Performance Modeling and Design Trade-offs of Wireless Communication Networks with Heterogeneous Service Requirements

Te-Kai Liu

CENG-96-15

Department of Electrical Engineering - Systems
University of Southern
Los Angeles, California 90089-2562
(213) 740-4579

August 1996

Performance Modeling and Design Trade-offs of Wireless Communication Networks with Heterogeneous Service Requirements

by

Te-Kai Liu

A Dissertation Presented to the

FACULTY OF THE GRADUATE SCHOOL

UNIVERSITY OF SOUTHERN CALIFORNIA

In Partial Fulfillment of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY

(Computer Engineering)

August 1996

UNIVERSITY OF SOUTHERN CALIFORNIA
THE GRADUATE SCHOOL
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA 90007

This dissertation, written by

Te-Kai Liu

under the direction of his Dissertation Committee, and approved by all its members, has been presented to and accepted by The Graduate School, in partial fulfillment of requirements for the degree of

DOCTOR OF PHILOSOPHY

Alfred C. Ladner

Dean of Graduate Studies

Date June 10, 1996

DISSERTATION COMMITTEE

Chairperson

To my parents and my wife,
for their love, encouragement and support.

Acknowledgements

I would like to express my sincere appreciation to Professor John A. Silvester, my dissertation chairman, for his valuable advice and financial support throughout the course of this work. I am also grateful to Professor Andreas Polydoros for his insightful comments, encouragement, and research assistantship during the first few years of my Ph.D. study at USC. I would like to thank the other members of my committee: Professors Robert Scholtz, Richard Arratia, and Sandeep Gupta.

Many thanks to my friends and colleagues in the Electrical Engineering Department at USC for their friendships and various discussions, among whom I mention Dr. Kuo-Chun Lee, Dr. Albert Long, Dr. Nelson Fonseca, Dr. Alex Sun, Dr. Stanley Wang, Dr. Gary Yang, Achilleas Anastasopoulos, Gent Paparisto, Prokopios Panagioutou, Moe Win, Jieh-Chian Wu, Gilberto Mayor, Tien-Chien Yu, Jeen Hasik, and Jinsung Choi. Sincere thanks also go to Bill Bates, Diane Demetras, Milly Montenegro, Regina Morton, and Mary Zittercob for their administrative help.

Finally, I must thank my parents and my dear wife Wan-Ching. This work could never be completed without their love and support.

Contents

List of Figures

List of Tables

Abstract

In this dissertation we study several wireless communication networks with different service requirements. Our approach is to first develop performance models, either analytical or simulation, for the wireless communication network of interest and then investigate the trade-offs between the system performance and various network control schemes.

For mobile data networks employing the slotted ALOHA protocol, we present an analytical model that can be used to investigate the fairness issue which arises due to the near-far effect, taking into account the effect of the physical layer parameters such as channels, modulation and coding. The performance of two retransmission control policies are evaluated and compared. The impact of the propagation effects of mobile radio channels on the single-user and network performance is studied by a simulation model where four fading channels are implemented and their performance is compared.

For distributed packet radio networks with hard real-time requirements, we define a system reliability measure and analyze the performance of the reservation ALOHA protocol. Both the worst-case scenario and the steady-state scenario are studied by Markovian models and simulation. Various trade-offs between the system performance and the system parameters such as channel data rate, packet error rate, etc. are investigated.

For wireless CDMA networks supporting integrated voice/video/data services, we propose a set of interface parameters that summarize the effect of physical layer parameters and develop markovian models for the integrated traffic to study the performance of various congestion control policies. The performance under different types of feedback information and the effects of various system parameters is investigated by the proposed model. The congestion control policies allow data traffic to fully utilize the channel capacity while guaranteeing the quality of service of voice and video.

Chapter 1

Introduction

1.1 Wireless Communication Networks

Wireless communications networks have been attracting much attention because of the flexibility and versatility of the services that can be provided. The research and development on wireless communications networks have been active for several decades since the Defense Advanced Research Projects Agency (DARPA) initiated the research on packet radio networks [24]. The design objective of DARPA's packet radio network was to provide an efficient means of data communications for mobile users in a dynamic ground radio environment. After the successful deployment of DARPA's packet radio networks, many other radio networks have been researched and designed to operate in diverse environments such as satellite [3], [35], conventional macrocellular [48], upcoming microcellular [39], and indoor wireless [56] environments. These networks are similar in that they all apply packet switching technology and are designed to operate in wireless environments. Nevertheless, they are quite different from the viewpoint of 1) the characteristics of radio channels of the environments, 2) the services/applications that are

to be supported, e.g., personal communication systems (PCS) [11], [31], intelligent vehicle highway systems (IVHS) [46], [51], wireless data systems [41], and 3) the detailed network architecture such as signaling format, access control, error control, network topology, routing control, flow control, etc. For reviews on some of the key issues, see [55].

In general, the design and performance evaluation of wireless communications networks need to consider the characteristics of radio channels in the given environment, the services to be supported, and the network architecture. The relationships between wireless environment, service requirements, and network architecture are illustrated in Fig. 1.1. Basically, the design of the network architecture (including the physical layer and the access layer) is to satisfy the service requirements subject to 1) the channel impairment due to the physical environment, 2) the interference due to co-channel interference and multi-user interference, and 3) design constraints such as radio spectrum bandwidth and hardware complexity.

Wireless Environments and Channel Models

One of the challenging aspect of wireless communications networks is the dynamic nature of the mobile radio channel [42], [44] in wireless environments. The signals in mobile radio channels are subject to propagation path loss and multipath fading. Propagation path loss refers to the phenomenon that the mean strength of radio signals decreases as the distance between transmitters and receivers increases due to free space loss and groundwave loss. The mean strength of the received signal will vary if transmitters, receivers or both are mobile. Multipath fading is the phenomenon that the strength of the

received signal fluctuates rapidly in time. It occurs because the received signal is the superposition of a large number of scattered and reflected waves due to the surrounding trees, buildings and

