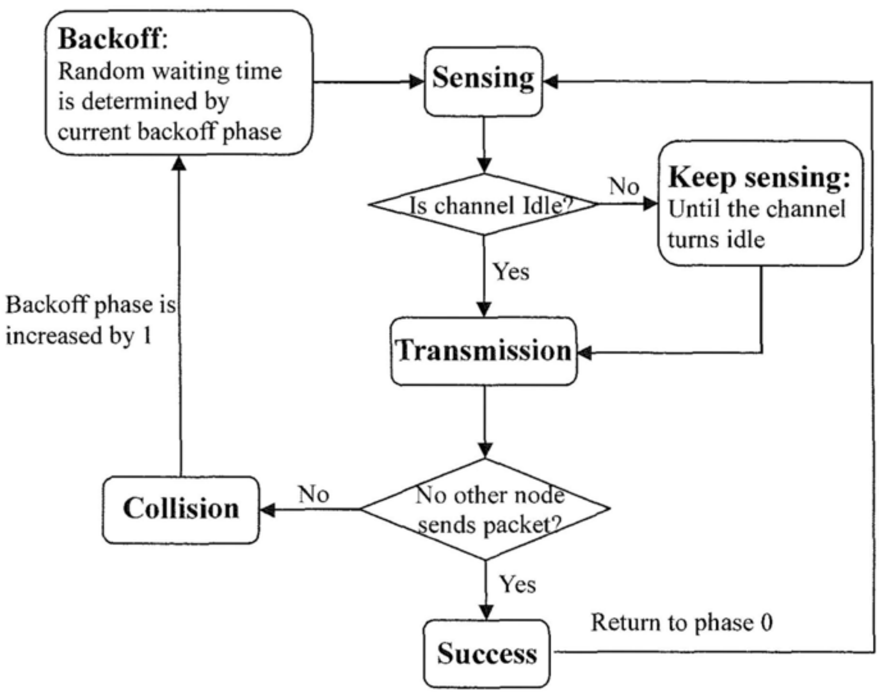


Fig. 3.5. Flow chart of access behaviors for HOL packets of the 1P-CSMA.

Based on the flow chart of the 1P-CSMA protocol shown in Fig. 3.5, the corresponding state transition diagram with the *K*-Exponential Backoff algorithm is shown in Fig. 3.6. An HOL packet is in one of the three fundamental states: sensing (S_i), successful transmission, collision (C_i), and waiting (W_i), $i = 0, \ldots, K$, at any time. Note that the successful transmission state is split into two sub-states: Suc and Suc', corresponding to the two types of successful transmissions, respectively.

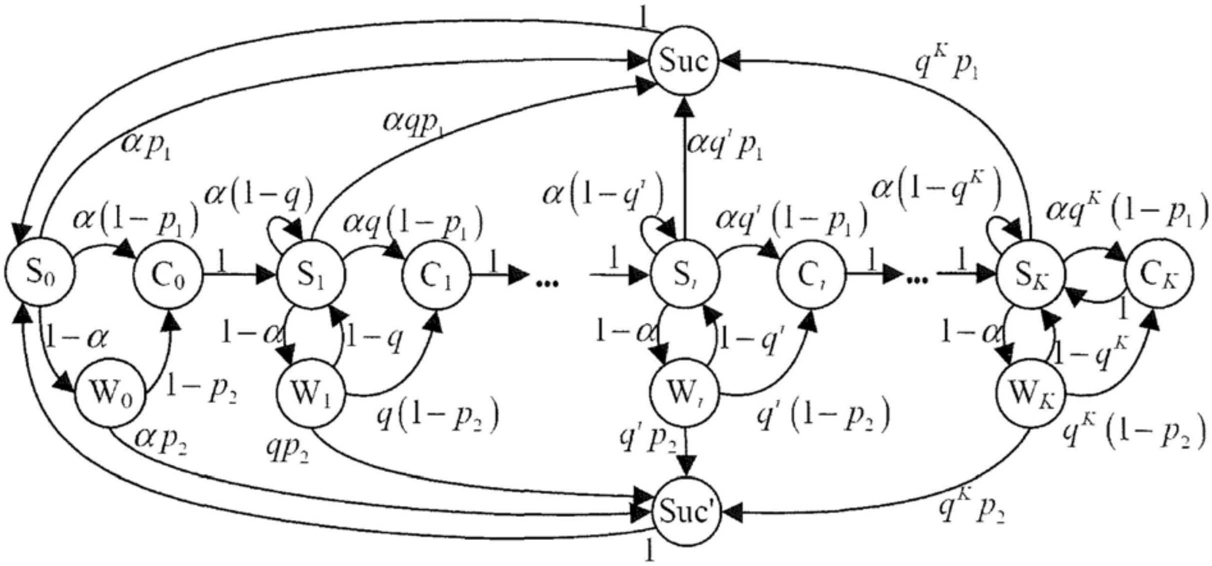


Fig. 3.6. Markov chain of HOL packet of the 1P-CSMA.

If an idle channel is detected in the sensing state S_i with probability α, an HOL packet is sent with probability q^i. With probability p_1, the packet is successfully sent in Suc state; otherwise, collision occurs in the collision state C_i. If a busy channel is detected, the corresponding node continually monitors the channel activity until an idle channel is sensed. This persistent sensing action is represented by the waiting state W_i. When the channel turns idle, the node sends the packet as the Type II transmission with probability q^i. If the packet is successfully sent with probability p_2 in the Suc' state, a fresh HOL packet starts with the initial sensing state S_0; otherwise, the collided packet in the C_i state moves to phase $i+1$ and repeats the above process starting from the sensing state S_{i+1}.

Moreover, for the sake of simplicity, we assume that a new packet at an empty buffer arrives at the beginning of the transmission period if the channel is busy, so that the waiting time is 1 slot time for any phase i. The transition probabilities of the Markov chain are derived below.

The parameter α is a probability that the channel is available for transmission, that is, the channel is in pure idle state. Then it can be calculated from the limiting probabilities given by (3.8) as follows:

$$\begin{aligned}\alpha &= \frac{a\cdot\pi_{Idle}}{a\cdot\pi_{Idle}+(1+a)(\pi_{Suc1}+\pi_{Suc2})+(x+a)\pi_{Col}} \\ &= \frac{ae^{-(x+a)G}\left(1-Ge^{-(1+a)G}+Gxe^{-(1+a)G}\right)}{D_{CH}}\end{aligned} \quad . \tag{3.13}$$

It is illustrated in Fig. 3.3 that the Type I transmission is successful only if all other nodes do not attempt to send packets in the first mini-slot of the busy period. Thus, the probability of a successful Type I transmission is

$$p_1 = e^{-aG} \quad . \tag{3.14}$$

On the other hand, the success of the Type II transmission requires that no other nodes attempt to send packets in the previous transmission period including successful transmission and collision. Hence, the probability of a successful Type II transmission is a conditional probability which conditions on the kind of the previous transmission. The probability that a transmission is success p_s is expressed as,

$$\begin{aligned} p_s &= \frac{(1+a)(\pi_{Suc1}+\pi_{Suc2})}{(1+a)(\pi_{Suc1}+\pi_{Suc2})+(x+a)\pi_{Col}} \\ &\frac{(1+a)e^{-(x+a)G}G(a+x-xe^{-aG})}{(1+a)e^{-(x+a)G}G(a+x-xe^{-aG})+(x+a)\left[1-e^{-aG}-(1+a-e^{-aG})Ge^{-(1+a)G}\right]} \end{aligned} .(3.15)$$

Based (3.15), the probability of successful Type II transmission is given by,

$$p_2 = p_s e^{-(1+a)G} + (1-p_s)e^{-(x+a)G}. \quad (3.16)$$

The Type I transmission lead the busy period; therefore, the probability that a transmission period is Type I is equivalent to the probability α given by (6) that the system is in a pure idle period. It follows that the probability of successful transmission of 1P-CSMA is given by

$$\begin{aligned} p &= \Pr\{\text{Type 1 transmission}\}p_1 + \Pr\{\text{Type 2 transmission}\}p_2 \\ &= \alpha p_1 + (1-\alpha)p_2 \end{aligned} \quad (3.17)$$

Substitute (2.9), (3.14) and (3.16) into (3.17), and the probability of successful transmission can be expressed by the following function of the attempt rate G:

$$p(G) = \frac{e^{-(x+a)G}\left(a+x-xe^{-aG}\right)}{D_{CH}}. \quad (3.18)$$

It should be noted that the probability of success given by (3.18) is consistent with (3.12) , because $p = \frac{\hat{\lambda}}{G}$ by definition. With x = 1, the probability of successful transmission p of the 1P-CSMA/CA can be simplified into the following expression,

$$p(G)=\frac{e^{-(1+a)G}\left(1+a-e^{-aG}\right)}{(1+a)\left(1-e^{-aG}\right)+ae^{-(1+a)G}}. \tag{3.19}$$

Let b_{S_i}, b_{C_i}, b_{W_i}, b_{Suc}, and $b_{Suc'}$ be the respective limiting probabilities of states S_i, C_i, W_i, Suc, and Suc' of the Markov chain shown in Fig. 3.6, from which we obtain the following set of state equations:

$$\begin{aligned}
b_{S_i} &= \begin{cases} b_{Suc}+b_{Suc'}, & \text{for } i=0 \\ \alpha\left(1-q'\right)b_{S_i}+\left(1-q'\right)b_{W_i}+b_{C_{i-1}}, & \text{for } i=1,\dots,K-1 \\ \alpha\left(1-q^K\right)b_{S_K}+\left(1-q^K\right)b_{W_K}+b_{C_{K-1}}+b_{C_K}, & \text{for } i=K \end{cases} \\
b_{C_i} &= \alpha q'\left(1-p_1\right)b_{S_i}+q'\left(1-p_2\right)b_{W_i}, \text{ for } i=0\dots,K \\
b_{W_i} &= (1-\alpha)b_{S_i}, \text{ for } i=0,\dots,K \\
b_{Suc} &= \sum_{i=0}^{K}\alpha q' p_1 b_{S_i} \\
b_{Suc'} &= \sum_{i=0}^{K} q' p_2 b_{W_i}.
\end{aligned} \tag{3.20}$$

It can be proven from (3.20) that if $p+q>1$, then all states of the Markov chain are positive recurrent and aperiodic (i.e. an ergodic system). Thus, the time-average probabilities of those states can be determined from (3.20) with the sojourn time $t_{S_i}=a$, $t_{C_i}=x$, $t_{W_i}=1$, $t_{Suc}=1$, and $t_{Suc'}=1$ of state S_i, C_i, W_i, Suc, and Suc', respectively, for $i=0,\dots,K$ as follows:

$$\tilde{b}_{S_i} = \begin{cases} aD_{HOL}^{-1}, & \text{for } i=0 \\ a(1-p)^i q^{-i} D_{HOL}^{-1}, & \text{for } i=1,\dots,K-1 \\ a(1-p)^K q^{-K} p^{-1} D_{HOL}^{-1}, & \text{for } i=K \end{cases}$$

$$\tilde{b}_{C_i} = \begin{cases} x(1-p) D_{HOL}^{-1}, & \text{for } i=0 \\ x(1-p)^{i+1} D_{HOL}^{-1}, & \text{for } i=1,\dots,K-1 \\ x(1-p)^K p^{-1} D_{HOL}^{-1}, & \text{for } i=K \end{cases}$$

$$\tilde{b}_{W_i} = \begin{cases} (1-\alpha) D_{HOL}^{-1}, & \text{for } i=0 \\ (1-\alpha)(1-p)^i q^{-i} D_{HOL}^{-1}, & \text{for } i=1,\dots,K-1 \\ (1-\alpha)(1-p)^K q^{-K} p^{-1} D_{HOL}^{-1}, & \text{for } i=K \end{cases} \tag{3.21}$$

$$\tilde{b}_{Suc} = \alpha p_1 p^{-1} D_{HOL}^{-1}$$

$$\tilde{b}_{Suc'} = (1-\alpha) p_2 p^{-1} D_{HOL}^{-1},$$

where

$$D_{HOL} = \frac{q(1+a-\alpha)}{p+q-1}\left[1-\frac{(1-q)}{p}\left(\frac{1-p}{q}\right)^{K+1}\right]+\frac{x}{p}+1-x. \tag{3.22}$$

The offered load ρ of each input queue is the probability that the queue is non-empty. Since the input rate λ of the Bernoulli arrival process is the probability of finding a packet arrived at input in any time slot and each input packet will visit the successful transmission states Type I or Type II for exactly one slot time, the input rate λ should be equal to $\rho\left(\tilde{b}_{Suc}+b_{Suc'}\right)$. With the time-average probability of states suc and suc' given in (3.21), the expression of the offered load ρ can be obtained as follows:

$$\begin{aligned}\rho &= \frac{\lambda}{\tilde{b}_{Suc} + \tilde{b}_{Suc'}} \\ &= \lambda D_{HOL} \\ &= \frac{\lambda q(1+a-\alpha)}{p+q-1}\left[1 - \frac{(1-q)}{p}\left(\frac{1-p}{q}\right)^{K+1}\right] + \lambda\left(\frac{x}{p} + 1 - x\right).\end{aligned} \tag{3.23}$$

We show in the next theorem that the network throughput derived from the service time of HOL packets is the same as (3.10), which was previously obtained from the channel model shown in Fig. 3.4. Obviously, the service time is dependent on the retransmission factor q of the backoff scheduling algorithm; but the network throughput is invariant with respect to q. This gives evidence that the stability of the system cannot be analyzed by the characteristic equation of throughput alone.

Theorem 1. For buffered 1-persistent CSMA with K-Exponential Backoff, the throughput in equilibrium is given by

$$\hat{\lambda}_{out} = \frac{-p \ln p_1}{a} = \frac{Ge^{-(1+a)G}\left(x + a - xe^{-aG}\right)}{D_{CH}}. \tag{3.24}$$

Proof: A particular HOL packet is ready to be sent only if an idle channel is detected. The probability of successful Type I transmission p_1 for a desired node is the conditional probability that none of the other nodes accesses the channel, given that all nodes sense the channel is idle, which is given by

$$\begin{aligned} p_1 &= \Pr\{\text{none of other } n-1 \text{ nodes access the channel} \mid \text{channel is sensed idle}\} \\ &= \frac{\Pr\{\text{none of other } n-1 \text{ nodes access the channel}\}}{\Pr\{\text{channel is sensed idle}\}}. \end{aligned} \tag{3.25}$$

If no one accesses the channel, it means that all the other $n-1$ nodes are either empty, or in sensing states but not accessing the channel. Thus, we have

$$\begin{aligned} &\Pr\{\text{none of other } n-1 \text{ nodes access the channel}\} \\ &= \left[\Pr\{\text{node is empty}\} + \Pr\{\text{node is in sensing but not scheduled to send packet}\}\right]^{n-1} \\ &= \left[(1-\rho)+\rho\sum_{i=0}^{K}\tilde{b}_{S_i}\left(1-q^i\right)\right]^{n-1} \overset{\text{for large n}}{=} \exp\left\{-n\rho\left[1-\sum_{i=0}^{K}\tilde{b}_{S_i}\left(1-q^i\right)\right]\right\}. \end{aligned} \tag{3.26}$$

The probability that the node senses an idle channel is given by

$$\begin{aligned} &\Pr\{\text{channel is sensed idle}\} = \left[\Pr\{\text{node is empty}\} + \Pr\{\text{node is in sensing state}\}\right]^{n-1} \\ &= \left[(1-\rho)+\rho\sum_{i=0}^{K}\tilde{b}_{S_i}\right]^{n-1} \overset{\text{for large n}}{=} \exp\left\{-n\rho\left(1-\sum_{i=0}^{K}\tilde{b}_{S_i}\right)\right\}. \end{aligned} \tag{3.27}$$

Substituting (3.21) into (3.26) and (3.27), the probability of successful transmission p_1 defined by (3.25) can be expressed as

$$\begin{aligned} p_1 &= \frac{\exp\left\{-n\rho\left[1-\sum_{i=0}^{K}\tilde{b}_{S_i}\left(1-q^i\right)\right]\right\}}{\exp\left\{-n\rho\left[1-\sum_{i=0}^{K}\tilde{b}_{S_i}\right]\right\}} \\ &= \exp\left\{-\frac{an\rho}{D_{HOL}}\left[\sum_{i=0}^{K-1}(1-p)^i+(1-p)^K p^{-1}\right]\right\}. \end{aligned} \tag{3.28}$$

Then, substituting (3.23) into the above expression, we have

$$p_1 = \exp\left\{\frac{-a\hat{\lambda}}{p}\right\}. \tag{3.29}$$

In equilibrium, the network throughput $\hat{\lambda}_{out}$ should be equal to the aggregate input rate $\hat{\lambda}$. It is easy to show that the network throughput in equilibrium (3.24) can be obtained from (3.29). □

Corollary 1. *For buffered 1-persistent CSMA/CA with K-Exponential Backoff, the throughput in equilibrium is given by*

$$\hat{\lambda} = \frac{Ge^{-(1+a)G}\left(1+a-e^{-aG}\right)}{(1+a)\left(1-e^{-aG}\right)+ae^{-(1+a)G}}. \tag{3.30}$$

Proof: Substituting $x = 1$ into (3.24), the equation (3.30) is immediately obtained. □

Corollary 2. *For buffered non-persistent CSMA/CD with K-Exponential Backoff, the throughput in equilibrium is given by*

$$\hat{\lambda} = Ge^{-(1+a)G}\left(1+a-e^{-aG}\right)\left\{\begin{array}{l}(1+a)Ge^{-(\gamma+a)G}\left(x+a-\gamma e^{-aG}\right) \\ +ae^{-(\gamma+a)G}\left[1-(1-\gamma)Ge^{-(1+a)G}\right] \\ +(\gamma+a)\left[1-e^{-aG}-\left(1+a-e^{-aG}\right)Ge^{-(1+a)G}\right]\end{array}\right\}^{-1}. \tag{3.31}$$

Proof: Substituting $x = \gamma$ into (3.24), we get (3.31). □

We obtain the same characteristic equation of network throughput whatever values of cut-off phase K we choose. Moreover the throughput equation given by (3.24) also agrees with that derived by our channel model in section 2 (3.10). For the CA scheme, the throughput expressed in (3.30) is consistent with that of Kleinrock and Tobagi [6]; while the throughput equation in (3.31) for the CD scheme coincides with the result of Rom and Sidi [8]. The consistency between these approaches indicates that it is appropriate to adopt the independence assumption among input buffers, as mentioned in [3] that the correlation among n input queues becomes weak for large n.