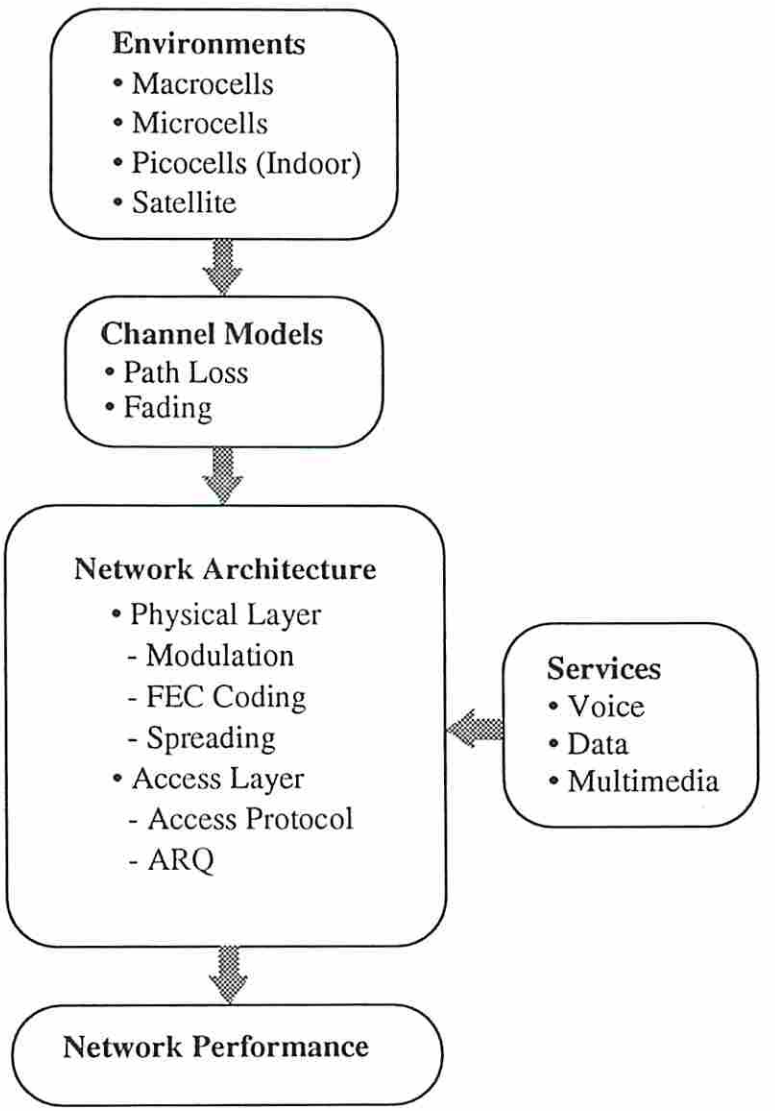


Fig. 1.1. The relationships between environments, services, network architecture and performance.

roads. The strength of the sum of the signals from different paths has fast variations because of the random path delays, which manifest themselves as random phases of the multipath signals.

Though all subject to propagation path loss and multipath fading, the channels in different wireless environments could have very different statistical characteristics. There has been a lot of interest in channel measurement and simulation for various wireless environment, e.g., macrocells [12], [19], [59], microcells [19], [50], [60], and indoor picocells [9], [48], [58]. The channel characteristics of the wireless environment impact the communication system through the physical layer. Good channel models are crucial for the design of physical-layer countermeasures, such as coded modulation [66], FEC coding [2], [59], equalization [63], [67] and diversity [31], [54] schemes.

Service Requirements

One of the reason for the fast proliferation of wireless communication networks can be attributed to the ease and flexibility in network deployment, as well as the versatility of services that can be provided, e.g., data, voice, and possibly video in the future personal communication systems (PCS). The requirements for the various types of services are application dependent and could be very different. For example, in the future IVHS the data communication between vehicles and base stations for vehicular traffic management does not impose stringent delay constraint, whereas the data communication between vehicles for automatic vehicular control have a very stringent delay constraint and require high reliability for safety. Another example is the PCS where voice, video and data services

are available for mobile users anytime and anywhere. In PCS, real-time services such as interactive voice and video services need guaranteed delay but can tolerate some loss of packets, whereas data services usually do not impose delay constraints but require high reliability. In addition, the bit rate (throughput) requirement of a video call is much higher than that of a voice call.

The different service requirements on delay and throughput have a strong influence on the design of the access layer, especially the channel access protocol (also known as multiple access protocol). The channel sharing among multiple users can be achieved in the frequency (FDMA), time (TDMA), or code (CDMA) domains. The design of channel access protocols over the 3 channel-sharing domains for various new applications (e.g., PCS and IVHS) is a complex design issue. However, the importance of this issue is growing due to the scarcity of the radio spectrum and the ever-increasing demand for high bit-rate services.

1.2 Performance Evaluation

From the above, we see that there is a strong interaction between the design of the physical/access layers of wireless communication networks, the operating environment, and the service requirements. To understand the trade-offs between the many complicated factors involved, we need to evaluate the performance of each design option. In general, there are 3 approaches for performance evaluation [29].

1) Analytical Approach. The approach is to extract the key parameters and then create analytical (mathematical) models to study the performance of the network being considered. To be able to represent a complicated wireless network mathematically, some simplifying assumptions have to be introduced. Though the performance estimate might not be very accurate, this approach can give designers a first impression on the network and help to identify the potential problems at the early stage of design.

2) Simulation Approach. This approach is to create a more realistic network model and evaluate its performance by computer simulation. Simulation can be used to study the behavior of the network in more detail with the help of the powerful computational capability of modern computers. Nevertheless, the computer run time may become prohibitively long as the model gets more and more complicated.

3) Hybrid Approach. By combining the merits of the first two approaches, this approach incorporates some analytical models (for parts of the network where good analytical models exist) into a simulation model to speed up the simulation time. This approach can give a reasonably accurate performance estimate in much less time than the pure simulation approach.

To develop a network model that takes into accounts all the complicated factors mentioned in Section 1.1 is next to impossible [9]. Usually, researchers have to focus on some aspects of the network and develop performance models (either analytical or simulation) that can capture the key features in these aspects of interest.

1.3 Thesis Outline

In this thesis, we study several wireless communication networks of different service requirements analytically and by simulation. The focus of analytical models is on the interface between the physical and access layers of wireless networks. Our approach is to define a set of interface parameters that are rich enough to summarize the effect of the physical layer parameters such as channels, modulation, coding, and spreading. Based on the set of interface parameters, the wireless network controlled by the given access control scheme is modeled by a Markov chain, which takes into account the dynamic behavior of traffic sources. With the analysis of the performance models, various design trade-offs and control parameter optimization are studied.

The performance of a mobile data network employing the slotted ALOHA protocol is studied in Chapter 2 and 3. An analytical model is proposed in Chapter 2 whereas a simulation model is developed in Chapter 3. A set of interface parameters that incorporate fading and near-far effect into the physical layer is formulated. The effect of two retransmission control policies, under both heavy and normal traffic scenarios, on the single-user and network performance is investigated. Based on the performance model, two optimization problems are formulated, namely the maximum throughput problem and the maximum balanced throughput problem. The solution of the first problem maximizes the total network throughput whereas the solution of the second problem maximizes the total network throughput subject to the constraint that all users have the same performance. A linear network in a Rician fading channel is analyzed by the proposed model as an example.

In Chapter 3, we study the impact of propagation effects on the performance of a mobile data network. Four fading channels, including Rayleigh, Rician, with and without shadowing, are investigated by a simulation model. User performance, as a function of the distance to the base station, and the throughput-delay characteristics of the network are evaluated.

Chapter 4 evaluates the performance of a mobile data network with a hard real-time communication requirement. In this distributed packet radio network, the communication requirement is that every user send a status packet to its partner at least every T sec, called the deadline period. We introduce a reliability measure, called the deadline failure probability, which is the probability that the status packet is not successfully delivered to the partner within the deadline period. Reservation ALOHA protocol is used as the access scheme. Two scenarios, the worst case and the steady-state scenarios, are considered. Various trade-offs between the system parameters, such as the channel data rate, the packet error probability, the deadline period, and the reliability requirement are studied by Markovian models.

Chapter 5 focuses on the quality of service (QoS) guarantee in wireless CDMA networks supporting integrated voice, video and data services. A set of interface parameters, called the conditional link quality, are defined. These interface parameters are rich enough to represent both direct sequence and frequency hopping CDMA system. Congestion control policies which maximize the throughput of data users while guaranteeing the QoS of voice and video users are developed and analyzed. Two types of

feedback information about the system state are considered. The system is then modeled by a 3-dimensional Markov chain whose equilibrium state probabilities are solved to find the system performance, i.e., the data throughput and the quality of service of voice/video services.

Chapter 6 concludes the thesis and discuss some possible future research.

Chapter 2

Retransmission Control and Fairness Issue in Mobile Slotted ALOHA Data Networks

In this chapter, the effects of different retransmission control policies on the performance of a mobile data system employing the slotted ALOHA protocol are investigated, with the emphasis on the unfairness between close-in and distant users due to the near-far effect. An analytical multi-group model is developed to evaluate both the user and the network performance of the mobile slotted ALOHA network under two classes of retransmission control policies, namely the Uniform Policy and the Nonuniform Policy. The Uniform Policy requires that all users adopt the same retransmission probability, whereas the Nonuniform Policy allows more distant users to have larger retransmission probability in order to compensate for the unfairness caused by the near-far effect. The performance of a slotted ALOHA network with a linear topology in a Rician fading channel under the two policies is compared by the multi-group model and simulation. The Nonuniform Policy is found to be more effective in alleviating the unfairness of user throughput over a wider range of the data traffic load than the Uniform Policy, which is effective only when the data traffic load is very light. Thus, a mobile data network can

enjoy the network performance improvement derived from the near-far effect while the unfairness between close-in and distant users can be greatly mitigated without resorting to power control.

2.1 Introduction

Mobile packet data networks have been attracting a lot of interest recently due to the flexibility and low cost of the services provided to users on the move. Typical mobile data services include paging and wireless electronic mail services. In addition to the services provided to human users, the signalling traffic generated for network control (e.g., tracking and registration of mobile users) also forms a major need for mobile data networks. Other innovative future systems such as Intelligent Vehicles Highway System (IVHS) require mobile data networks for the communication needs between vehicles and roadside base stations for vehicular traffic management and driver information services.

One of the major design issues in mobile data networks is the choice of the channel access protocol for the mobile-to-base station channel (uplink). The channel from the base station to mobile users (downlink) is not a multiple access channel; neighboring base stations can avoid the contention on the downlink by FDMA, TDMA, or other scheduling algorithms. Therefore, we will only consider the channel access problem on the uplink.

The bursty characteristics of the data traffic on the uplink lead us to consider random access schemes. Among the family of random access schemes, slotted ALOHA (S-

ALOHA) [53] is a simple channel access protocol that has received a great deal of attention since it appeared in early seventies. The throughput-delay performance in a broadcast channel was considered in [7], [28], where simultaneous transmissions are assumed to be lost and a single transmission is assumed to be successful. In a mobile radio environment, the propagation path loss and multipath fading can make the received power of packets differ by orders of magnitude allowing advantage to be taken of the capture effect, the phenomenon that the packet with the strongest power is successfully received even in the presence of other interfering packets. On the other hand, co-channel interference and thermal noise make the probability of successful reception of a single transmission less than one. Therefore, the capture effect improves network performance, whereas co-channel interference and thermal noise degrade performance. Many capture models have been proposed to evaluate network performance, e.g., [17], [18], [49], [68]. Some of them take into account the fading channel, most of which use a Rayleigh fading model due to the analytical tractability.

Stability is another important issue with ALOHA systems [7]. A simple approach to stabilize an ALOHA system is by retransmission control. Many adaptive retransmission control schemes have been considered for broadcast channels (i.e., no capture); see [23] for a recent study. There are some papers considering static retransmission control for S-ALOHA with capture. For example, the stability of S-ALOHA with capture is considered in [49]; the stability condition for a 2-group S-ALOHA system with capture can be found in [44].

Almost all of the published work on S-ALOHA focuses on network performance, with some exceptions such as [17], [49], [8], and [36] where single user performance is considered. In those papers considering single user performance, the near-far effect is found to cause unfair user throughput. To alleviate the unfairness, it was concluded in [8] that some form of power control is necessary, which will inevitably weaken the capture effect and therefore degrade network performance. However, [36] suggests the use of retransmission control to level out the unfairness and yet keep the capture effect intact.

In this chapter we will extend the work of [36] and investigate the effect of retransmission control on both single user and network performance, taking into account multipath fading and the near-far effect. The focus is on the inequity in user performance and the effectiveness of retransmission control to improve the fairness. The approach that we take is to abstract a mobile S-ALOHA network as a multi-group system where the user spatial distribution, channel fading characteristics, and physical layer parameters (such as modulation, coding, diversity, etc.) are matched properly to the system parameters of the analytical multi-group model. The accuracy of the analytical model is validated by a more detailed simulation model (see chapter 3.)