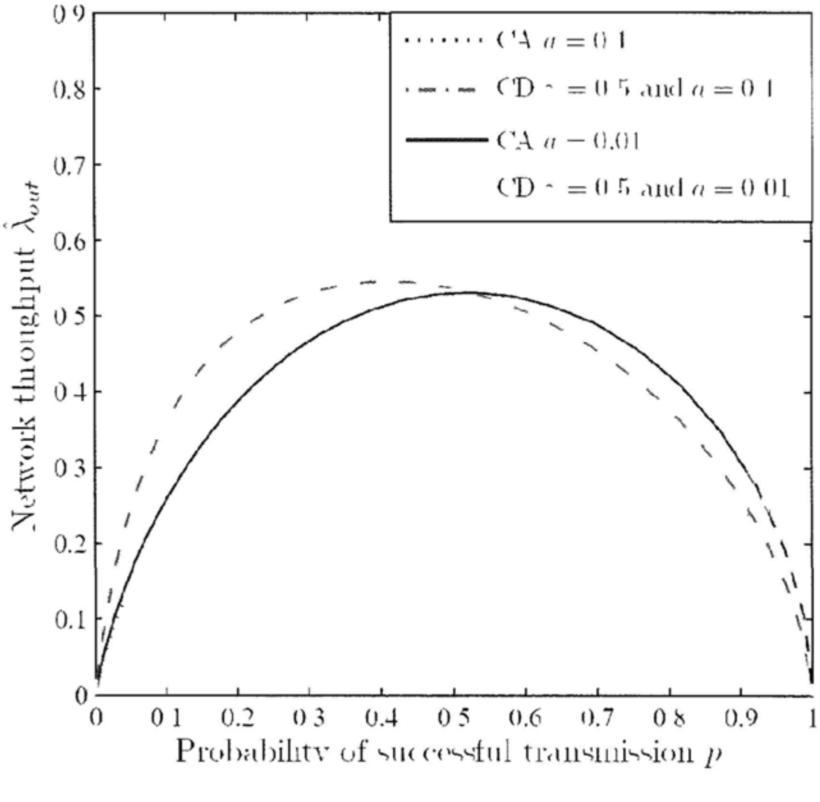


Fig. 3.7. Network throughput versus probability of successful transmission.

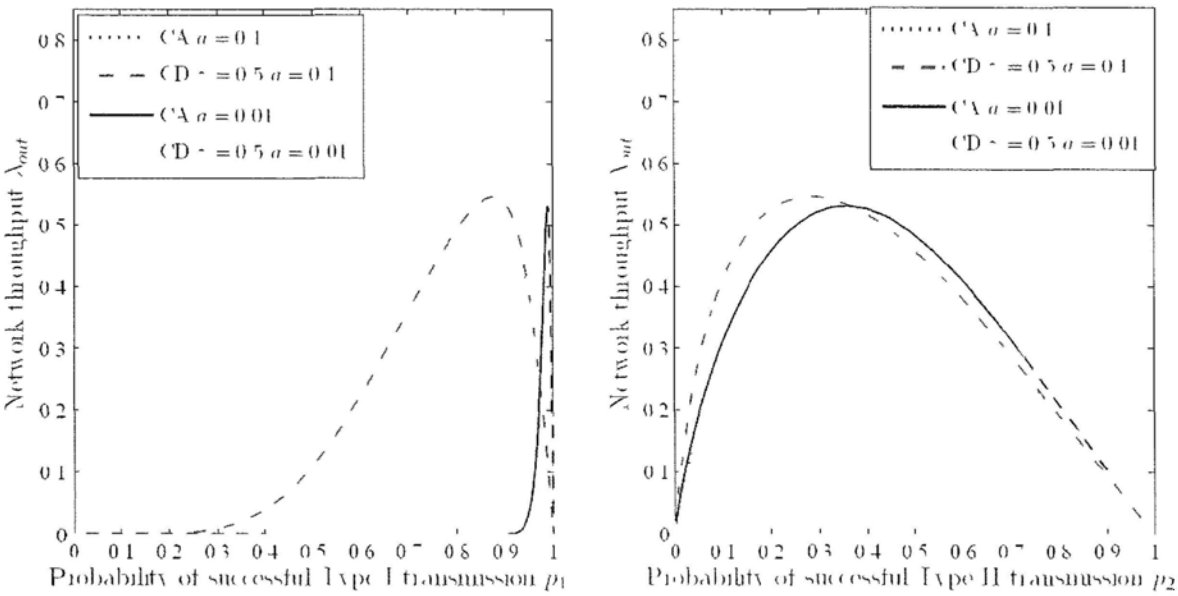


Fig. 3.8. Network throughput versus probability of successful Type I and Type II transmissions.

The characteristic equation of network throughput given in (3.24) describes the relationship between network throughput and probabilities of successful transmission (p_1, p_2, and p), as shown in Fig. 3.7 and Fig. 3.8. For CD scheme, the probability of successful transmission is larger than the counterpart. Moreover, the network throughput and the probability of successful transmission increases with the value of mini-slot size a. It can be explained by considering an un-slotted CSMA system $a = 0$. In this continuous-time system, the probability of successful transmission is equal to one, because the probability that two or more packets transmit simultaneously is equal to zero.

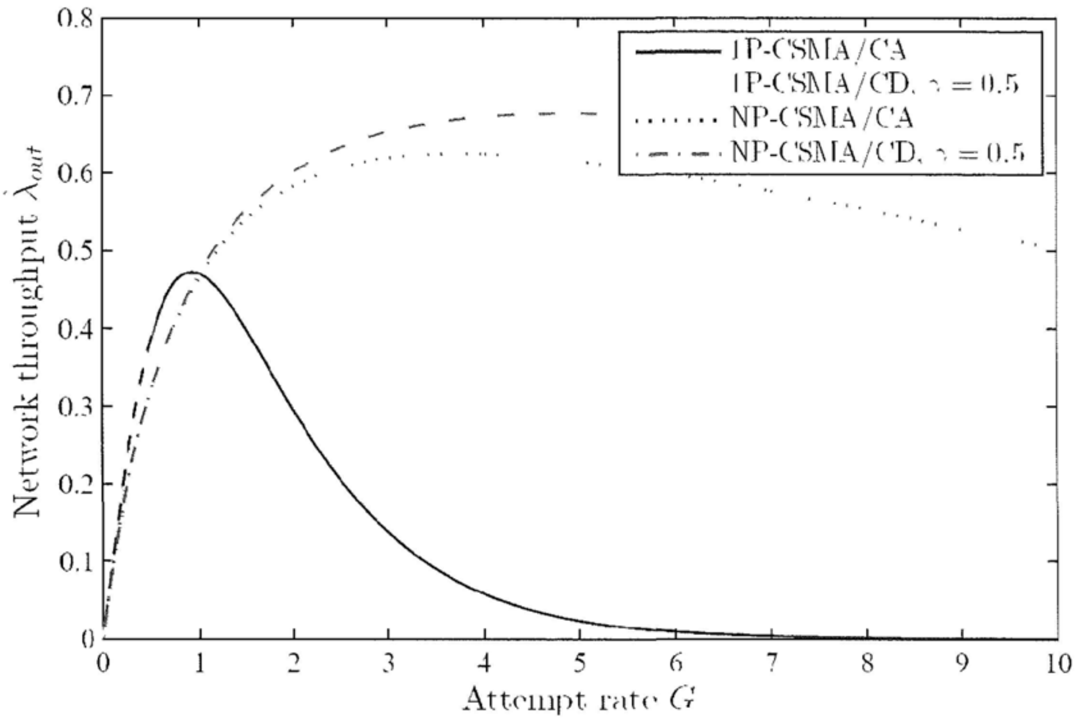


Fig. 3.9. Network throughput versus attempt rate G for 1P-CSMA and NP-CSMA.

Due to the persistent property of the 1P-CSMA, the rate that nodes access the channel is higher than that of NP-CSMA. For low attempt rate G, the network throughput of 1P-CSMA is larger than the NP-CSMA, because of the better channel utilization. When we increase the attempt rate, the probability of collision is also increased. As a result, the NP-CSMA is out-performed under a situation with higher attempt rate. This tradeoff is illustrated in Fig. 3.9.

Furthermore, let random variables S_i^*, C_i^*, W_i^*, Suc^*, and Suc'^* be the service completion time of an HOL packet, starting from the states state S_i, C_i, W_i, Suc, and Suc' respectively, until it is successfully transmitted. We assume, without loss of generality, that $M = a^{-1}$ and $X = xa^{-1}$ are integers. It is straightforward to show from the Markov chain of Fig. 3.6 that the generating functions $S_i(z)$, $C_i(z)$, $W_i(z)$, $Suc(z)$, and $Suc'(z)$ of these service completion times can be found by solving the following set of equations:

$$
\begin{aligned}
S_i(z) &= E\left[z^{S_i^*}\right] \\
&= \alpha\left(1-q'\right)zS_i(z) + \alpha q' p_1 z Suc(z) + \alpha q'\left(1-p_1\right)zC_i(z) + \left(1-\alpha\right)zW_i(z), \text{ for } \forall i \\
W_i(z) &= E\left[z^{W_i^*}\right] \\
&= q' p_2 z^M Suc(z) + q'\left(1-p_2\right)z^M C_i(z) + \left(1-q'\right)z^M \ \mathrm{S}_i(z), \text{ for } \forall i \\
C_i(z) &= E\left[z^{C_i^*}\right] \\
&= \begin{cases} z^X S_{i+1}(z), \text{ for } i = 0,\ldots,K-1 \\ Col_{K-1}(z), \text{ for } i = K \end{cases} \\
Suc(z) &= Suc'(z) = E\left[z^{Suc^*}\right] = z^M.
\end{aligned}
\tag{3.32}
$$

Since the service of each HOL packet always starts from the state S_0, the first and second moments of service time can be derived from the set of generating functions in (3.32) and are given in Appendix B. Note that the mean service time derived in Appendix B is consistent with the expression $\left(\tilde{b}_{Suc} + \tilde{b}_{Suc'}\right)^{-1}$ given in (3.21).

Based on the queueing model of the input buffer, we will investigate various stability issues of 1P-CSMA concerning throughput and delay in the following sections.

3.4 Stable Regions

In this section, we are going to find the stable throughput region and the bounded delay region with respect to the retransmission factor q. Through the study of the relationship between the attempt rate G, input rate λ, and retransmission factor q, we can obtain an equation of the retransmission factor q. We also define the stable conditions in the coming sub-sections. With the functional relationship obtained in sub-section 3.4.1 and the stable conditions, we can determine the stable regions.

3.4.1 The expression of attempt rate G

For a 1P-CSMA network with n nodes, suppose that there are a total of $n_b = \sum_{i=1}^{K} n_i$ HOL packets in a mini-slot, in which n_i packets are in the sensing state of phase i, for i = 0,…, K. The following HOL packets may desire to transmit in the mini-slot:

1) An empty node may send a newly arrived packet with probability $a\lambda$;
2) An HOL packet in phase 0 will be transmitted immediately;
3) A backlogged HOL packet in phase i will be transmitted with probability q^i, for i = 1,…, K.

Hence, the attempts rate per mini-slot, aG, can be expressed as

$$aG = a\lambda E\left[n - n_b\right] + \sum_{i=0}^{K} q^i E\left[n_i\right], \tag{3.33}$$

where $E[n - n_b]$ and $E[n_i]$ are the expected numbers of empty nodes and nodes in sensing state phases i respectively. The mean number of empty nodes in the system is

$$E[n-n_b]=n(1-\rho); \tag{3.34}$$

while the mean number of backlogged nodes in phase i is

$$E[n_i]=n\frac{\tilde{b}_{S_i}}{\sum_{j=0}^{K}\tilde{b}_{S_j}}. \tag{3.35}$$

It follows from (3.21) that the attempt rate in mini-slot defined in (3.33) can be given as follows:

$$\begin{aligned} aG &= a\hat{\lambda}(1-\rho)+\frac{n\rho\sum_{i=0}^{K}\tilde{b}_{S_i}q^i}{\sum_{j=0}^{K}\tilde{b}_{S_j}} \\ &= a\hat{\lambda}(1-\rho)+n\rho\frac{p+q-1}{pq+(p+q-1-pq)\left(\frac{1-p}{q}\right)^K}. \end{aligned} \tag{3.36}$$

By solving equation (3.36), the retransmission factor q can be expressed as a function of the attempt rate G, denoted as $q=h(G)$. For any finite cut-off phase K, this function $h(G)$ is on the order $1/n$, i.e., $q = O(1/n)$, which implies that the network is intrinsically unstable when the number of nodes n is large. This point can be explained by considering the saturated case of Geometric Retransmission ($K = 1$) when all nodes have backlogged HOL packets. In this worst-case scenario, the probabilities of successful transmission for the Type I and Type II are $p_1=e^{-anq}$ and $p_2=e^{-(1+a)nq}$. Consequently, either the retransmission factor q or the probabilities of successful transmission p_1 and p_2 approach zero when the number of nodes $n \to \infty$. This inherently instable property of Geometric Retransmission has also been previously reported in [58] – [60].

On the other hand, Exponential Backoff ($K \to \infty$) mitigates the contention problem by pushing packets to deeper phases. This is the reason that the retransmission factor q expressed in terms of $h(G)$ is on the order of $O(1)$, which suggests that the network can be stabilized even for an infinite population. For Exponential Backoff without cut-off phase ($K \to \infty$), the expression of G can be further simplified to

$$aG = a\hat{\lambda}\left(1-\rho\right)+n\rho\frac{p+q-1}{pq}. \tag{3.37}$$

The retransmission factor q can also be formulated as a function of attempt rate G by substituting (3.23) into (3.36) as follows:

$$q = h(G)$$

$$= \left(1-p\right)\frac{A\left(p\right)+2\hat{\lambda}\left(x+p-px\right)+\sqrt{A\left(p\right)^2+\dfrac{4a\hat{\lambda}^3p^2\left(x+p-px\right)\left(1+a-\alpha\right)}{n}}}{2A\left(p\right)+2\hat{\lambda}\left(x+p-px\right)-\dfrac{2a\hat{\lambda}^2p^2\left(1+a-\alpha\right)}{n}} \tag{3.38}$$

where $A\left(p\right)=ap^2\left(\hat{\lambda}-G\right)+\hat{\lambda}p\left(1+a-\alpha\right)-\dfrac{a\hat{\lambda}^2p\left(x+p-px\right)}{n}$. (3.39)

Obviously, the expressions of the retransmission factor q for both CA and CD schemes are easily obtained by substituting (3.1) into (3.38) and (3.39). It is easy to show that the function $q = h(G)$ in (3.38) monotonically increases with respect to G, as illustrated in Fig. 3.10 for several cases when population size $n = 10$.

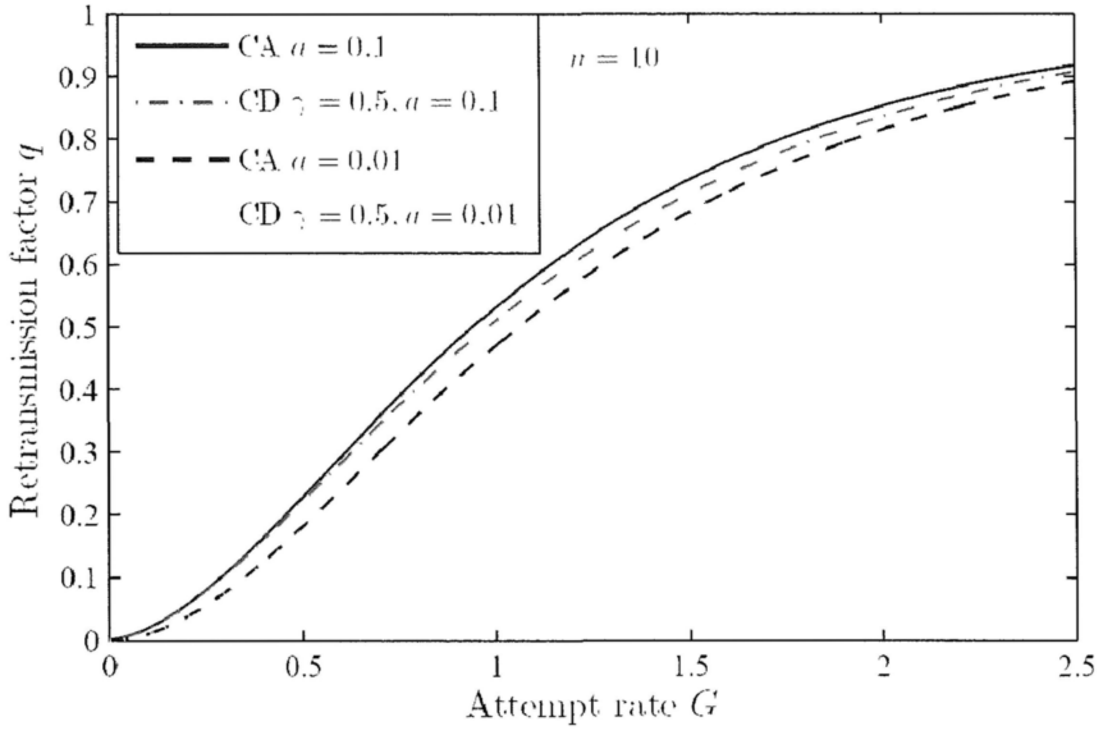


Fig. 3.10. Retransmission factor q versus attempt rate G.