In Section 2.2, the system parameters of the multi-group model are introduced, and the principle of matching a mobile S-ALOHA network with a multi-group system is discussed. Section 2.3 investigates the problem of finding the optimal retransmission control strategy under different objectives and constraints. In Section 2.4, the numerical results of the performance of an example network are presented. Various performance

measures such as single user performance, group performance, and network performance as functions of the retransmission control strategy are considered. Section 2.5 provides some concluding remarks.

2.2 The Multi-group System Model for Mobile S-ALOHA Networks

In this section, we first introduce the analytical system model, namely the multi-group S-ALOHA with capture, the objective of which is to represent a wide range of S-ALOHA networks. We then discuss how to set system parameters of the multi-group model for a given mobile S-ALOHA network, which could be a stand-alone network or a cellular structured network with co-channel interference.

2.2.1. The Multi-group Model

A. Terminal Model and Channel Access Control

Consider a S-ALOHA system with one central station and K groups of terminals (or users). Group i (i = 1, 2,..., K) consists of M_i single-buffered terminals that are identical and independent of each other. Each terminal is either in the idle state or in the backlogged state. When a terminal is in the idle state, a packet will be generated and transmitted in the next slot with probability σ_i, the *packet generation probability*. If the transmission is successful, the terminal will receive an error-free positive feedback right after the transmission and remain in the idle state. Whereas if the transmission is not successful, it

will enter the backlogged state and retransmit the packet with probability q_i, the *retransmission probability*, in each of the following slots until it succeeds. Although the assumption about instantaneous and error-free feedback is to simplify the analysis, the performance obtained serves as an upper bound for the system performance. The effect of imperfect and delayed feedback in a mobile S-ALOHA system can be investigated by simulation but is beyond the scope of this study.

The terminal model described above is general enough for us to consider the effect of channel access control strategies under 2 scenarios: the heavy traffic and the normal traffic scenarios. In the heavy traffic scenario, the packet retransmission probability q_i and the packet generation probability σ_i of group i are set to be the same. This scenario is of interest if one wants to find the capacity (the maximum throughput) of a network. If one is interested in average packet delay or stability of the system, the system performance should be examined over various q_i for a given σ_i, which corresponds to our normal traffic scenario.

The flexibility of the possibly different packet generation/retransmission probabilities (σ_i/q_i) of the multiple groups allows us to investigate more general channel access control strategies which can improve the fairness of the system. We distinguish 2 channel access strategies: the Uniform Policy and the Nonuniform Policy. Under the Uniform Policy, the packet generation probabilities and the retransmission probabilities are the same for all groups (i.e., $\sigma_i = \sigma \ \ \forall i$ and $q_i = q \ \ \forall i$), but σ is not necessarily equal

to q. For the Nonuniform Policy, σ_i and q_i can vary from group to group.

B. General Capture Model

Assuming that the radio channels from terminals in the same group to the central station are statistically the same and independent from each other, the probability that the central station can successfully receive a packet in a slot depends on the activity of all K groups, the interference that comes from the thermal noise, and any co-channel interference. Define the activity vector $\vec{a} = (a_1, a_2, ..., a_K)$, where a_i, $i = 1, 2, ..., K$, is the number of transmissions from group i in a slot. Given the activity vector in a slot, the probability $P_i(\vec{a})$ (called the *conditional capture probability* of group i) that one of terminals in group i successfully transmits a packet to the central station depends on factor such as fading characteristics, modulation, coding, diversity, and the statistics of the interference due to thermal noise and co-channel interference. We can think of $P_i(\vec{a})$ as a parameter that depends on the operating environment and the physical modem structure. For a given operating environment and a physical modem structure, $P_i(\vec{a})$ can be precomputed.

C. Network Performance Model

For a given normal traffic scenario[1], a channel access control policy (uniform or nonuniform), and a set of conditional capture probabilities, the network can be modeled by a K-dimensional Markov chain with the state being the number of backlogged users in each

1. In the heavy-traffic scenario, the network is memoryless from slot to slot, and therefore its performance can be evaluated without the effort of formulating a Markov chain (see Appendix B).

group. A decoupling approximation, which assumes that the state transition of the one-dimensional Markov chain for a group depends only on the equilibrium state occupancy probabilities of the other groups, can be used to avoid having to solve this K-dimensional Markov chain. The decoupling assumption was found to be sufficiently accurate when packets from stronger groups tend to successfully capture the receiver in the presence of the interfering packets from weaker groups [36]. The equilibrium state probabilities of the decoupled Markov chains can then be used to compute the throughput, delay, and stability of each group. Since the users in the same group are symmetric, the user throughput is just the group throughput divided by the number of users in the group; the packet average delay of a user is the same as the average packet delay of the corresponding group. The details of the computational procedure can be found in [36] or Appendix A.

2.2.2 Model Parameter Selection

The system parameters K, $\vec{M} = (M_1, M_2, ..., M_K)$, and $\{P_i(\vec{a})\}_{i=1}^{K}$ need to be determined before network performance can be evaluated. The first parameter to be determined is K, the number of groups. A larger K allows us to see in more detail how user (or group) performance behaves at the cost of the higher complexity of the resulting model and the higher computational cost in determining $\{P_i(\vec{a})\}_{i=1}^{K}$, but K cannot be too large in order to keep the decoupling approximation accurate. After K is determined, users in the network can be put into the same group if they have similar likelihood of being captured at

the central station. One way is to divide the range of the mean received powers at the central station into K (not necessarily equal) intervals and put the users that fall in the same interval into the same group. The distance of group i relative to the central station after grouping is determined so that group i will produce the same average interference power at the central station as the sum of the power of the M_i individual users. For an inverse square power law, the distance between group i and the central station, r_{G_i}, satisfies

$$M_i \frac{1}{r_{G_i}^2} = \sum_{j=1}^{M_i} \frac{1}{r_j^2},$$

where $r_1, \ldots, r_{M_i}$ are the location of the users in group i before grouping.

It should be noted that the above grouping method (for determining $\vec{M}$ of the multi-group model) is independent of the actual spatial distribution (which can be linear, such as vehicles in highways, or planar, such as users in cities.) After the group performance is found, the individual user performance as a function of the distance relative to the central station can be obtained by interpolating group performance. Compared to the conventional approach for the analysis of a S-ALOHA system, where the spatial users distributions have to be restricted to some specific forms due to analytical tractability [6], the proposed grouping method is more general.

2.3 The Optimal Retransmission Control

For a mobile S-ALOHA network which is equipped with a particular physical modem structure and operating in a given environment, the system parameters (i.e., $\vec{M} = (M_1, M_2, ..., M_K)$ and $\{P_i(\vec{a})\}_{i=1}^{K}$) of the corresponding K-group S-ALOHA system can first be determined. Then, there is the question of how to set the retransmission probabilities of the K groups (q_i's) for given packet generation probabilities (σ_i's). Depending on the objective of the network design, we can formulate different optimization problems, some of which are discussed below.

2.3.1 The Maximum Throughput Problem

If the objective is to maximize the total network throughput, the retransmission probabilities $\vec{q} = (q_1, q_2, ..., q_K)$ should be set to the solution of the multivariate[1] nonlinear constrained optimization problem given by

$$\max_{q_1, ..., q_K} S$$

$$\text{subject to } 0 \le q_i \le 1, i = 1, ..., K,$$

where S, the network throughput, is the sum of the throughputs of the K groups. In general, we are unable to solve this optimization problem due to the highly nonlinear relationship

1. For the Uniform Policy, the optimization problem degenerates to a univariate optimization problem.

between $\vec{q}$ and S. However, in [37] we analytically solved a special case of the problem where the group with stronger power operates as if there were no other (weaker) groups, an assumption that has been used for studying the performance multi-group S-ALOHA system with infinite population, e.g. [41], or with finite population, e.g. [35]. The key steps of the analysis of the special case also can be found in Appendix B. In the numerical example presented later, numerical algorithms based on [22] are used to solve the above optimization problem.

The solution to this maximum throughput problem usually results in a scenario where most of the network throughput comes from the users in the close-in region of the central station. This phenomenon is acceptable if the data traffic is not delay-sensitive and users who are distant from the central station will move closer at some time. If the data traffic is delay-sensitive (i.e., the packet cannot wait until users move to the neighborhood of the central station), we may want to consider the maximum balanced throughput problem below.

2.3.2 The Maximum Balanced Throughput Problem

To ensure that the user throughput is the same throughout the system while the network throughput is maximized, we need to solve another constrained optimization problem given by

$$\max_{q_1, \cdots, q_K} \quad S$$

subject to $0 \leq q_i \leq 1$, $i = 1, \ldots, K$, and

$$\frac{S_1}{M_1} = \frac{S_2}{M_2} = \ldots = \frac{S_K}{M_K},$$

where S_i and M_i are the throughput and the population of group i, respectively. There are $(K - 1)$ more constraints here than in the maximum throughput problem, and so we can expect that the maximum balanced throughput will be smaller than the maximum throughput for the unbalanced case. The reduction in throughput is the cost paid to achieve fairness.

Similar to the maximum throughput problem, the maximum balanced throughput problem can also be solved by many numerical algorithms. In [37], we analytically solve a special case which is the same as in Section 2.3.1 except for the additional $(K - 1)$ constraints. The key steps are summarized in Appendix B.

2.4 Application

Most of the work on the analysis of mobile S-ALOHA networks considers Rayleigh fading. In order to demonstrate the flexibility of our model, we consider a linear mobile S-ALOHA network in a Rician fading channel. In addition, the performance of the uniform and nonuniform retransmission control policies are also examined.

2.4.1 The Network

A. Topology

Consider a S-ALOHA network with 50 users equally spaced in an interval of [$d_{\min}$, 1] as shown in Fig. 2.1. $d_{\min}$, which is greater than zero, reflects the reality that users cannot approach too close to the central receiver. Here $d_{\min}$ is chosen such that the maximum mean receiving power is 30 dB (i.e., $d_{\min} = 0.0316$) with the power of user 50 as the reference.

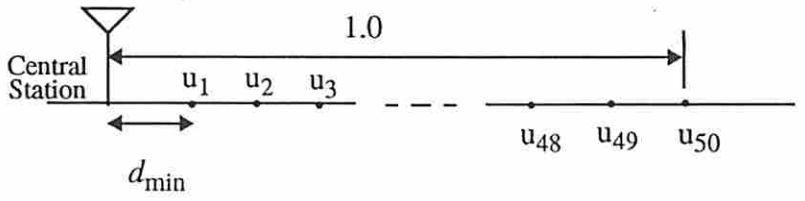


Fig. 2.1. The topology of a S-ALOHA network with 50 users equally spaced in [$d_{\min}$, 1.0]

B. Channel Model

All users transmit with the same power. The path loss exponent of the channel from mobile users to the central station ranges from 1.2 to 4.0 in micro cellular environments [65]. In the present study, the exponent is chosen to be 2. In addition to the path loss, the signals from mobile users to the central station suffer Rician fading with the instantaneous signal power following a noncentral chi-square distribution given by [34]

$$f(x|P, K_r) = \frac{(1+K_r)\, e^{-K_r}}{P} exp\left(-\frac{1+K_r}{P}x\right) I_0\left(\sqrt{4K(1+K_r)\frac{x}{P}}\right),$$

where P is the mean signal power (proportional to r^{-2}), K_r is the Rician factor, and I_0 is the modified Bessel function of the first kind of order zero. Measurement results show that K_r ranges from 7 dB to 12 dB [6]. In our study, K_r is chosen to be 10 dB. The instantaneous signal power is assumed to be constant over a packet, i.e., slow fading. Note that the shadowing effect is not considered here.

C. Capture Criterion

The receiver at the central station is assumed to be able to successfully receive a packet in the presence of other simultaneous transmissions if the ratio of the power of the desired signal and the sum of interfering powers is greater than a threshold R, called the capture threshold. The value of R depends on the modulation and coding schemes used [34] and is chosen to be 4 (6 dB) in the present study.