3.4.2 Stable throughput region

The characteristic equation of the throughput (3.24) versus attempt rate G is a curve that first increases and then decreases with G as plotted in Fig. 3.11 for the 1P-CSMA/CA, in which $\hat{\lambda}_{\max}$ denotes the maximum throughput of equation (3.24). This indicates that to have optimal throughput, the attempt rate G cannot be too small or too large. For a throughput smaller than the maximum throughput, $\hat{\lambda} < \hat{\lambda}_{\max}$, the throughput equation (3.24) has two roots; the smaller and larger roots are denoted as $G_S(\hat{\lambda})$ and $G_L(\hat{\lambda})$, respectively. Similarly, examples plotted in Fig. 3.12 also illustrate that the characteristics equation of network throughput contains two roots for any given aggregate input rate $\hat{\lambda}$ less than the maximum throughput $\hat{\lambda}_{\max}$.

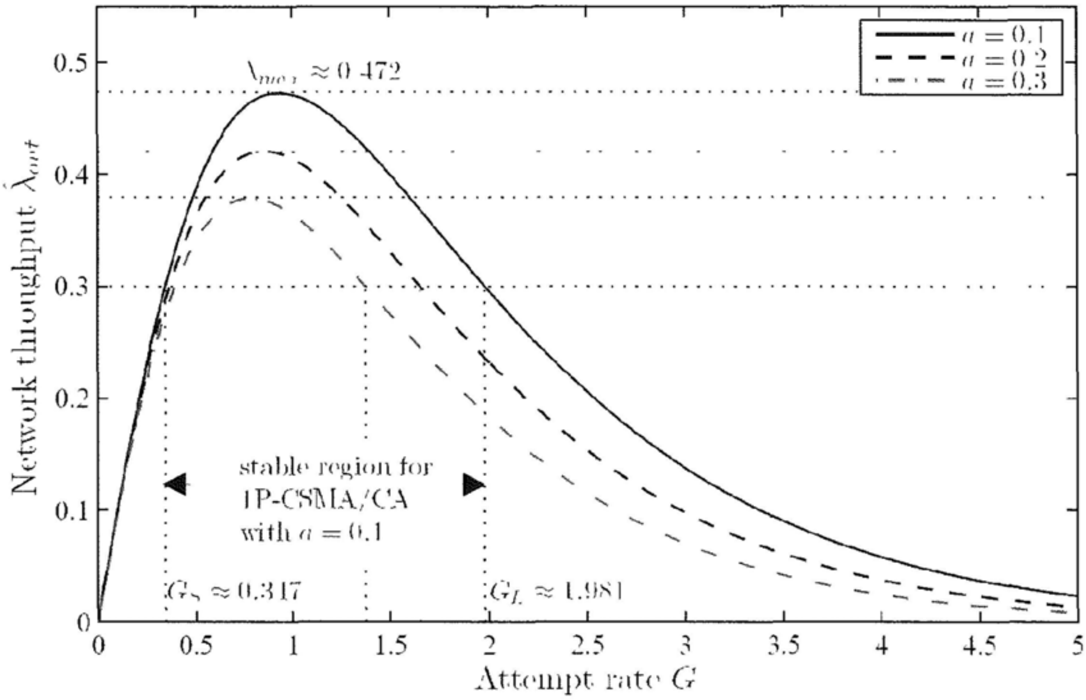


Fig. 3.11. Network throughput versus attempt rate G for 1P-CSMA/CA.

Considering the tradeoff between G and p, we know that G should be bounded in the range between $G_S(\hat{\lambda})$ and $G_L(\hat{\lambda})$ to ensure a stable network throughput where $\hat{\lambda}_{out} = \hat{\lambda}$. Examples in Fig. 3.11 illustrate this point. For a given aggregate input rate $\hat{\lambda} = 0.3$ and a = 0.1, the network has a stable throughput $\hat{\lambda}_{out} = \hat{\lambda} = 0.3$ when G is within the range [≈

0.347, ≈ 1.981]. In Fig. 3.12 for CD scheme γ = 0.5, $\hat{\lambda} = 0.3$ and a = 0.1, the required range of attempt rate G is [≈ 0.345, ≈ 4.22], which is larger than that in Fig. 3.11. Thus, a necessary condition of stable throughput of the entire system can be stated as follows:

Stable Throughput Condition:

STC. For any aggregate input rate $\hat{\lambda} < \hat{\lambda}_{\max}$, the attempt rate G should satisfy

$$G_S(\hat{\lambda}) \le G \le G_L(\hat{\lambda}) . \tag{3.40}$$

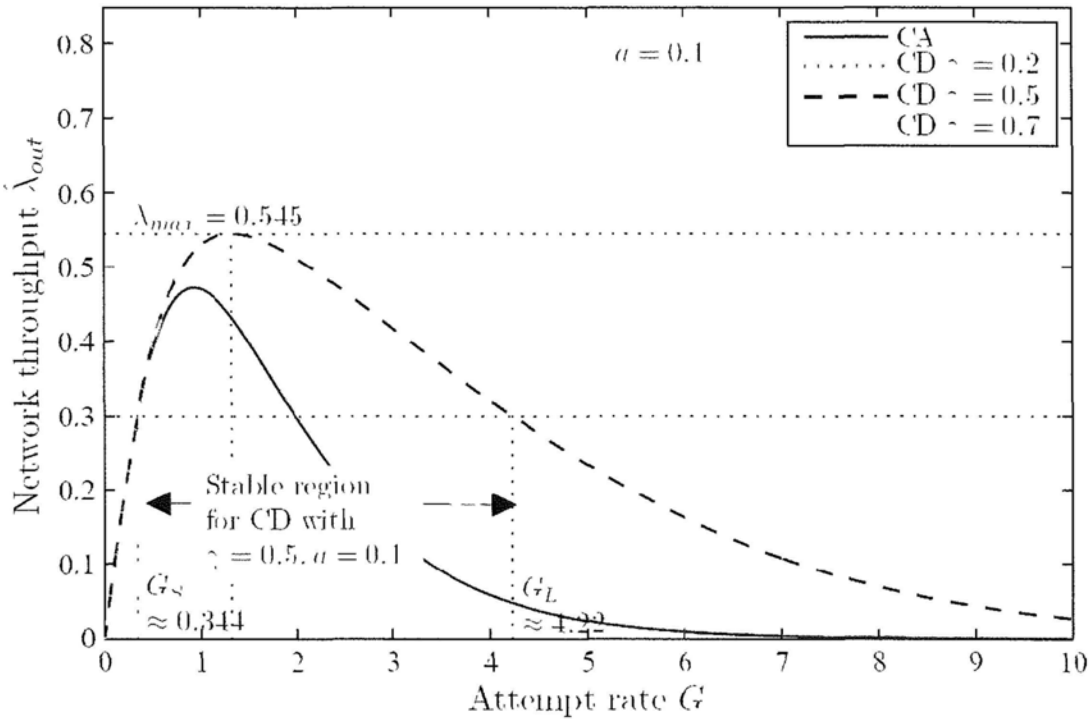


Fig. 3.12. Network throughput versus attempt rate G for 1P-CSMA/CA and CD.

As mentioned in previous chapters, the attempt rate G is an implicit function of the retransmission factor q associated with the underlying scheduling algorithm, the expression is given in (3.38). This functional relationship, together with the inequality (3.40), determines a stable region of q. Since the function (3.38) is a monotonic function with the attempt rate G, the stable throughput region R_T of the retransmission factor q corresponding to the stable throughput condition **STC** can be defined as

$$q \in R_T = \left[h(G_S), h(G_L)\right]. \tag{3.41}$$

Furthermore, the network throughput is defined by $\hat{\lambda}_{out} = \min\{n\left(\tilde{f}_0 + \tilde{f}_0'\right), \hat{\lambda}\}$ and the stable throughput condition **STC** ensures that $\hat{\lambda}_{out} = \hat{\lambda}$. It follows that the **STC** implies $n\left(\tilde{f}_0 + \tilde{f}_0'\right) \geq \hat{\lambda} = n\lambda$, which means the offered load $\rho \leq 1$. It is able to show the offered load ρ in (2.15) is a monotonically increasing function of the retransmission factor q if the attempt rate is bounded in the range $G_S \leq G \leq G_L$. When the offered load $\rho = 1$, the attempt rate G will reach G_L. As a result, if the retransmission factor q is chosen from $q \in \left[h(G_S), h(G_L)\right)$, the offered load ρ of Geo/G/1 queue of each input buffer is strictly less than 1, which is simply the stable condition of any queueing system, that the arrival rate should be strictly less than the service rate.

The stable regions of Exponential Backoff with n = 10 and a = 0.1 are shown Fig. 3.13 under different situations. The area under the lower bound $h(G_S)$ and upper bound $h(G_L)$ is the stable throughput region. For aggregate input rate $\hat{\lambda} = 0.3$, the network throughput can be stabilized when the retransmission factor q is selected within the ranges [≈0.135, ≈0.894], [≈0.127, ≈0.973], [≈0.129, ≈0.929], and [≈0.131, ≈0.897] for Fig. 3.13 (a), (b), (c), and (d) respectively. The sizes of stable throughput regions are in the following ascending order: (b) > (c) > (d) > (a). This implies that a shorter collision period γ gives a greater network throughput and a larger stable throughput region.

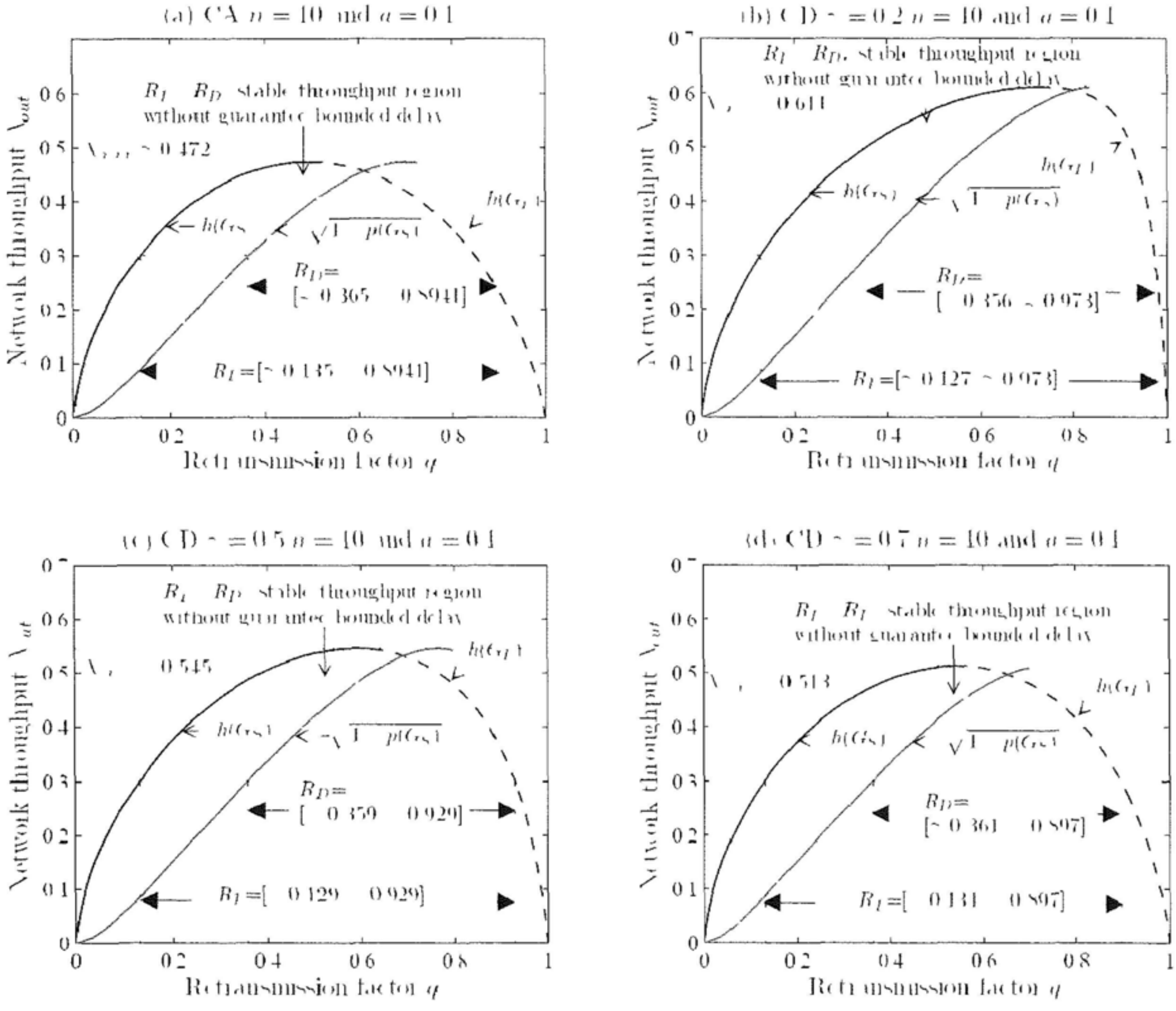


Fig 3 13 Stable throughput and bounded delay regions of Exponential Backoff

Our stability analysis is confirmed by the simulation results shown in Fig. 3.14 with n = 10 and a = 0.1, which closely follows the protocol details for 1-persistent CSMA. The 95% confidence intervals are shown for all the simulations points. For fixed aggregate input rate $\hat{\lambda} = 0.3$, Fig. 3.14 displays that stable throughput can be achieved if the retransmission factor q is properly chosen from the stable throughput region R_T = [≈0.135, ≈0.849]. The simulation results match exactly with the analytical one when the retransmission factor q is within the stable region. While outside the stable region, the throughput immediately drops and the size of the confidence intervals increases due to the unstable queueing behavior of the input buffers.

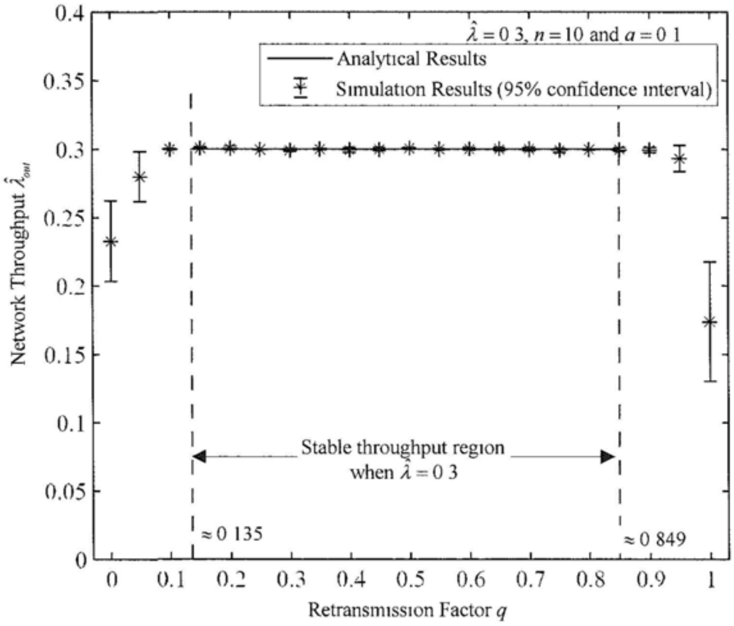


Fig. 3.14. Simulation results of stable throughput region for the 1P-CSMA/CA.

Other simulation results are given in Fig. 3.15 with various values of γ, where the simulated network throughput agrees with our predicted one when the retransmission factor q lies inside the stable region.