2.4.2 Approximation by a K-group System

A. Selection of K

The range of the mean receiving power is between 0 dB (from user 50) and 30 dB (from user 1). One way to group users is to divide the 30 dB range into K equal intervals (in dB) and then put the users falling in the same interval into the same group. For a 2-group approximation, users 1 to 8 are in group 1, whereas users 9 to 50 are in group 2, i.e., $\vec{M} = (8, 42)$. Similarly, we have $\vec{M} = (4, 11, 35)$ for a 3-group approximation, $\vec{M} = (3, 5, 12, 30)$ for a 4-group approximation, and $\vec{M} = (2, 3, 7, 12, 26)$ for a 5-group approximation. As K increases, the accuracy of the multi-group approximation increases (provided that

decoupling approximation is still good) at the cost of the increasing complexity in the conditional capture probabilities.

B. Computation of the Conditional Capture Probabilities for Rician Fading

Given the activity vector $\vec{a} = (a_1, a_2, ..., a_K)$, the conditional capture probability of group i is $P_i(\vec{a}) = a_i \cdot P_r[X > R \cdot Y]$, where X is the received power of a particular packet of group i, and Y is the sum of the received powers of all other packets. Given $a_i \neq 0$, X is a noncentral chi-square random variable with parameters $P = 1/r_i^2$ and K_r. The probability density function (pdf) of Y is the convolution of the pdf's of all other transmissions, each of which is a noncentral chi-square random variable with the same K_r, but different P. The pdf of Y does not have a closed form expression. In our study, Y is approximated by a noncentral chi-square random variable with the same mean and variance as the sum of all interfering non-central chi-square random variables. With this approximation, $P_i(\vec{a})$ can be numerically evaluated by the double integral

$$\int_0^\infty \left[\int_0^{\frac{x}{R}} f_Y(y)\, dy \right] f_X(x)\, dx .$$

Note that this approximation is exact when there is only one interferer in a slot.

In this analysis, we have made 3 approximations: one being the grouping approximation, another being the decoupling approximation for the K-dimensional Markov chain, and the third being the chi-square approximation for the distribution of the sum of

the power of all interfering packets. The accuracy of the system performance obtained by these approximations made here is validated by simulation results.

2.4.3 Numerical Results and Discussion

The unbalanced user throughput as a function of the distance relative to the central station in heavy traffic for the network given in Section 2.4.1 is shown in Fig. 2.2. It shows that close-in users have significantly higher throughput than distant users, although all users transmit with the same probability ($\sigma = q = 0.02$). The figure demonstrates that the prediction of the user throughput by the multi-group model closely matches simulation results, which shows that the grouping approximation and the chi-square approximation are good as far as system performance is concerned. Note that the ability of the multi-group model to predict the user throughput as a function of the distance relative to the central station improves as K increases.

To demonstrate the accuracy of the multi-group approach for a wide range of network loading, we plot, in Fig. 2.3, the network throughput as a function of the packet generation probability for various grouping approximations. This figure shows that the 2-group approximation is not as accurate as the 3-group or 5-group approximations. Note that all of these grouping approximations are good when the network is operating at a point below maximum capacity. The discrepancy between the multi-group models and the simulation model is primarily due to the grouping method that lumps together all users in a group, which results in a capture phenomenon weaker than in the real network. With the similarity between the 3-group and 5-group model results, the simpler 3-group model will be used for

the rest of the numerical examples (unless mentioned otherwise.)

Fig. 2.4 shows the impact of the choices of packet generation probability and retransmission probability on the stability of the system. As expected, for a particular packet generation probability, there is a maximum retransmission probability, beyond which the network throughput will drop very fast–the *unstable region*. The figure shows that a smaller retransmission probability can ensure stability of the system for a wider range of packet generation probabilities at the cost of higher packet delay.

Fig. 2.5-2.10 explore the effect of the Uniform Policy on network and user throughput. We consider 3 retransmission probabilities: $q = 0.04$ in Fig. 2.5-2.6, $q = 0.08$ in Fig. 2.7-2.8, and $q = 0.12$ in Fig. 2.9-2.10. In Fig. 2.5, the network throughput increases monotonically as the packet generation probability σ increases, whereas the group-3 throughput shows a behavior of increase initially and then decrease. This indicates that higher overall network throughput does not necessarily mean that all users experience higher throughput. In fact, the weaker groups are sacrificed to the stronger groups when the channel load is high. Fig. 2.6 shows that the unfairness between user throughput deteriorates as σ increases. In fact, it can be seen that after σ exceeds some threshold (close to 0.01 in this case,) the group-3 user throughput starts to saturate and then decrease, whereas the group-1 user throughput continues to increase due to capture.

In Fig. 2.7-2.8, we increase q from 0.04 to 0.08. Now the network throughput does not increase monotonically as σ increases. From Fig. 2.7, we see that the network throughput drops because of the breakdown of group 3, which consists of most of the users

in the network. But the network throughput picks up again later since the group-1 throughput keeps increasing and the group-2 throughput does not decrease too fast. Fig. 2.8 shows the behavior of the user throughput as σ increases. As in Fig. 2.6, the fairness of the user throughput begins to degrade when σ exceeds about 0.01. Fig. 2.9-2.10 show similar behavior as Fig. 2.7-2.8 except that the breakdown of group 3 is more dramatic because of the higher retransmission probability $q = 0.12$. From Fig. 2.6, 2.8, and 2.10, we find that the fairness of user throughput exhibits a threshold phenomenon, and that the unfairness beyond the threshold gets more critical for larger retransmission probabilities.

From the results in Fig. 2.5-2.10, we see that there is a trade-off between fairness and efficiency for a S-ALOHA network with the Uniform Policy. To show the trade-off, we define two indices: the fairness index and the efficiency index. The *fairness index* is defined as the ratio of the user throughput of the weakest group to the user throughput of the strongest group, which is always less than one for the Uniform Policy, and the *efficiency index* is defined as the ratio of the network throughput to the maximum network throughput for a given retransmission probability. To avoid clutter, Fig. 2.11 shows the design trade-off for only two values of the retransmission probability q. In both cases, as the packet generation probability σ increases, the fairness index decreases whereas the efficiency index increases up to the point where the maximum throughput is reached. The point where the two indices are equal suggests a good trade-off between the network efficiency and the fairness of the user throughput.

Fig. 2.12-2.13 explore the effect of the Nonuniform Policy on network and user

throughput. The retransmission probabilities of the 3 groups, $\vec{q}$ = (2.8E-3, 2.3E-2, 0.1), are chosen to give the maximum balanced throughput for $\sigma = 0.01$. Although the total network throughput shown in Fig. 2.12 looks similar to the one in Fig. 2.7, the distribution of the network throughput for the 3 groups are different. In Fig. 2.13, with the smaller retransmission probability, the group-1 user throughput does not increase as fast as in the case of the Uniform Policy (see Fig. 2.6, 2.8 and 2.10.)

The comparison between the Uniform Policy and the Nonuniform Policy is highlighted in Fig. 2.14-2.15. Fig. 2.14 shows that the network throughput is not sensitive to the retransmission probability before the maximum throughput (which depends on the retransmission probability) is reached. As expected, a lower retransmission probability will give a larger maximum network throughput. Fig. 2.15 compares the fairness indices of different retransmission control policies. It shows that the Nonuniform Policy is able to maintain better fairness over a wider range of network loading compared to the Uniform Policy. Note that the fairness index of the Nonuniform Policy in Fig. 2.15 is slightly greater than 1.0 for σ smaller than 0.01. This indicates that the user throughput of group 3 is slightly larger than that of group 1 when σ is less than 0.01. We also observe that the fairness indices of the 4 cases considered are all above 0.8 as long as σ is smaller than 0.01, which means that the unfairness is not serious if the packet generation probability is always less than 0.01.

Having discussed the effect of the retransmission control policies, we compare the numerical results of the maximum throughput and the maximum balanced throughput of

several multi-group systems that are derived from the S-ALOHA network introduced in Section 2.4.1 in heavy traffic (i.e., $\sigma_i = q_i$), see Table 2.1. These results are obtained by numerical optimization. Both the maximum throughput and the maximum balanced throughput increase as the number of groups (K) increases due to the improved capture capability under the near-far effect. This strongly suggests that it is very inefficient to achieve balanced throughput by perfect power control (corresponding to $K = 1$) where all users adjust their transmission power so that their packets will arrive at the central station with equal strength. On the other hand, retransmission control (corresponding to higher values of K) should be adopted to take the advantage of the capture phenomenon due to the near-far effect. In fact, it can be inferred that if each user can fine tune its retransmission probability by interpolating the optimal retransmission probabilities obtained by the 5-group system, the system performance can be further improved.

We also note that as the network throughput is maximized, the fairness is very poor among close-in and distant users. In order to enforce the fairness among users' throughputs, the network has to sacrifice some capacity. The difference between the maximum throughput and the maximum balanced throughput is the cost of the fairness of user throughput.

2.5 Conclusion

An analytical multi-group model is developed to evaluate the performance of mobile

slotted ALOHA networks under two retransmission control policies. The effectiveness of the policies in improving the fairness of user performance in a mobile environment with fading and near far effects are examined. The Uniform Policy is simple and can achieve reasonable fairness if the packet generation probability is kept below a certain threshold (exceeding which will result in a highly unfair system.) The Nonuniform Policy, however, is able to provide better fairness over a wider range of packet generation probabilities.

The cost of applying retransmission control to achieve balanced user throughput is considered by comparing the numerical solutions of the maximum throughput problem and the maximum balanced throughput problem. The results indicate that power control is not as efficient as retransmission control for improving fairness of the system.

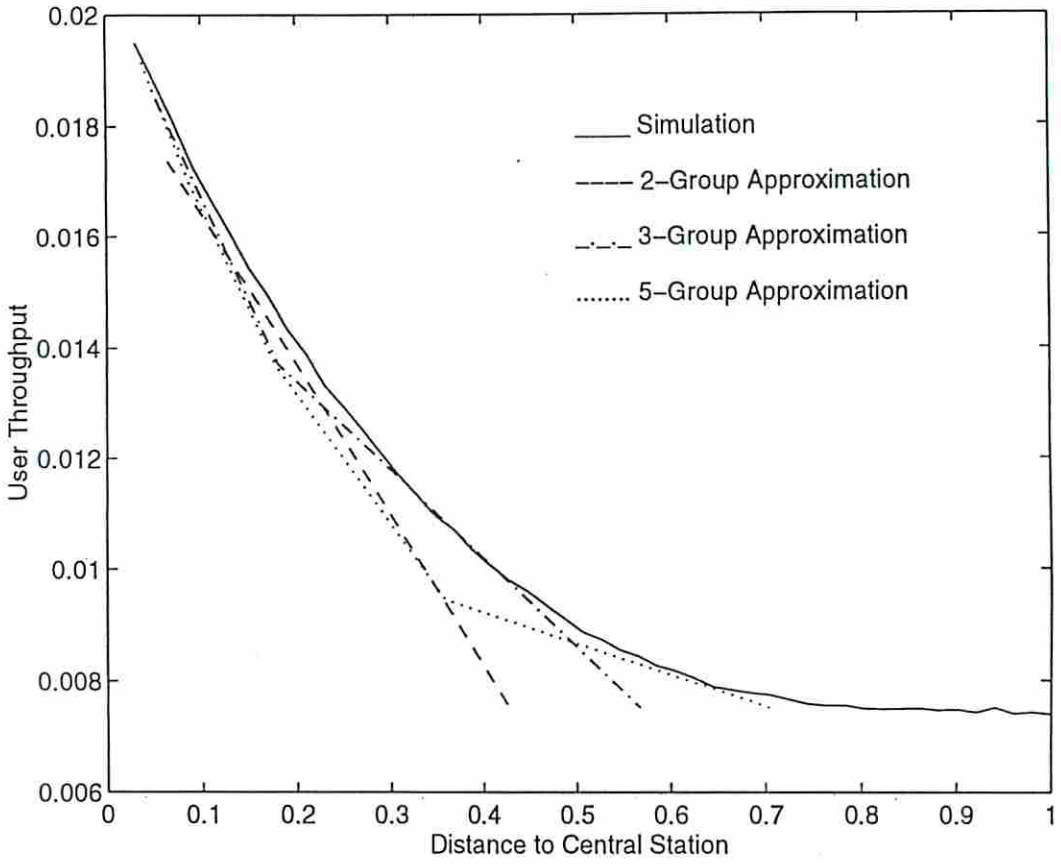


Fig. 2.2. The unfair user throughput as a function of the distance relative to the central station for the slotted ALOHA network with Rician fading given in Section 2. 4.1. The transmission probability $\sigma = q = 0.02$ and the Rician factor $K_r = 10$ for all users.