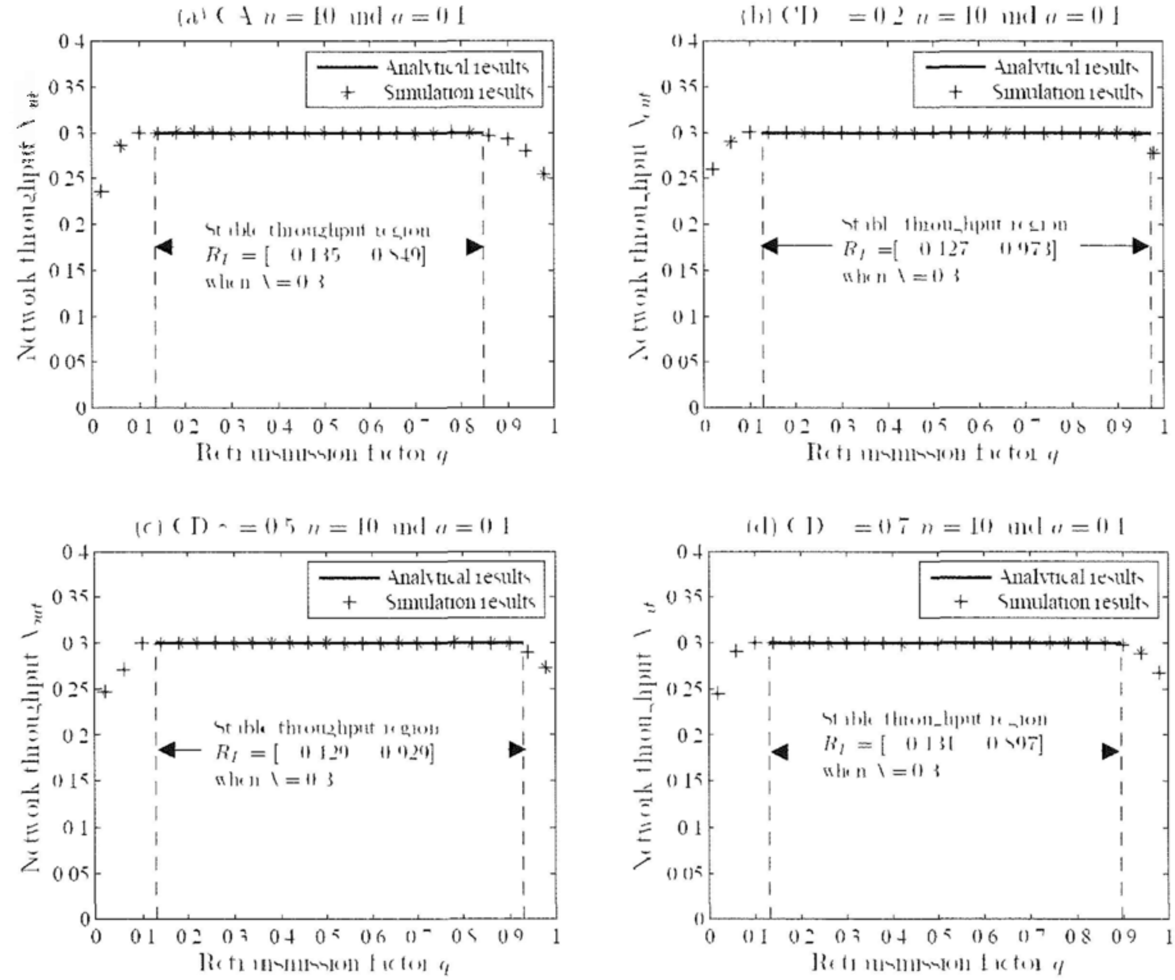


Fig 3 15 Simulation results of stable throughput region for the 1P-CSMA/CA and CD

When n goes to infinity, the expression (3.38) of retransmission factor q can be further reduced to

$$q = 1 - p. \tag{3.42}$$

Hence, from (3.41), we immediately obtain the following non-empty stable throughput region:

$$R_T = [1 - p(G_S), 1 - p(G_L)], \tag{3.43}$$

which coincides with Song et al.'s results on Exponential backoff reported in [58] – [60] that the network throughput can be non-zero even when the number of nodes goes to infinity. This is also demonstrated in Fig. 3.16 that the stable throughput regions are similar for different values of n.

3.4.3 Bounded delay region

The stable throughput condition (3.40) cannot give us any insight about the packet queueing delay. We deduce an additional constraint from our queueing model of the input buffer and the Pollaczek-Khinchin formula to guarantee bounded queueing delay.

Bounded Delay Condition:

BDC. *The Pollaczek-Khinchin formula for mean delay $E[T]$ of Geo/G/1 queue* [68]

$$E[T] = E[X] + \lambda \frac{E[X^2] - E[X]}{2(1 - \lambda E[X])} \tag{3.44}$$

requires bounded second moment of service time $0 < E[X^2] < \infty$.

The condition **BDC** is stronger than the stable throughput condition **STC.** We define a bounded delay region R_D the retransmission factor q is a region with of guarantees bounded mean queueing delay. This region can be specified by the second moment of service time $E[X^2]$, and it is a subset of the stable throughput region R_T.

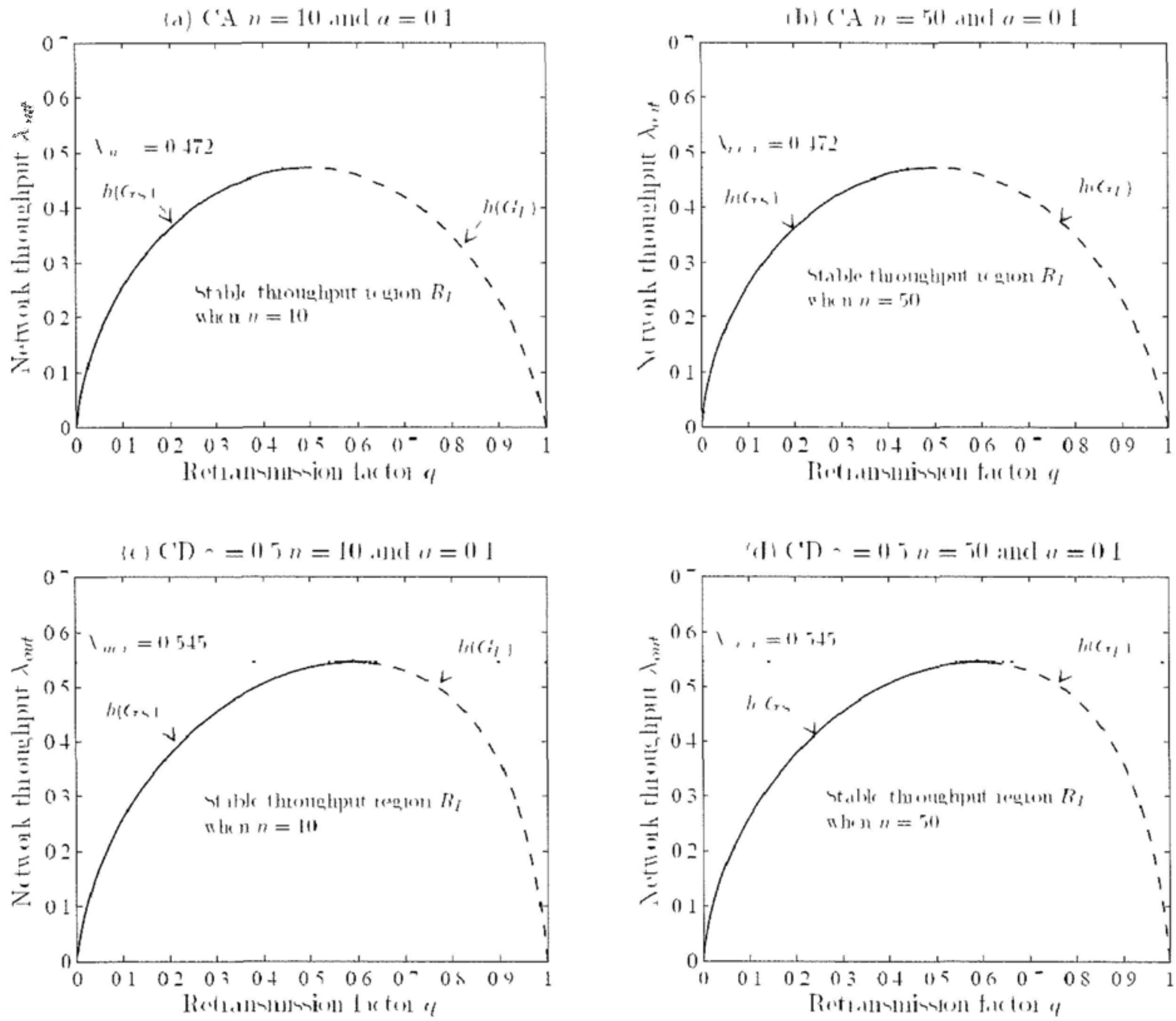


Fig. 3.16. Stable throughput region of Exponential Backoff under different population sizes *n*.

The second moment of service time of the Exponential Backoff scheme is given by (B2) in Appendix B as follows:

$$E[X^2] = B(p,q)\lim_{K\to\infty}\left(\frac{1-p}{q^2}\right)^{K-1} + C(p,q), \tag{3.45}$$

where $B(p,q)$ and $C(p,q)$ are two polynomials given by (B3) and (B4), respectively, in Appendix B. Thus, the retransmission factor should satisfy the following condition to guarantee a bounded second moment of service time $E[X^2]$:

$$q > \sqrt{1-p}\,. \tag{3.46}$$

The bounded delay region R_D of Exponential Backoff can be determined by the combination of the above condition (3.46) and the stable throughput region R_T excluding the point of upper bound, and given as follows:

$$R_D = \left[\sqrt{1-p(G_S)}, h(G_L)\right). \tag{3.47}$$

From Appendix B, we can derive the mean access delay $E[X]$ and queueing delay $E[T]$:

$$E[X] = a\left[M - X + \frac{X}{p} + \frac{q(M+1-\alpha M)}{(p+q-1)}\right] \tag{3.48}$$

and

$$
\begin{aligned}
E[T] = E[X] - \frac{\lambda E[X]}{2(1-\lambda E[X])} + \frac{a^2\lambda}{2(1-\lambda E[X])}[(2X-M)(M+1-\alpha M)+\alpha M^2 \\
+\frac{(X-1)X+2(1-\alpha)p_2(M-X)M}{p}+\frac{2q^3(M+1-\alpha M)^2}{(p+q^2-1)(p+q-1)}-(X+1)X \\
+q\frac{(-2p+2\alpha pM+pM-2pX+2X)(M+1-\alpha M)-\alpha pM}{p(p+q-1)} \\
+2\left[X+p(M-X-\alpha M)-(1-\alpha)p_2M\right]\left[\frac{pM-pX+X}{p^2}+\frac{q(M+1-\alpha M)}{(p+q-1)^2}\right]]
\end{aligned}
\tag{3.49}
$$

Although access delay $E[X]$ in (3.48) is same as that of the NP-CSMA in (2.43), the probability of success p of the 1P-CSMA in (3.48) contains p_1 and p_2 which is different from that of the NP-CSMA. At equilibrium, the maximum achievable throughput of the NP-CSMA protocol is larger than that of the 1P-CSMA as illustrated in Fig. 3.17. This can also be observed by comparing the stable throughput regions of the two protocols in Fig. 2.9 and Fig. 3.13.

The bounded delay region R_D of exponential backoff is a subset of the stable throughput region R_T. As shown in Fig. 3.13, the shaded region is a bounded delay region; outside this region, i.e., $R_T \setminus R_D$, the system has stable throughput but cannot guarantee bounded delay. In this undesired region, predominating backlogged packets are pushed to deep phases with very low retransmission probabilities. If a node tries to send its HOL packet, the successful probability will be very high. Once the backlogged HOL packet is cleared, then the channel may be "captured" by subsequent packets in the input buffer, which are all in phase 0, until the queue is cleared. During this period, it appears that the network throughput is still stable, but the variance of the service time of each individual packet can be infinitely large due to this unfairness of services caused by the capture effect described in [67].

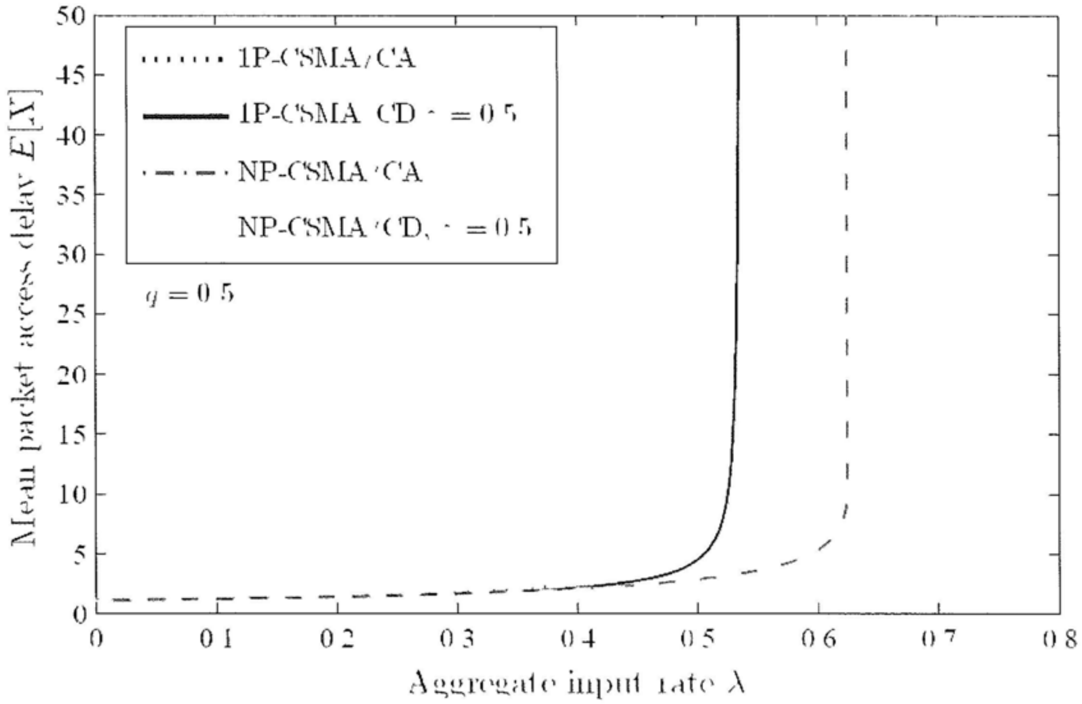


Fig. 3.17. Packet access delay verses aggregate input rate for the 1P-CSMA and NP-CSMA.

Fig. 3.18 and Fig. 3.19 demonstrate the packet-queueing delay $E[T]$ of the 1P-CSMA versus the retransmission factor q for the 1P-CSMA/CA and CD protocols respectively. Our analysis is confirmed by simulation as shown in both figures. If the retransmission factor q is chosen within the region of bounded delay R_D, then the mean packet queueing delay keeps small. However, if the retransmission factor q lies outside the bounded delay region, the packet queueing delay becomes larger even though it is still within the stable throughput region. For any retransmission factor q outside the stable throughput region, the queueing delay immediately jumps to an unacceptable level. It also shows that the bounded delay region of the 1P-CSMA/CD with smaller value γ of is larger than that of the CA scheme.

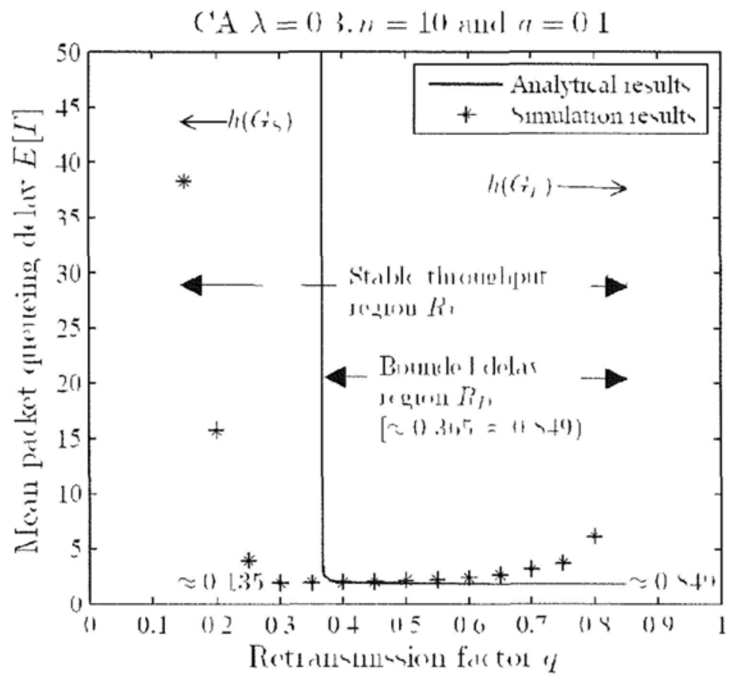


Fig. 3.18. Packet queueing delay verses retransmission factor q for the 1P-CSMA/CA.

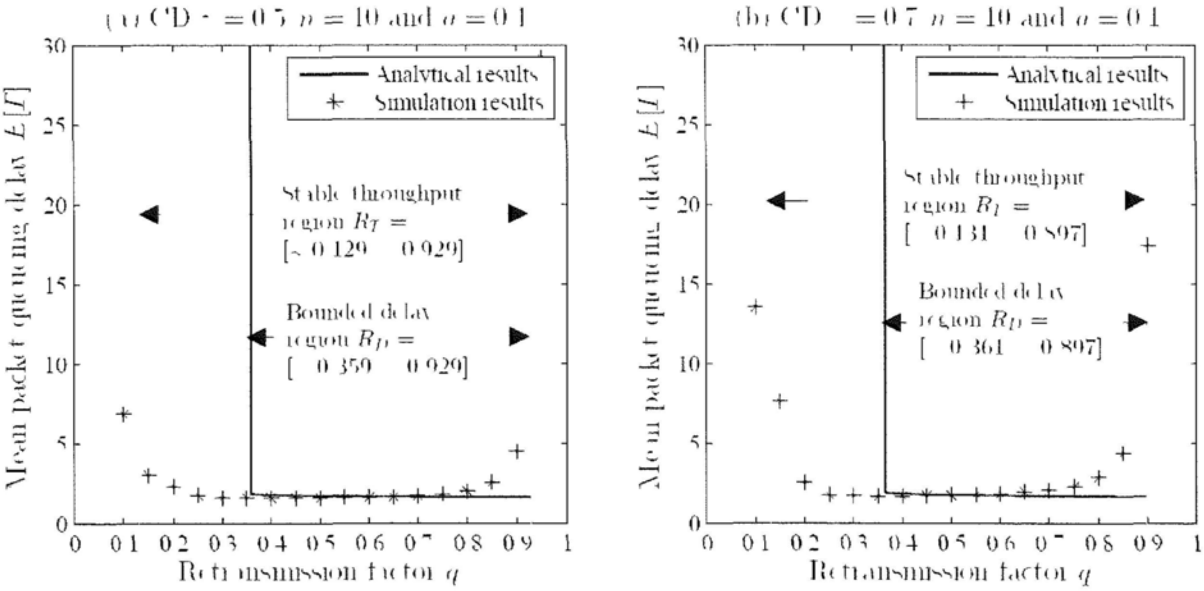


Fig 3 19 Packet queueing delay verses retransmission factor q for the 1P-CSMA/CD

On the other hand, Fig. 3.20 exhibits the mean packet queueing delay $E[T]$ versus aggregate input rate $\hat{\lambda}$ for fixed retransmission factor $q = 0.5$. It can be clearly seen that when aggregate input rate $\hat{\lambda}$ lies inside the bounded delay region, the analytical results match with simulation. The queueing model of input buffers becomes unstable outside the stable region, and we cannot expect the system behavior to be accurately predicted by any analysis, because the Markov chain that describes HOL packets is only valid when the queueing process is stationary. Moreover, from the figure, we can observe that the queueing delay increases and eventually becomes unacceptably large when aggregate input rate is outside the stable region. This result coincides with [58] in which the authors claimed that the exponential backoff scheme can be stable if the aggregate input rate is sufficiently small.

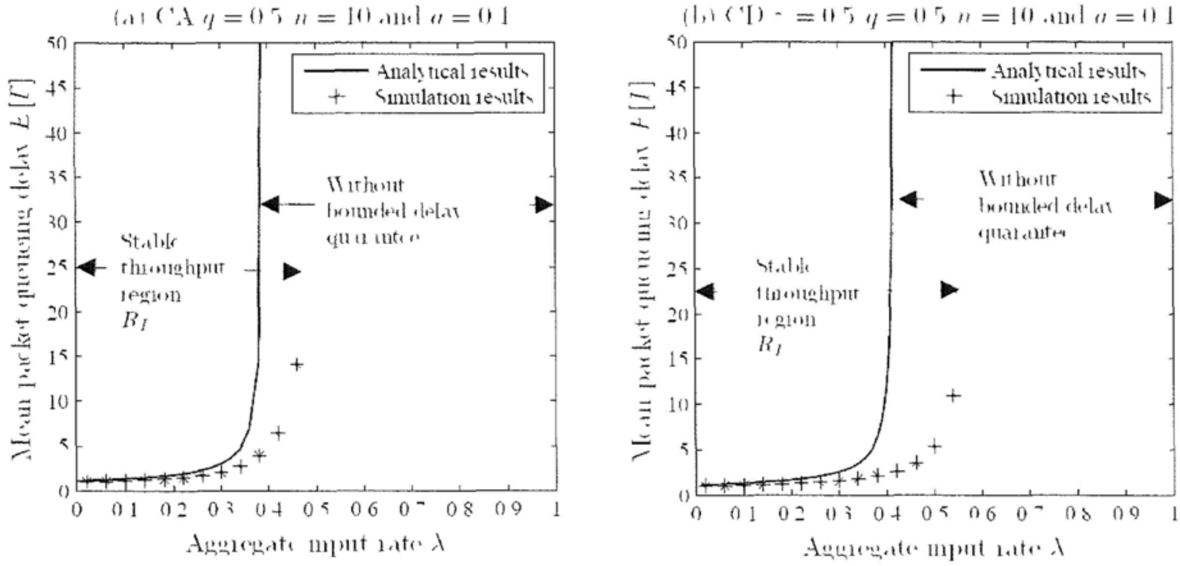


Fig. 3.20. Packet queueing delay verses aggregate input rate.

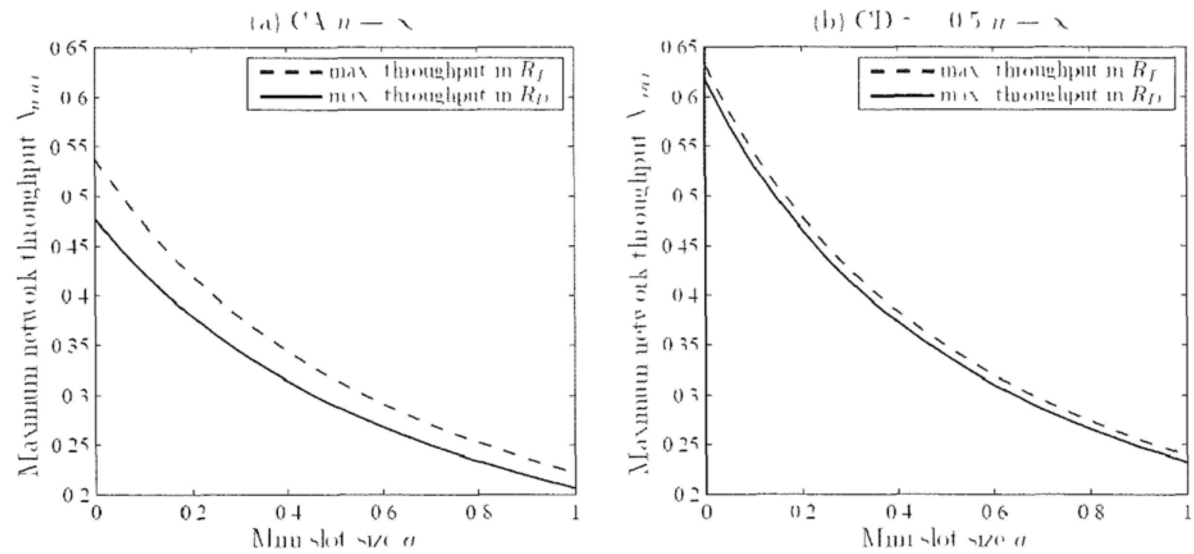


Fig. 3.21. Maximum throughputs in the stable throughput and bounded delay regions.

Since the bounded delay region R_D specified by (3.47) is a subset of the stable throughput region R_T, the maximum throughput $\hat{\lambda}_{\max}$ may not be achievable if $q \in R_D$. As shown in Fig. 3.13 for the 1P-CSMA/CA, the network throughput increases with the lower bound of the bounded delay region $\sqrt{1-p(G_S)}$, and decreases with the upper bound $1-p(G_L)$. Therefore, the maximum throughput within the bounded delay region R_D can be achieved when the lower bound and upper bound of the retransmission factor q given by (30) are equal. That is,

$$\sqrt{1-p(G_S)} = 1-p(G_L). \tag{3.50}$$

In Fig. 3.21, we plot the respective maximum throughput within the stable throughput region and bounded delay region. For $0 < a < 1$, the maximum throughput $\hat{\lambda}_{\max}$ given in (3.24) is always larger than that obtained from the bounded delay region. In other words, the absolute maximum throughput $\hat{\lambda}_{\max}$ in the stable throughput region cannot be achieved with bounded mean delay guarantee.

3.5 Conclusion

We have analyzed the stability conditions in terms of stable throughput and bounded delay for slotted 1-persistent CSMA/CA networks. The queueing model of input buffer with the K-Exponential Backoff collision resolution algorithm is proposed to conduct the throughput analysis of the entire system and the performance of each individual input buffer. Based on this model, the exponential backoff scheduling algorithm has been used to establish the stable throughput region and bounded delay region with respect to aggregate input rate and retransmission factor q. It is shown that the network throughput of the exponential backoff scheme can always be stabilized with the proper selection of q; while the region shrinks remarkably subject to the bounded delay requirement. Consequently, the maximum achievable throughput of the network within the bounded delay region is slightly smaller than the absolute maximum throughput. The proposed methodology can also be applied to other CSMA-based networks, such as IEEE 802.14.5, in the future.

Chapter 4

A Generalized model of NP-CSMA Protocols with Window-based Binary Exponential Backoff

We focus on the window-based backoff scheduling algorithm of NP-CSMA protocols and show that the window-based algorithm and the probabilistic-based algorithm in chapter 3 and 4 are equivalent.

4.1 Window-based verse Probability-based Exponential Backoff Scheduling Algorithms

In practical, any collapsed packets are backlogged to alleviate the traffic loads and so as to increase the probability of successful transmission. For exponential backoff scheduling algorithm, packets encounter collisions i times are in backoff phase i. The backlogged packet will be scheduled for retransmission in which the waiting time depends on the current backoff phase.

In probability-based backoff scheduling algorithms (e.g. in chapter 3 and 4), the backlogged packet in phase i has retransmission probability q^i, where q is the backoff factor with value between 0 and 1. While the window-based backoff scheduling algorithm consists of backoff timer with uniformly distributed from a range of (0, CW-1), where the window size CW, called the ***Contention Window.*** The contention window in phase i is equal to

$$CW = W_i . \tag{4.1}$$

For the binary exponential backoff, the contention window in phase i is expressed as,

$$W_i = 2^i W , \tag{4.2}$$

where W is the minimum contention window with value larger or equal to 1.

4.2 Queueing Model of Input Buffer

Due to the similarity between window-based and probability-based backoff algorithms, we adopt the notations in chapter 2 as listed in TABLE 4.1. To consider the window-based backoff algorithm, the one-dimensional Markov chain introduced in chapter 3 and 4 is no longer suitable here. Instead, we use a two-dimensional Markov chain to describe the behavior of the HOL packet including the backoff phase and also the backoff counter.

We consider the NP-CSMA protocol with binary exponential backoff. The input buffer of each node is modeled as a Geo/G/1 with Bernoulli arrival process of rate λ

packets/timeslot. A fresh HOL packet is sent only when the idle channel is being sensed, and the packet is scheduled for retransmission at a later time if a collision occurred. A backlogged HOL packet is in phase i if it has encountered collisions i times. Collapsed packets in phase i are pushed into the next phase with backoff time randomly chosen from a range of (0, $W_{i+1} - 1$), where is expressed in (4.2). Let K be the cut-off phase, then the contention window size increases exponentially with the number of collisions, up to W_K, experienced by the HOL packet.

a	A mini-slot size
γ	Duration of collision for CD scheme
x	Duration of collision, in which $x = \begin{cases} 1 & \text{for CA scheme} \\ \gamma & \text{for CD scheme} \end{cases}$
λ	Packet arrival rate in a node
G	Aggregate attempt rate
n	Number of nodes in the system
$\hat{\lambda}$	Aggregate input rate $\hat{\lambda} = n\lambda$
ρ	Offered load of each node

TABLE 4.1. Notations

It is demonstrated in the previous chapters that the packet service distribution highly depends on the employed backoff scheduling algorithm; however, it does not affect the channel activities. Therefore, both probability-based and window-based backoff algorithms have the same probability in finding an idle channel α and probability of successful transmission p which is derived in chapter 3:

$$\alpha = \frac{a}{aGe^{-aG} + a + x\left(1 - e^{-aG} - aGe^{-aG}\right)}, \tag{4.3}$$

$$p = e^{-aG}. \tag{4.4}$$

The Markov chain of HOL packet should also contains the four fundamental states: successful transmission (Suc), waiting, sensing (S_i), and collision (C_i) states, where $i = 0, 1, \ldots, K$. In order to represent backoff phases and backoff time, the waiting state should be two-dimensional $\{i, j\}$, where i is the backoff phase $= 0, 1, \ldots, K$ and j is the backoff time $= 1, 2, \ldots, W_i - 1$. The state transition diagram is shown in Fig. 4.1.

If a busy channel is sensed, the packet is delayed for a random time uniformly chosen from the interval $[0, W_0 - 1]$. In other words, after facing a busy channel, the fresh packet will be equally likely in phase 0 sensing or waiting state. If an idle channel is detected in the sensing state S_i, then the packet will be transmitted. With probability p, it is successfully transmitted in the state Suc. After a successful transmission, a fresh HOL packet starts with the initial sensing state S_0. On the other hand, collision occurs with probability $1 - p$, the collided packet in the state C_i is moved to sensing state or other waiting states phase $i + 1$ with equally likely probability, that is, W_i^{-1}. After collision, the above process is repeated in the next phase.

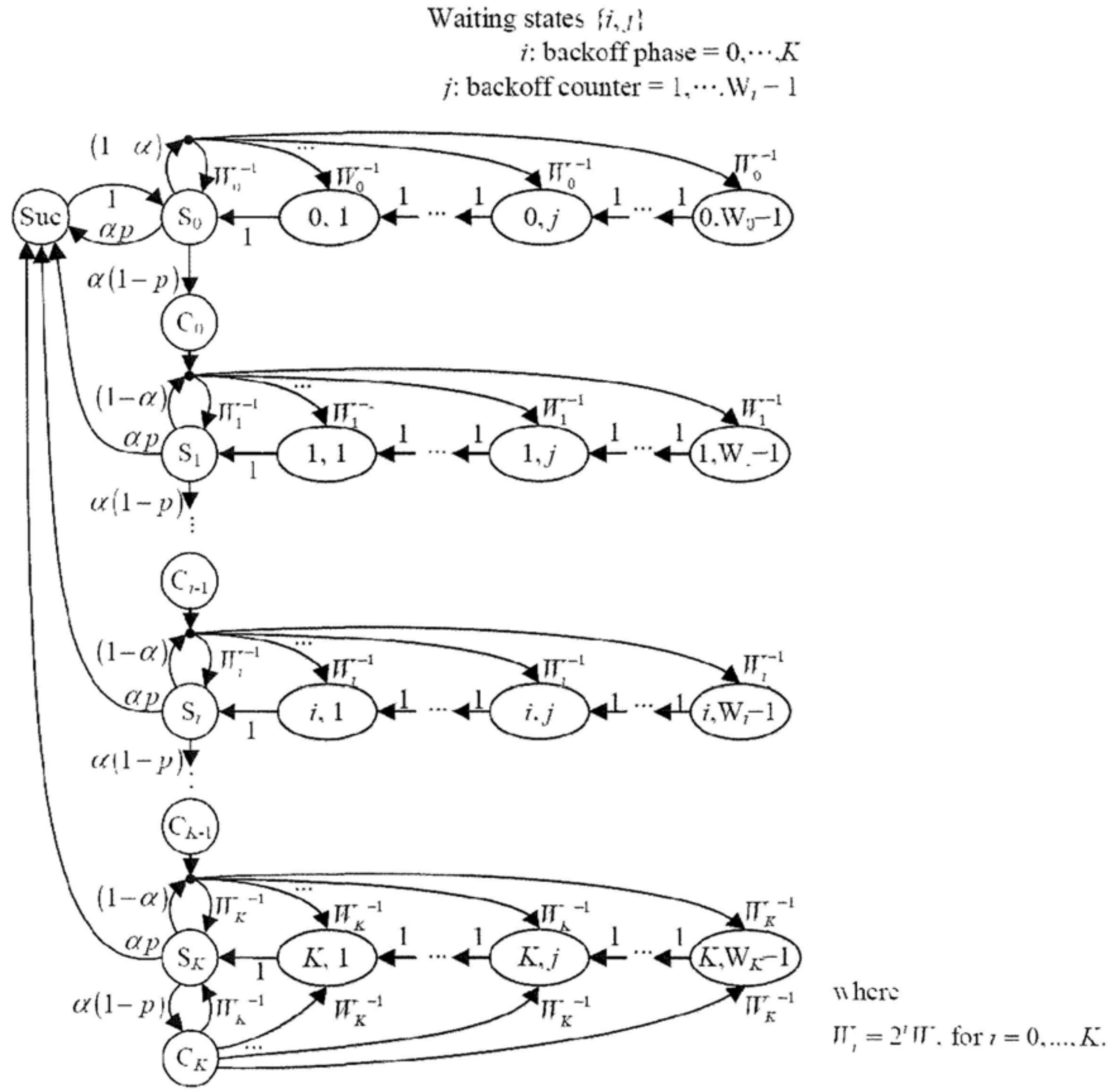


Fig. 4.1. Markov chain of HOL packet with binary exponential backoff.

Let b_{Suc}, b_{S_i}, b_{C_i}, and $b_{i,j}$ be the limiting probabilities of states Suc, S_i, C_i, and waiting state $\{i, j\}$ respectively. From the Markov chain described in Fig. 4.1, we obtain the following set of state equations:

$$\begin{cases} b_{Suc} = \alpha p\left(b_{S_0} + b_{S_1} + \cdots + b_{S_K}\right) \\ b_{S_0} = b_{Suc} + (1-\alpha)W_0^{-1}b_{S_0} + b_{0,1} \\ b_{S_K} = (1-\alpha)W_K^{-1}b_{S_K} + W_K^{-1}b_{C_{K-1}} + W_K^{-1}b_{C_K} + b_{K,1} \\ b_{S_i} = (1-\alpha)W_i^{-1}b_{S_i} + W_i^{-1}b_{C_{i-1}} + b_{i,1}, & \text{for } i = 1,\dots,K-1 \\ b_{C_i} = \alpha(1-p)b_{S_i}, & \text{for } i = 0,\dots,K \\ b_{0,W_0-1} = (1-\alpha)W_0^{-1}b_{S_0} \\ b_{0,j} = b_{0,j+1} + (1-\alpha)W_0^{-1}b_{S_0}, & \text{for } j = 1,\dots,\mathrm{W}_0-2 \\ b_{i,W_i-1} = W_i^{-1}b_{C_{i-1}} + (1-\alpha)W_i^{-1}b_{S_i}, & \text{for } i = 1,\dots,K-1 \\ b_{i,j} = b_{i,j+1} + W_i^{-1}b_{C_{i-1}} + (1-\alpha)W_i^{-1}b_{S_i}, & \text{for } i = 1,\dots,K-1 \text{ and } j = 1,\dots,\mathrm{W}_i-2 \\ b_{K,W_K-1} = W_K^{-1}b_{C_{K-1}} + W_K^{-1}b_{C_K} + (1-\alpha)W_K^{-1}b_{S_K} \\ b_{K,j} = b_{K,j+1} + W_K^{-1}b_{C_{K-1}} + W_K^{-1}b_{C_K} + (1-\alpha)W_K^{-1}b_{S_K}, \text{for } j = 1,\dots,W_K-2 \end{cases} \quad . \tag{4.5}$$

It can be observed from (4.5) that if

$$2(1-p) < 1, \tag{4.6}$$

all states of the Markov chain are positive recurrent and aperiodic. From (4.6), the ergodic condition is independent of the minimum contention window size W. The time-average probability of each state can be determined from (4.5) with the sojourn time $t_{Suc} = 1$, $t_{S_i} = a$, $t_{C_i} = x$, and $t_{i,j} = a$ of states Suc, S_i, C_i, and $\{i, j\}$, respectively, for $i = 0,1\dots,K$ and $j = 1,\dots,W_i - 1$ as follows:

$$\tilde{b}_{suc} = \frac{\alpha p}{D_{HOL}}$$

$$\tilde{b}_{s_i} = \begin{cases} \frac{ap}{D_{HOL}}(1-p)^i & \text{for } i = 0,\ldots K-1 \\ \frac{a}{D_{HOL}}(1-p)^K & \text{for } i = K \end{cases}$$

$$\tilde{b}_{c_i} = \begin{cases} \frac{x\alpha p}{D_{HOL}}(1-p)^{i+1} & \text{for } i = 0,\ldots K-1 \\ \frac{x\alpha}{D_{HOI}}(1-p)^{K+1} & \text{for } i = K \end{cases}$$

$$\tilde{b}_{i,j} = \begin{cases} \frac{ap}{D_{HOI}}(1-\alpha)\left(1-\frac{j}{W}\right) & \text{for } i = 0, \text{and } j = 1,\ldots,W_0 - 1 \\ \frac{ap}{D_{HOL}}(1-p)^i\left(1-\frac{j}{2^i W}\right) & \text{for } i = 1,\ldots K-1, \text{and } j = 1,\ldots,W_i - 1 \\ \frac{a}{D_{HOL}}(1-p)^K\left(1-\frac{j}{2^K W}\right) & \text{for } i = K, \text{and } j = 1,\ldots,W_K - 1 \end{cases} , \quad (4.7)$$

where $D_{HOL} = \alpha p + x\alpha(1-p) + \frac{a(1+\alpha p-\alpha pW)}{2} + \frac{apW}{2(2p-1)} - \frac{aW2^K(1-p)^{K+1}}{2(2p-1)}$.

The offered load ρ of each input queue is the probability that the queue is non-empty. It is the basic measurement for analyzing the performance of each input buffer. The input rate λ of the Bernoulli arrival process can be interpreted as the probability of finding a packet arrived at input in any time slot. Each input packet will eventually become a fresh HOL packet, and visit the successful transmission state (Suc) for one slot time. Therefore, the input rate λ should equal $\rho\tilde{b}_{suc}$ the probability of finding an HOL packet in the successful transmission state in any time slot. With the time-average probability of state Suc given in (4.7), the expression of the offered load ρ can be obtained as follows:

$$\rho = \frac{\lambda}{\tilde{b}_{suc}} = \lambda\left[1 + \frac{x}{p} - \frac{a(W-1)}{2} + \frac{a}{2\alpha p} + aW\frac{p - 2^K(1-p)^{K+1}}{2\alpha p(2p-1)}\right]. \qquad (4.8)$$

The offered load can be also regarded as λE[X], in which $E[X]$ is the mean service time of each HOL packets. Then $E[X]$ should be equal to $\tilde{b}_{suc}{}^{-1}$.

Furthermore, let random variables Suc^*, S_i^*, C_i^* and $B_{i,j}^*$ be the service completion time of an HOL packet, starting from the states Suc, S_i, C_i, and $\{i, j\}$, respectively, until it is successfully transmitted. We assume, without loss of generality, that $M = a^{-1}$ and $X = xa^{-1}$ are integers. It is easy to show from the Markov chain of Fig. 4.1 that the generating functions $Suc(z)$, $S_i(z)$, $C_i(z)$, and $B_{i,j}(z)$ of these service completion times can be found by solving the following set of equations:

$$\begin{aligned}
S_i(z) &= E\left[z^{S_i^*}\right] \\
&= \begin{aligned}&\alpha pzSuc(z) + \alpha(1-p)zC_i(z) + (1-\alpha)W_i^{-1}zS_i(z) \\ &+ (1-\alpha)W_i^{-1}z\left(B_{i,1}(z) + \ldots + B_{i,W_i-1}(z)\right), \qquad \text{for } i = 0,\ldots,K\end{aligned} \\
Suc(z) &= E\left[z^{Suc^*}\right] = z^M \\
C_i(z) &= E\left[z^{C_i^*}\right] \\
&= \begin{cases} z^X W_{i+1}{}^{-1} S_{i+1}(z) + z^X W_{i+1}{}^{-1}\left(B_{i+1,1}(z) + \ldots + B_{i+1,W_{i+1}-1}(z)\right), & \text{for } i = 0,\ldots,K-1 \\ C_{K-1}(z), & \text{for } i = K \end{cases} \\
B_{i,j}(z) &= E\left[z^{B_{i,j}^*}\right] = \begin{cases} zS_i(z), & \text{for } j = 1 \text{ and } i = 0,\ldots,K \\ zB_{i,j-1}(z), & \text{for } j = 2,\ldots,W_i - 1 \text{ and } i = 0,\ldots,K \end{cases}
\end{aligned} \qquad . \quad (4.9)$$

Let X denote the service time of a HOL packet. Since the service of each HOL packet starts from state S_0, therefore the first and second moments of service time, $E[X]$ and $E[X^2]$, can be derived from the generating function $S_0(z)$ and the expressions of them are given in Appendix C. It is confirmed by the equation (C1) that the expression of mean service time $E[X]$ is same as the expression $\tilde{b}_{suc}^{-1}$ given by (4.7). By means of the queueing model of the input buffer, we will investigate various system performance of the NP-CSMA concerning throughput and delay in the following sections.

4.3 System Performance

4.3.1 Network Throughput

The characteristic equation of network throughput for the NP-CSMA protocol with window-based binary exponential backoff scheduling algorithm can be obtained from the queueing model of input buffer presented in section 4.2. Since a node is ready to send an HOL packet only if an idle channel has been detected, the probability of successful transmission p from a desired node is conditioned on an idle channel, that is,

$$p = \frac{\Pr\{\text{none of other } n-1 \text{ nodes access the channel}\}}{\Pr\{\text{channel is sensed idle}\}}. \tag{4.10}$$

If no one access the channel, it means that all other $n-1$ nodes are either empty, or in sensing states but not accessing the channel. Thus, we have

$$
\begin{aligned}
&\Pr\{\text{none of other } n-1 \text{ nodes access the channel}\} \\
&= \left[\Pr\{\text{node is empty}\} + \Pr\{\text{node is in sensing but not scheduled to send packet}\}\right]^{n-1} \\
&= \left[(1-\rho) + \rho \sum\nolimits_{i=0}^{K} \tilde{b}_{i,J}\right]^{n-1} \\
&\overset{\text{for large n}}{=} \exp\left\{-n\rho\left(1 - \sum\nolimits_{i=0}^{K} \tilde{b}_{i,J}\right)\right\}.
\end{aligned}
\tag{4.11}
$$

The probability that the node senses an idle channel is given by

$$
\begin{aligned}
\Pr\{\text{channel is sensed idle}\} &= \left[\Pr\{\text{node is empty}\} + \Pr\{\text{node is in sensing state}\}\right]^{n-1} \\
&= \left[(1-\rho) + \rho\left(\sum\nolimits_{i=0}^{K} \tilde{b}_{i,J} + \sum\nolimits_{i=0}^{K} \tilde{b}_{S_i}\right)\right]^{n-1} \\
&\overset{\text{for large n}}{=} \exp\left\{-n\rho\left(1 - \sum\nolimits_{i=0}^{K} \tilde{b}_{i,J} - \sum\nolimits_{i=0}^{K} \tilde{b}_{S_i}\right)\right\}.
\end{aligned}
\tag{4.12}
$$

Substituting (4.7) into (4.11) and (4.12), then the probability of successful transmission p defined by (4.10) can be expressed as

$$
p = \exp\left(-\frac{a\hat{\lambda}}{\alpha p}\right). \tag{4.13}
$$

After substituting the expression of α from (2.9) into (4.13), we have

$$
p = \exp\left\{-\frac{\hat{\lambda}(1+a-p)}{p}\right\}. \tag{4.14}
$$

A network is in equilibrium when the aggregate packet input rate equals the departure rate: $\hat{\lambda} = \hat{\lambda}_{out}$. Then from (4.14), the network throughput in equilibrium is given as below,

$$
\hat{\lambda}_{out} = \frac{-p\alpha \ln p}{a} = \frac{aGe^{-aG}}{aGe^{-aG} + a + x\left(1 - e^{-aG} - aGe^{-aG}\right)}. \tag{4.15}
$$

It should be noted that the throughput given by (4.15) is the same as that derived in the theorem of chapter 2. The consistency between these approaches indicates that the backoff scheduling algorithm is independent of the network throughput.

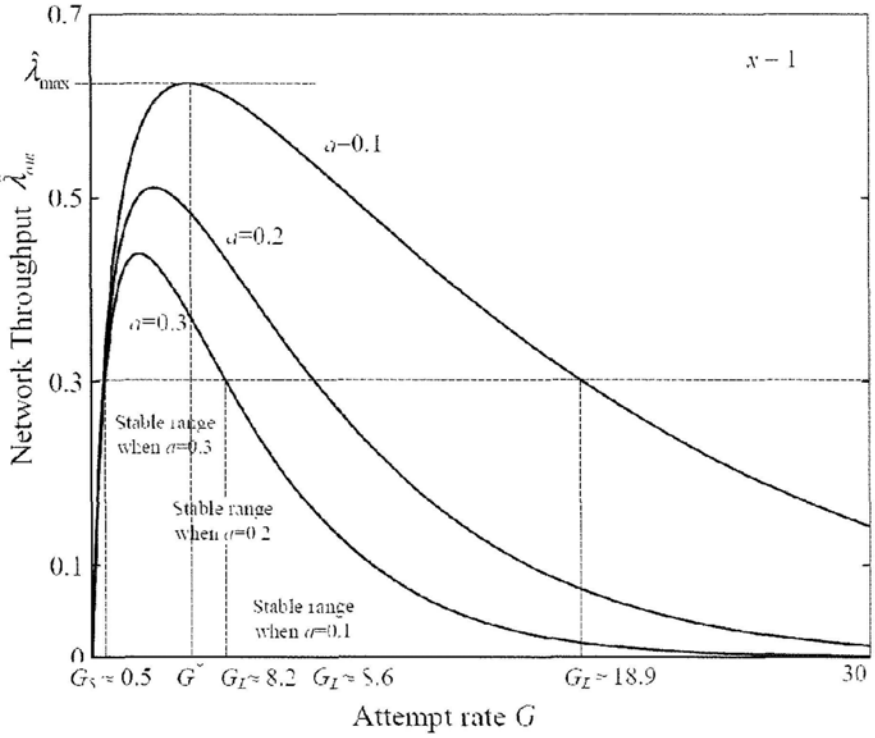


Fig. 4.2. Throughput versus the attempt rate for NP-CSMA.

As network throughput is independent of backoff algorithms, the stable throughput condition for the window-based backoff scheme is the same as that in the chapter 2. Considering the tradeoff between G and p, we know the attempt rate G should be bounded in the range between $G_S(\hat{\lambda})$ and $G_L(\hat{\lambda})$ to ensure a stable network throughput $\hat{\lambda}_{out} = \hat{\lambda}$. As plotted in Fig. 4.2, the network has a stable throughput of $\hat{\lambda}_{out} = \hat{\lambda} = 0.3$ when G is within the range [$G_S \approx 0.5$, $G_L \approx 18.9$]. We have the following stable throughput condition.

Stable Throughput Condition:

STC. *For any aggregate input rate* $\hat{\lambda} < \hat{\lambda}_{\max}$, *the attempt rate G should satisfy*

$$G_S(\hat{\lambda}) \le G \le G_L(\hat{\lambda}) . \tag{4.16}$$

We can derive the expression of the attempt rate G from the queueing model of input buffer, as below:

Considering the NP-CSMA network with n nodes, a desired HOL packet can be transmitted successfully with probability e^{-aG}, when all other nodes are inactive in the mini-slot prior to its transmission, that is, they are either in sensing or waiting states. Suppose that there are a total of $n_b = \sum_{i=0}^{K} n_i$ HOL packets in a mini-slot, in which n_i packets are in the sensing state of phase i, for $i = 0,\ldots, K$. The following HOL packets may desire to transmit during the mini-slot of a sensing state:

1) An empty node may send a newly arrived packet with probability aλ;
2) An HOL packet in sensing states will be transmitted immediately;

Hence, the attempts rate per mini-slot, aG, can be expressed as

$$aG = (n - n_b)a\lambda + \sum_{i=0}^{K} n_i . \tag{4.17}$$

Let ϕ_i be the probability that an HOL packet is in the sensing state of phase i given that the channel is idle, where for $i = 0,1,\ldots,K$. The probability ϕ_i is expressed as

$$\phi_l = \frac{\tilde{b}_{s_l}}{\sum_{k=0}^{K}\left(\tilde{b}_{s_k} + \sum_{j=1}^{W_k-1}\tilde{b}_{k,j}\right)}. \tag{4.18}$$

The following attempt rate per mini-slot can be obtained from (4.7) and (4.17):

$$\begin{aligned} aG &= n(1-\rho)a\lambda + n\rho\sum_{l=0}^{K}\phi_l \\ &= a\hat{\lambda}(1-\rho) + \frac{n\rho\sum_{l=0}^{K}\tilde{b}_{s_l}}{\sum_{l=0}^{K}\tilde{b}_{s_l} + \sum_{l=0}^{K}\sum_{j=1}^{W_l-1}\tilde{b}_{l,j}} \\ &= a\hat{\lambda}(1-\rho) + \frac{2n\rho(2p-1)}{[1-\alpha p(W-1)](2p-1) + pW - W2^K(1-p)^{K+1}}. \end{aligned} \tag{4.19}$$

For Exponential Backoff ($K \to \infty$), the expression of G can be further simplified into

$$aG = a\hat{\lambda}(1-\rho) + \frac{2n\rho(2p-1)}{(1+\alpha p)(2p-1) + pW(1-2\alpha p+\alpha)}. \tag{4.20}$$

From the above expression, we show that the attempt rate G is formed from newly arrived packet and all re-scheduled packets which are backlogged packets but ready for retransmission. It agrees with the literatures [19] – [23], the minimum contention window size W affects the system performance due to the functional relationship of the attempt rate G and W in (4.20).

4.3.2 Packet Transmission Delay

The packet transmission delay consists of two types, namely packet access delay and packet queueing delay. The access delay is the time for an HOL packet to complete its service; while the queueing delay is the time for a newly arrived packet to complete its service.

The packet access delay is simply the first moment of service time distribution $E[X]$ which is obtained from (4.9) as given in appendix. For the exponential backoff ($K \to \infty$), we have

$$E[X] = 1 + \frac{a}{2} - \frac{aW}{2} + \frac{x(1-p)}{p} + \frac{a}{2\alpha p} + \frac{aW}{2\alpha(2p-1)}. \tag{4.21}$$

Examples of the packet access delay $E[X]$ against the aggregate input rate $\hat{\lambda}$ for different value of x are plotted in Fig. 4.3 with a = 0.1. The access delay grows with the aggregate input rate and becomes unbounded if the ergodic condition is not satisfied.

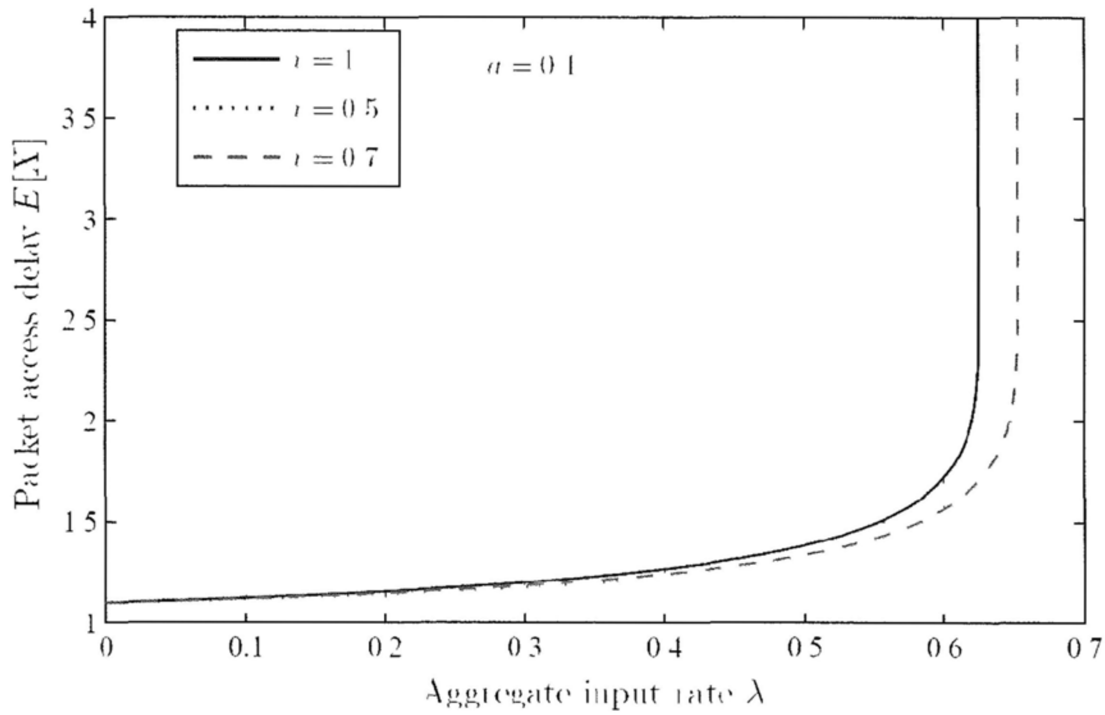


Fig 4 3 Packet access delay versus aggregate input rate for NP-CSMA

To find the packet queueing delay, we always have to consider situations that the delay is bounded. From the Pollaczek-Khinchin formula, we have the following condition.

Bounded Delay Condition:

BDC. *The Pollaczek-Khinchin formula for mean delay* $E[T]$ *of Geo/G/1 queue* [68]

$$E[T] = E[X] + \lambda \frac{E[X^2] - E[X]}{2(1 - \lambda E[X])} \tag{4.22}$$

requires bounded second moment of service time $0 < E[X^2] < \infty$.

The condition **BDC** is more restrictive than the condition **STC.** The second moment of service time of the Exponential Backoff scheme is given (C2) by in Appendix C as follows:

$$E[X^2] = B(p,W) \lim_{K \to \infty} 2^{2K} (1-p)^{K-1} + C(p,W), \tag{4.23}$$

where $B(p,W)$ and $C(p,W)$ are two polynomials given by (C3) and (C4) in Appendix C, respectively,. Thus, the retransmission factor should satisfy the following condition to guarantee a bounded second moment of service time $E[X^2]$:

$$2^2 (1-p) < 1. \tag{4.24}$$

From the above condition (4.24), it is observed that the bounded delay condition is independent of the minimum contention window size W.

With the above bounded delay condition, we can find the packet queueing delay based on the P-K formula in [68] and the expressions of first and second moments of service time distribution in appendix. For $K \to \infty$, the packet queueing delay can be reduced to

$$E[T]=E[X]+\frac{\lambda}{2(1-\lambda E[X])}\{\frac{aW(1+\alpha)}{\alpha p(2p-1)}\left[p+x(1-p)+\frac{ap}{2}+\frac{a}{2\alpha}\right]-\frac{2a^2W^2(1-\alpha p)}{3\alpha(4p-3)}-\frac{ax(1+\alpha)}{\alpha p}$$
$$-2ax-x^2\frac{1-p}{p}+\frac{1+2\alpha X}{\alpha p^2}\left[ap+ax(1-p)+\frac{a^2p}{2}+\frac{a^2}{2\alpha}\right]+\frac{a^2W^2(1+\alpha)(1+\alpha-2\alpha p)}{2\alpha^2(2p-1)(4p-3)}$$
$$+1-a^2\frac{5+\alpha p}{6\alpha p}+\frac{a(a+x2\alpha)W(1+\alpha-2\alpha p)}{2\alpha^2(2p-1)^2}-\frac{aW\left[2x(1+\alpha-2\alpha p)+4\alpha p+3a+a\alpha\right]}{2\alpha(2p-1)}\}. \tag{4.25}$$

Fig. 4.4 plots the packet queueing delay against the aggregate input rate for different values of collision duration x. In equilibrium (that is, network throughput equals aggregate input rate), the smaller values of x we choose, the larger the maximum throughput it supports.

In chapter 2, we have shown that the local maximum throughput in bounded delay region is smaller than the global maximum throughput. Here we further demonstrate this fact by comparing the packet access and queueing delay. From the ergodic condition, bounded access delay requires the probability of successful transmission $p > 0.5$. Based on the condition (4.24), the required probability of successful transmission p for bounded queueing delay is $p > 0.75$ which is stronger than the counterpart. As illustrated in Fig. 4.3 and Fig. 4.4, the maximum achievable throughput with bounded queueing delay is smaller than that with bounded access delay (bounded service time).

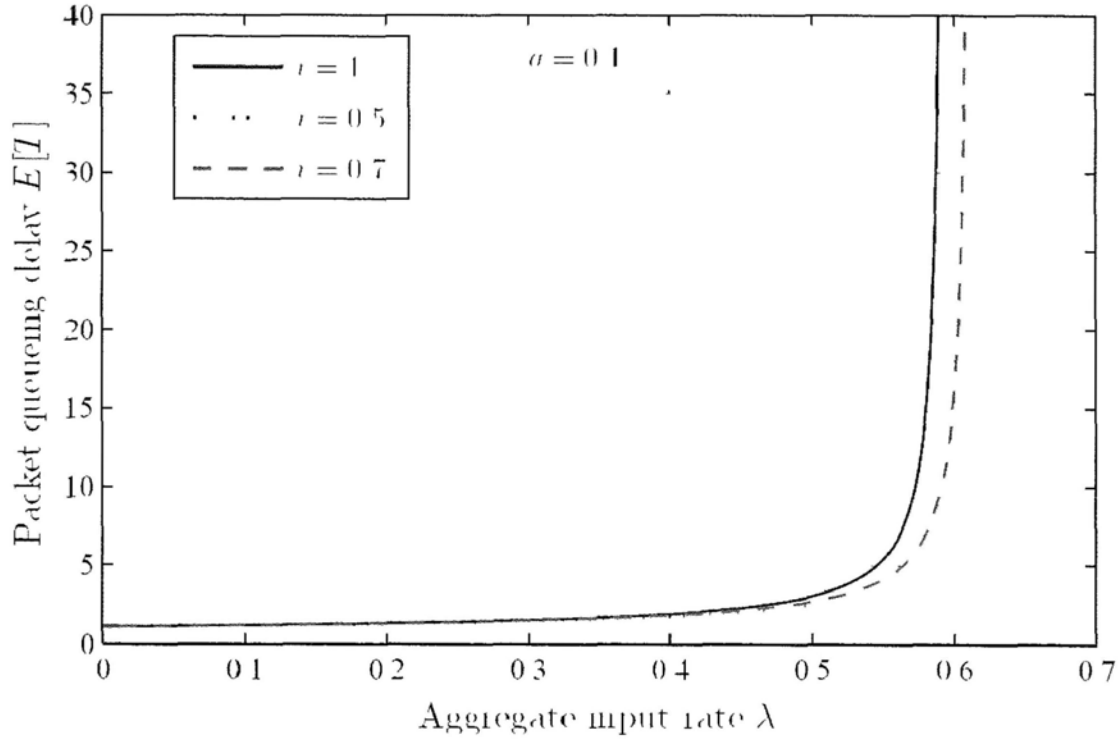


Fig 4 4 Packet queueing delay versus aggregate input rate for NP-CSMA

4.4 Similarity between Probability-based and Window-based backoff scheduling algorithms

The equivalence between probability-based and window-based backoff scheduling algorithms is summarized in TABLE 4.2. For the probability-based algorithm in chapter 2, the retransmission probability is represented by a factor q When q =0.5, the backoff algorithm is exactly the binary exponential backoff algorithm. In this case, the ergodic and bounded delay conditions of both algorithms are the same as in TABLE 4.2. Since the network throughput is invariant to the backoff algorithms, the two algorithms have the same network throughput as well as the stable throughput condition. More the constraint of bounded queueing delay is stronger than that of bounded service time; therefore, the bounded delay region is a subset of stable throughput region.

	Window-based	**Probability-based**
Ergodic condition	$1-p<1/2$	$1-p<q$
Stable throughput condition	$G_S(\hat{\lambda}) \le G \le G_I(\hat{\lambda})$	$G_S(\hat{\lambda}) \le G \le G_L(\hat{\lambda})$
Bounded delay condition	$1-p<1/4$	$1-p<q^2$
Network throughput	$\dfrac{aGe^{-aG}}{aGe^{-aG}+a+x\left(1-e^{-aG}-aGe^{-aG}\right)}$	$\dfrac{aGe^{-aG}}{aGe^{-aG}+a+x\left(1-e^{-aG}-aGe^{-aG}\right)}$

TABLE 4.2. Properties of window-based and probability-based algorithms.

Chapter 5

Summary and Future works

The first part in this chapter is a short summary which describes our achievements in single-hop networks. We believe that the proposed methodology can be extended to multi-hop networks, which is exactly our future works as given in section 5.2.

5.1 Summary

This thesis mainly concentrates on evaluating the performance of media access control (MAC) protocols. The MAC protocol plays an important role in wireless communication systems, but a comprehensive performance analysis of MAC protocol remains an open issue. We show that backoff scheduling algorithms and input traffic characteristics are key ideas in studying MAC protocol. With consideration of these issues, we develop a generic queueing model to completely analyze the performance of MAC protocols. The model is well demonstrated by focusing on unsaturated non-persistent and 1-persistent CSMA protocols with exponential backoff scheduling algorithm.

By modeling the service time distribution of each individual head-of-line (HOL) packet as a Markov chain, we obtain the characteristic equation of network throughput, the packet transmission delay as well as the stable throughput and bounded delay conditions with admissible input traffic.

Based on the stable conditions, we specify the stable throughput and bounded delay regions with respect to the retransmission factor and input rate. In particular, the bounded delay region is only a subset of the stable throughput region. The achievable maximum throughput inside the bounded delay region is slightly smaller than the absolute maximum throughput. For networks operating inside the stable throughput region but outside the bounded delay region, the variance of service time is very large due to the capture effect. With the stable regions, the network performance can be substantially improved by properly selecting the value of retransmission factor.

A consensus on the stability issue of the exponential backoff scheme cannot be concluded from controversial results reported in previous studies. We proved that the exponential backoff scheme is stable under some specific conditions even for infinite populations. While other algorithms with finite backoff phase are intrinsically instable. We also study the binary exponential backoff algorithm in window-based manner. The stable throughput and bounded delay conditions of the window-based scheme is equivalent to that of the probability-based scheme.

Last but not least, this queueing model can be systematically generalized to investigate various other types of MAC protocols, such as the 802.11 DCF and 802.15.4 protocol. We have already extended our work to the 802.11 DCF protocol in [*], which is very popular in the field of wireless communications. To sum up, the proposed model is useful to design a more robust MAC protocols for future development.

5.2 Future Works

In this thesis, we only focus on the analysis of single-hop network. Extending the works to multi-hop networks is a challenging task. Ad hoc Networks which is a type of multi-hop networks, it can be applied in various areas, such as wireless sensor networks (WSN), mobile ad hoc networks (MANET), VANETs (Vehicular Ad hoc Networks) and etc.

The ad hoc network is decentralized, infrastructure-less and highly dynamic, in which users can join the network anytime anywhere. Although the decentralized property makes ad hoc network attractive, it is not feasible for far-distance communications due to power limitation. Since users can join the ad hoc network anytime anywhere; it is rather difficult to guarantee system performance in this unpredictably changing topology. To tackle these problems, most of the present researches are focused on developing efficient routing algorithms for ad hoc networks, for instance, the dynamic source routing (DSR) and ad hoc on-demand distance-vector (AODV) protocols [72]. Despite that abundant routing algorithms have been proposed, the efficiency of these algorithms is still unknown due to the lack of a suitable methodology in analyzing the system performance.

For performance analysis, researchers so far played little attention to devote to this issue. It takes several decades to tackle the single-hop problem, but the problem in multi-hop ad hoc networks is much more complicated. There are plenty studies of multi-hop networks available [73] – [78], but none of them can give a complete story of multi-hop.

To achieve better performance in ad hoc networks, there is a need for a comprehensive analysis of the system. For example, it is essential to decide how to allocate channel resources to achieve better throughput; likewise, the system should be able to help schedule connections between users; but at the same time, it should also guarantee the delay of each user. These concerns may look similar to those encountered in single-hop networks, but they are not. Indeed, the size of the network, the infeasible far-distance communications, the unpredictably changing topology, and the unavoidably slow deployment make the problem very novel and challenging.

Our future work is to develop a model for performance analysis of multi-hop ad hoc and mesh networks. Although this is an open problem for several decades, we believe that it is feasible to develop a generic model based on our previous works in the field.

Based on this model, we expect to generate the following results: We will complete a comprehensive performance analysis of multi-hop networks, which includes the characteristic equation of network throughput, link capacity, mean number of travelled hops (packet transmission delay) and mean packet queueing delay. Moreover, the model can also be generalized to various other types of multi-hop networks, such as the WSB and MANET networks. A comprehensive analysis can provide us not only the useful information of implementing existing routing algorithms, but also shed some light on the deployment of future networks.

APPENDIX A. SERVICE TIME DISTRIBUTION FOR NP-CSMA PROTOCOL

For the Exponential Backoff scheme, the actual first moment $E[X]$ and second moment $E[X^2]$ of service time of HOL packets can be derived from (2.25) as follows:

$$E[X]=\frac{aq\left(M+1-M\alpha\right)}{\alpha\left(p+q-1\right)}+\frac{aX+ap\left(M-X\right)}{p}. \tag{A26}$$

$$E[X^2]=B\left(p,q\right)\lim_{K\to\infty}\left(\frac{1-p}{q^2}\right)^{K-1}+C\left(p,q\right), \tag{A27}$$

where

$$B(p,q)=\frac{2a^2\left(1-p\right)\left(M+1-\alpha M\right)^2}{\alpha^2 pq}\left(\frac{1}{pq}-\frac{1-p+q}{p+q^2-1}\right) \tag{A28}$$

and

$$\begin{aligned} C(p,q)=a^2[&\frac{p\left(M+1\right)\left(M-2\right)+\left(1-p\right)\left(X+1\right)\left(X-2\right)+2}{p}-\frac{\alpha qM}{\alpha\left(p+q-1\right)}\\ &+\frac{2q^3\left(M+1-\alpha M\right)^2}{\alpha^2\left(p+q^2-1\right)\left(p+q-1\right)}+\frac{q\left(M+1-\alpha M\right)\left(3pM+2X-2pX-2p\right)}{\alpha p\left(p+q-1\right)}\\ &+\frac{2q\left(1-p\right)X\left(M+1-\alpha M\right)}{\alpha\left(p+q-1\right)^2}+\frac{2\left(1-p\right)X\left(pM+X-pX\right)}{p^2}]+E\left[X\right]. \end{aligned} \tag{A29}$$

APPENDIX B. SERVICE TIME DISTRIBUTION FOR NP-CSMA PROTOCOL

For the Exponential Backoff scheme, the actual first moment $E[X]$ and second moment $E[X^2]$ of service time of HOL packets can be derived from (3.32) as follows:

$$E[X] = a\left[M - X + \frac{X}{p} + \frac{q(M+1-\alpha M)}{(p+q-1)}\right]. \tag{B1}$$

$$E[X^2] = B(p,q)\lim_{K\to\infty}\left(\frac{1-p}{q^2}\right)^{K-1} + C(p,q), \tag{B2}$$

where

$$B(p,q) = 2a^2(M+1-\alpha M)^2\left[\frac{(1-p)}{p^2q^2} - \frac{q^2}{(p+q^2-1)} - \frac{(1-p)^2(1+q)}{qp(p+q^2-1)}\right] \tag{B3}$$

and

$$\begin{aligned}C(p,q) = {} & E[X] + a^2\{(2X-M)(M+1-\alpha M) + \alpha M^2 - (X+1)X \\ & + \frac{(X-1)X + 2(1-\alpha)p_2(M-X)M}{p} + \frac{2q^3(M+1-\alpha M)^2}{(p+q^2-1)(p+q-1)} \\ & + q\frac{(-2p+2\alpha pM+pM-2pX+2X)(M+1-\alpha M)-\alpha pM}{p(p+q-1)} \\ & + 2\left[X + p(M-X-\alpha M) - (1-\alpha)p_2M\right]\left[\frac{pM-pX+X}{p^2} + \frac{q(M+1-\alpha M)}{(p+q-1)^2}\right]\}.\end{aligned} \tag{B4}$$

APPENDIX C. SERVICE TIME DISTRIBUTION FOR NP-CSMA PROTOCOL WITH WINDOW-BASED BINARY EXPONENTIAL BACKOFF

For the Exponential Backoff scheme, the actual first moment $E[X]$ and second moment $E[X^2]$ of service time of HOL packets can be derived from (6.15) as follows:

$$E[X] = a[M + \frac{X(1-p)}{p} + \frac{1}{2} - \frac{W}{2} + \frac{1}{2\alpha p} + \frac{W}{2\alpha(2p-1)} - \frac{W2^K(1-p)^{K+1}}{2\alpha p(2p-1)}] . \quad \text{(C1)}$$

$$E[X^2] = B(p,W)\lim_{K\to\infty} 2^{2K}(1-p)^K + C(p,W), \quad \text{(C2)}$$

where

$$B(p,W) = \frac{a^2W^2(1+\alpha)(2p-3+\alpha p)}{2\alpha^2 p(4p-3)} + \frac{a^2W^2(3-\alpha p)(1-\alpha p)}{6\alpha^2 p^2} \quad \text{(C3)}$$

and

$$\begin{aligned} C(p,q) = a^2[M^2 &- \frac{1-p}{p}X^2 - 2MX - \frac{(1+\alpha)X}{\alpha p} - \frac{5+\alpha p}{6\alpha p} - \frac{2W^2(1-\alpha p)}{3\alpha(4p-3)} \\ &+ \frac{1+2\alpha X}{\alpha p^2}\left[pM + (1-p)X + \frac{p}{2} + \frac{1}{2\alpha}\right] + \frac{W^2(1+\alpha)(1+\alpha-2\alpha p)}{2\alpha^2(2p-1)(4p-3)} \\ &+ \frac{(1+2\alpha X)W(1+\alpha-2\alpha p)}{2\alpha^2(2p-1)^2} + \frac{W(1+\alpha)}{\alpha p(2p-1)}\left[pM + (1-p)X + \frac{p}{2} + \frac{1}{2\alpha}\right] \\ &- \frac{W\left[2(1+\alpha-2\alpha p)X + 4\alpha pM + 3 + \alpha\right]}{2\alpha(2p-1)}] + E[X]. \end{aligned} \quad \text{(C4)}$$

Bibliography

[1] E. Cinlar and R. A. Agnew, "On the Superposition of Point Processes", *Journal of the Royal Statistical Society,* Series B, vol. 30, No. 3, pp. 576-581, 1968.

[2] D. J. Daley and D. Vere-Jones, An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods, *Springer publisher*, 2003.

[3] J. Y. Hui, *Switching and Traffic Theory for Integrated Broadband Networks.* Kluwer Academic Publishers, 1990.

[4] T. T. Lee and L. Dai, "Stability and Throughput of Buffered Aloha with Backoff," *Technical Report*, 2008. http://arxiv.org/pdf/0804.3486.

[5] N. Abramson, "The Aloha System – Another Alternative for Computer Communication," *Proc. Fall Joint Compet. Conf.*, AFIP Conference, vol. 44, pp. 281-285, 1970.

[6] L. Kleinrock and F.A. Tobagi, "Packet Switching in Radio Channels: Part 1-Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics," *IEEE Trans. Commun.*, vol. COM-23, pp. 1400-1416, 1975.

[7] H. Takagi and L. Kleinrock, "Throughput analysis for persistent CSMA systems," *IEEE Trans. Commun.*, vol. 33, no. 7, pp. 627-638, Jul. 1985.

[8] R. Rom and M. Sidi, *Multiple Access Protocols – Performance and analysis*. Springer-Verlag New York, June 1990.

[9] Y. Tay and K. Chua, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wireless Networks,* vol. 7, no. 2, pp. 159-171, March, 2001.

[10] Galtier, J., "Analysis of the Slotted Non-persistent CSMA Protocol with Poissonian Packet Size Using a Semi-Markov Graph Representation", *ICTON.*, pp. 258 – 262, 2006.

[11] A. Sugihara, K. Enomoto and I. Sasase, "Throughput Performance of a slotted nonpersistent CSMA with an Adaptive Array," *Sixth IEEE International Symposium on*, vol. 2, pp. 633-637, 1995.

[12] R. Bruno, M. Conti, E. Gregori, "Optimization of efficiency and energy consumption in p-persistent CSMA-based wireless LANs," *IEEE Transactions on Mobile Computing*, vol.1, no.1, pp. 10-31, 2002.

[13] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent packet traffic," *IEEE/ACM Trans. Networking*, vol. 4, pp. 209–223, Apr. 1996.

[14] X. Tan, S.C. Jue, E. Lo and R.H.S. Hardy, "Transient performance of a CSMA system under temporary overload conditions," *Local Computer Networks 15th Conf. on*, pp. 93-101, 1990.

[15] L. Jereb and M.A. Marsan, "Transient performance analysis of 1-persistent CSMA and CSMA-CD protocols," *IEEE/INFCOM*, pp. 246-252, 1990.

[16] T.S. Ho and K.C. Chen, "Performance analysys of IEEE 802.11 CSMA/CA medium access control protocol", *in Proc. IEEE PIMRC*, pp.407 - 411, 1996.

[17] H. Chhaya and S. Gupta, "Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocols," *Wireless Networks*, vol. 3, pp. 217–234, 1997.

[18] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, *IEEE standards 802.11*, January 1997.

[19] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Funtion," *IEEE Journal of selected areas in COMMUN.*, Vol. 18, No. 3, pp. 535-547, 2000.

[20] P. Chatzimisios, V. Vitsas, and A.C. Boucouvalas, "Throughput and delay analysis of IEEE 802.11 protocol," *in Proc. 5th IEEE Workshop Networked Appliances*, Oct. 2003, pp. 168–174.

[21] Z. Hadzi-Velkov and B. Spasenovski, "Saturation throughput: delay analysis of IEEE 802.11 DCF in fading channel," *in Proceedings of IEEE ICC2003*, May, 2003.

[22] 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," *in Proceedings of IEEE INFOCOM2002*, June 2002.

[23] H. Ma, X. Li; H. Li, P. Zhang, S. Luo, and C. Yuan, "Dynamic optimization of IEEE 802.11 CSMA/CA based on the number of competing stations," *in Proc. IEEE International Conf. on Commun.*, vol.1, pp. 191- 195, June 2004.

[24] Y. Yang and T.S.P. Yum, "Delay Distributions of Slotted ALOHA and CSMA," *IEEE Trans. Commun.*, vol. 51, No. 11, pp.1846-1875, 2003.

[25] P. Raptis, V. Vitsas, K. Paparrizos, P. Chatzimisios, A. C. Boucouvalas, and P. Adamidis, "Packet delay modeling of IEEE 802.11 Wireless LANs," *in Proc. Intl. Conf. on Cyber. Info. Tech. Sys. Apps. CITSA 2005*, July 2005.

[26] P. Raptis, V. Vitsas, K. Paparrizos, P. Chatzimisios, and A. C. Boucouvalas, "Packet delay distribution of the IEEE 802.11 distributed coordination function," *Sixth IEEE International Symposium on World of Wireless Mobile and Multimedia Networks (WoWMoM) 2005,* pp. 299- 304, 13-16 June 2005.

[27] E. Ziouva and T. Antonakopoulos, "CSMA/CA performance under high traffic conditions: Throughput and delay analysis", *Computer Commun.*, vol. 25, pp.313-321, 2002.

[28] C.H. Foh and M. Zukerman, "Performance analysis of the IEEE802.11 MAC protocol," *in Proceedings of the European Wireless 2002 Conference*, pp. 184-190, Florence, Italty, Feb. 2002.

[29] M.M. Carvalho and J.J. Garcia-Luna-Aceves, "Delay analysis of IEEE 802.11 in single-hop networks," *in Proceedings of the 11^{th} IEEE International Conference on Network Protocols,* p. 146, Nov. 2003.

[30] T. Sakurai and H.L. Vu, "MAC access delay of IEEE 802.11 DCF," *IEEE Trans. Wireless Communications*, vol. 6, no. 5, pp. 1702–1710, May 2007.

[31] H. Wu, S. Cheng, Y. Peng, K. Long, and J. Ma, "IEEE 802.11 distributed coordination function (DCF): analysis and enhancement," *in Proceedings of IEEE ICC2002*, pp. 605–609, 2002.

[32] C. Wang, B. Li, and L. Li, "A New Collision Resolution Mechanism to Enhance the Performance of IEEE 802.11 DCF," *IEEE Trans. Vehicular Technolgy,* vol. 53, no. 4, pp. 1235-1246, 2004.

[33] T.R. Park, T.H. Kim, J.Y. Choi, S. Choi, and W.H. Kwon, "Throughput and energy consumption analysis of IEEE 802.15.4 slotted CSMA/CA," *Electronics Letters*, vol.41, no.18, pp. 1017-1019, Sept. 2005.

[34] L.C. Wang, S.Y. Huang, and A. Chen, "On the Throughput Performance of CSMA-based Wireless Local Area Network with Directional Antennas and Capture Effect: A Cross-layer Analytical Approach," *IEEE Communications Society*, pp.1879-1884, 2004.

[35] L.C. Wang, S.Y. Huang, and A. Chen, "On the throughput performance of CSMA-based wireless local area network with directional antennas and capture

effect: a cross-layer analytical approach," in Proc. IEEE Wireless Commun. Networking Conf. (WCNC)., vol.3, pp. 1879- 1884, March 2004.

[36] F. Babich, M. Comisso, "Throughput and Delay Analysis of 802.11-Based Wireless Networks Using Smart and Directional Antennas," *IEEE Trans. on Commun.*, vol.57, no.5, pp.1413-1423, May 2009.

[37] J. He, D. Kaleshi, A. Munro, Y. Wang, A. Doufexi, J. McGeehan, and Z. Fan, "Performance investigation of IEEE 802.11 MAC in multihop wireless networks," *In MSWiM*, 2005.

[38] M. Ergen and P. Varaiya, "Throughput analysis and admission control in IEEE 802.11a," Mobile Networks and Applications, vol. 10, no. 5, pp. 705–706, Oct. 2005.

[39] K. Duffy, D. Malone, and D.J. Leith, "Modeling the 802.11 distributed coordination function in non-saturated conditions," *IEEE Commun. Lett.*, vol 9, pp. 715, 2005.

[40] D. Malone, K. Duffy, and D.J. Leith, "Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions," *IEEE/ACM Transactions on Networking (TON)*, v.15 n.1, p.159-172, February 2007.

[41] G. R. Cantieni, Q. Ni, C. Barakat, and T. Turletti, "Performance analysis under finite load and improvements for multirate 802.11," *Comput.Commun.*, vol. 28, no. 10, pp. 1095–1109, Jun. 2005.

[42] S. Pollin, M. Ergen, S.C. Ergen, B. Bougard, L. Van der Perre, F. Catthoor, I. Moerman, A. Bahai, and P. Varaiya, "Performance analysis of slotted carrier sense IEEE 802.15.4 medium access layer," *in proc. IEEE GLOBECOM*, 2006.

[43] S. Pollin, M. Ergen, S. Ergen, B. Bougard, L. Der Perre, I. Moerman, A. Bahai, P. Varaiya, and F. Catthoor, "Performance Analysis of Slotted Carrier Sense IEEE 802.15.4 Medium Access Layer," *IEEE Transactions on Wireless Commun.*, vol.7, no.9, pp.3359-3371, September 2008.

[44] I. Ramachandran , A.K. Das , S. Roy, "Analysis of the contention access period of IEEE 802.15.4 MAC," *ACM Transactions on Sensor Networks (TOSN)*, v.3 n.1, p.4-es, March 2007.

[45] P. Park, P. Di Marco, P. Soldati, C. Fischione, and K.H. Johansson, "A generalized Markov chain model for effective analysis of slotted IEEE 802.15.4," *IEEE 6th International Conference on Mobile Adhoc and Sensor Systems, 2009. MASS '09*, pp.130-139, Oct. 2009.

[46] 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, February 2003.

[47] O. Tickoo and B. Sikdar, "Queueing Analysis and Delay Mitigation in IEEE 802.11 Random Access MAC based Wireless Network," *in Proc. IEEE INFOCOM Conf.*, Mar. 2004.

[48] O. Tickoo and B. Sikdar, "A Queueing Model for Finite Load IEEE 802.11 Random Access MAC," *IEEE Commun. Society*, pp. 175-179, 2004.

[49] M.E. Rivero-Angeles, D. Lara-Rodriguez, and F.A. Cruz-Perez, "Differentiated Backoff Strategies for Prioritized Random Access Delay in Multiservice Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 381-397, Jan. 2009.

[50] A. Nasipuri, J. Zhuang and S.R. Das, "A Multichannel CSMA MAC Protocol for Multihop Wireless Networks," *in Proc. of IEEE Wireless Commun. and Networking Conf. (WCNC)*, Sep. 1999.

[51] G. Bianchi, L. Fratta, and M. Olivefi, "Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LANs," *in Proc. PIMRC*, pp. 392-3.96, Oct. 1996.

[52] X.M. Chen and G.S. Hong, "A simulation of Study of the Predictive *p*-persistent CSMA Protocol," *proceeding of the 35th annual Sim. Sym.*, 2002.

[53] A. Nasipuri, J. Zhuang and S.R. Das, "A multichannel CSMA MAC protocol for multihop wireless networks," *wireless commun and Netw. Conf.*, pp. 1402-1406, 1999.

[54] S. Choi, J. del Prado, S. Nandgopalan and S. Mangold, "IEEE 802.11e contention-based channel access (EDCF) performance evaluation," *in Proc. IEEE ICC,* Anchorage, Ak, May 2003.

[55] A. Koubaa, M. Alves, and E. Tovar, "A comprehensive simulation study of slotted CSMA/CA for IEEE 802.15.4 wireless sensor networks," *IEEE International Workshop on Factory Commun. Systems*, pp.183-192, 2006.

[56] D. J. Aldous, "Ultimate instability of exponential backoff- protocol for acknowledgment-based transmission control of random access communication channels," *IEEE Trans. Inf. Theory 33 (2)*, pp. 219-223, 1987.

[57] J. Goodman, A. G. Greenberg, N. Madras, and P. March, "Stability of binary exponential backoff," *J. ACM 35 (3)*, pp. 579-602, 1988.

[58] N.O. Song, B. J. Kwak and L. E. Miller, "On the Stability of Exponential Backoff," *Journal of Research of the National Institute of Standards and Technology*, vol. 108, pp. 289-297, 2003.

[59] B-J Kwok, N-O Song and L. E. Miller, "Analysis of the stability and performance of exponential backoff," *IEEE WCNC*, vol.3, pp. 1754 – 1759, 2003

[60] B-J Kwok, N-O Song and L. E. Miller, "Performance analysis of exponential backoff," *IEEE/ACM Trans. Networking*, pp. 343-335, April 2005.

[61] K. Medepalli and F. A. Tobagi, "On optimization of CSMA/CA based wireless LANs: Part I – impact of Exponential Backoff," Proc. IEEE Globecom, 2006.

[62] I. Vukovic and N. Smavatkul, "Delay Analysis of Different Backoff Algorithms in IEEE 802.11," *in Proc. IEEE Vehicular Technology Conference (VTC)*, Los Angeles CA, September 2004.

[63] D. Xu, T. Sakurai, and H. L. Vu, "An analysis of different backoff functions for an IEEE 802.11 WLAN,"*in Proceedings of VTC2008*, 2008

[64] S. Mehta, and and K. S. Kwak, "Performance Analysis of Binary Exponential Backoff and Improved Backoff,"*EURASIP Journal on Wireless Communications and Networking,* 2010

[65] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 protocol: Design and performance evaluation of an adaptive backoff mechanism," *IEEE J. Select.Areas Commun.*, vol. 18, pp. 1774–1786, Sept. 2000.

[66] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, pp. 329–359, 1996.

[67] K. K. Ramakrishnan and H. Yang, "The Ethernet capture effect: analysis and solution," *Proc. 19th Local Computer Networks Conf.*, Oct. 1994.

[68] Hideaki Takagi, *Queueing Analysis, A Foundation of Performance Evaluation, Volume 3: Discrete-Time Systems*, NORTH-HOLLAND, 1993.

[69] Pui King Wong, Dongjie Yin, and Tony T. Lee, "Analysis of Non-Persistent CSMA Protocols with Exponential Backoff Scheduling," *Technical Report*, 2010. http://arxiv.org/pdf/1005.0178.

[70] Pui King Wong, Dongjie Yin, and Tony T. Lee, "Performance Analysis of Markov Modulated 1-Persistent CSMA/CA Protocols with Exponential Backoff Scheduling," Technical Report, 2010. http://arxiv.org/pdf/1008.1628.

[71] Dongjie Yin, Pui King Wong, and Tony T. Lee, "Generalized Bianchi's Model of IEEE 802.11 Distributed Coordination Function," *Technical Report*, 2011. http://arxiv.org/pdf/1103.2212.

[72] Charles E. Perkins, *Ad hoc networking: an introduction*, Ad hoc networking, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 2001.

[73] G. Ferrari, and O.K. Tonguz, "Performance of ad hoc wireless networks with Aloha and PR-CSMA MAC protocol," *IEEE Global Telecommunications Conference (GLOBECOM) 2003*, vol.5, pp. 2824- 2829, Dec. 2003.

[74] X. Wang and K. Kar, "Throughput modelling and fairness issues in CSMA/CA based ad hoc networks," *in Proc. IEEE INFOCOM*, vol.1, pp. 23- 34, March 2005.

[75] S. Biswas and R. Morris, "Opportunistic routing in multi-hop wireless networks," ACM SIGCOMM computer communication review, vol. 34, issue 1, Jan. 2004.

[76] M. Gerla, R. Bagrodia, Lixia Zhang; K. Tang, Lan Wang, "TCP over wireless multi-hop protocols: simulation and experiments ," *IEEE International Conference on Communications, 1999*, vol.2, pp.1089-1094, 1999.

[77] Y. Liu, R. Sheahan, H.P. Schwefel, "Towards analytic modeling for CSMA/CA based MAC protocol in wireless multi-hop networks - A simulation study," the *13th GI/ITG Conference Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB 2006),* vol. 1-16, pp. 27-29, March 2006.

[78] Lei Cao, Jinjing Tao, Hang Shen, Guangwei Bai, "Simulation Study of TCP Performance over Multi-Hop Ad Hoc Networks," *the 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM 2010)*, vol. 1-4, pp. 23-25, Sept. 2010.