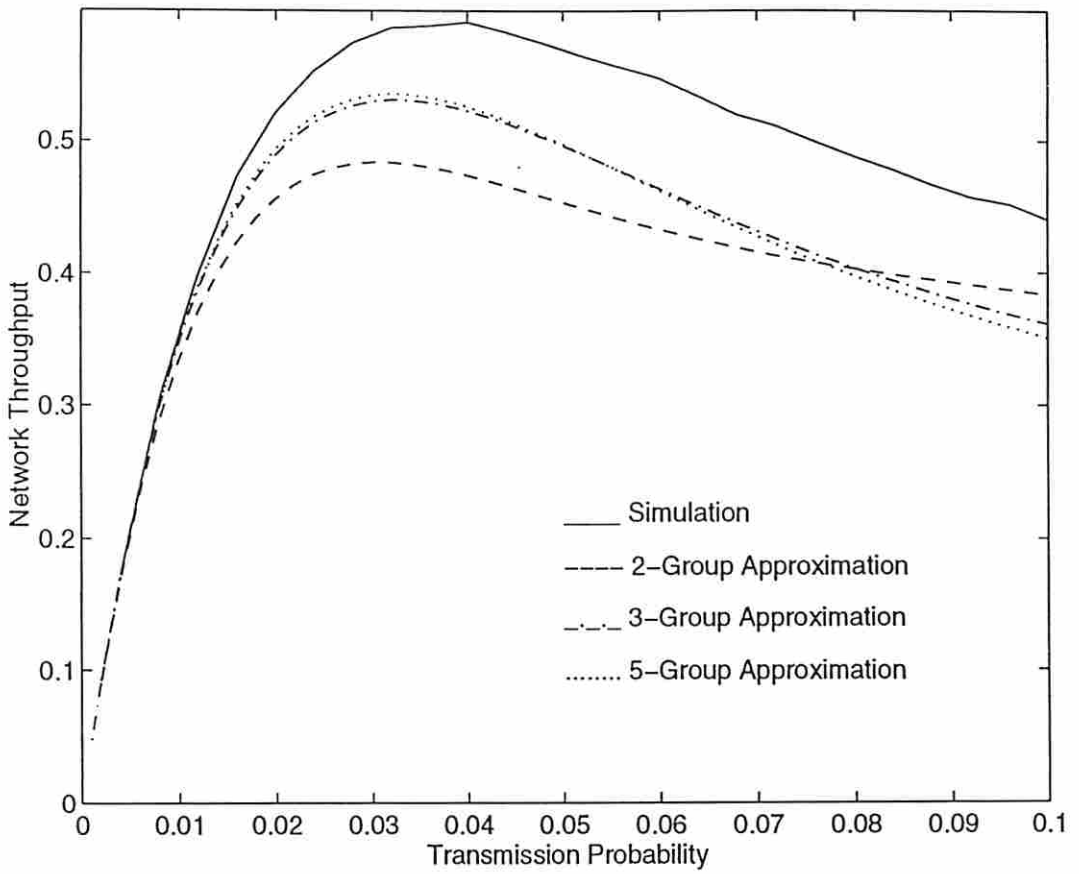


Fig. 2.3. Network throughput versus transmission probability in heavy traffic ($\sigma = q$) under the Uniform Policy for different grouping approximations compared against simulation.

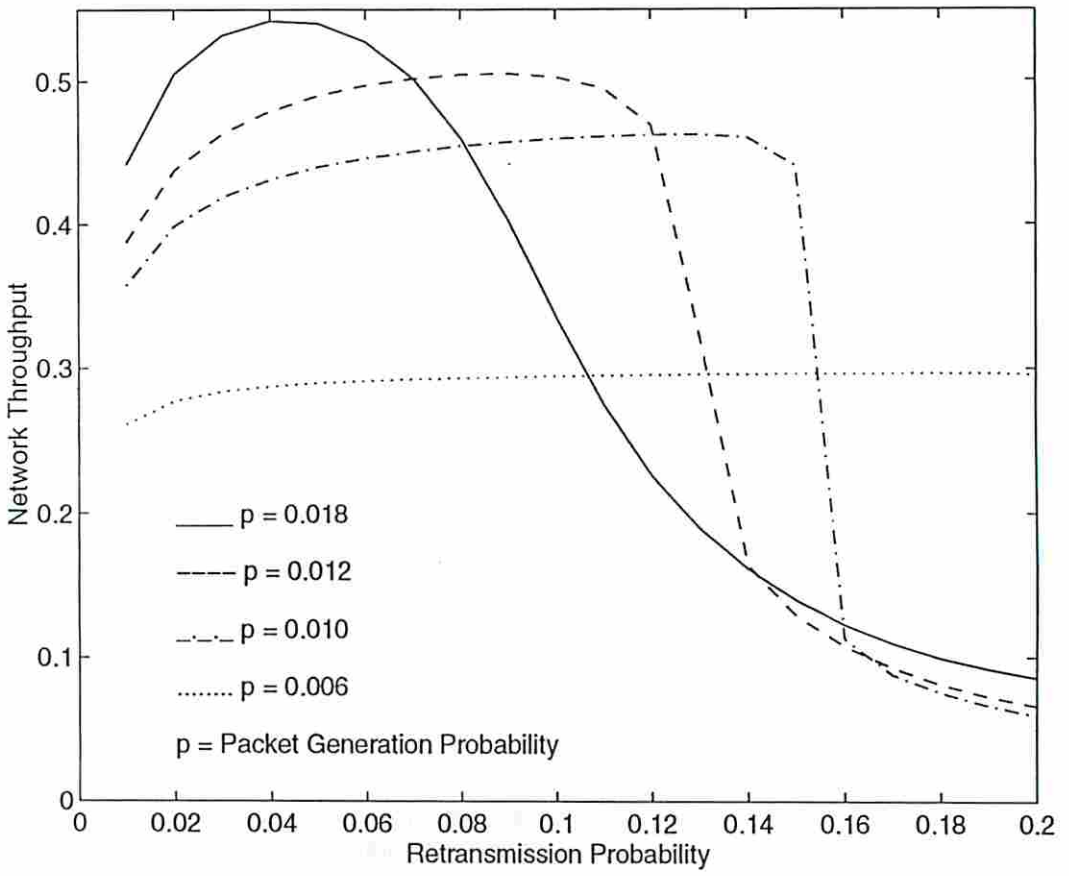


Fig. 2.4. Joint effect of σ and q on network throughput. The breakdown of network for q exceeding some threshold shows the unstable nature of S-ALOHA system.

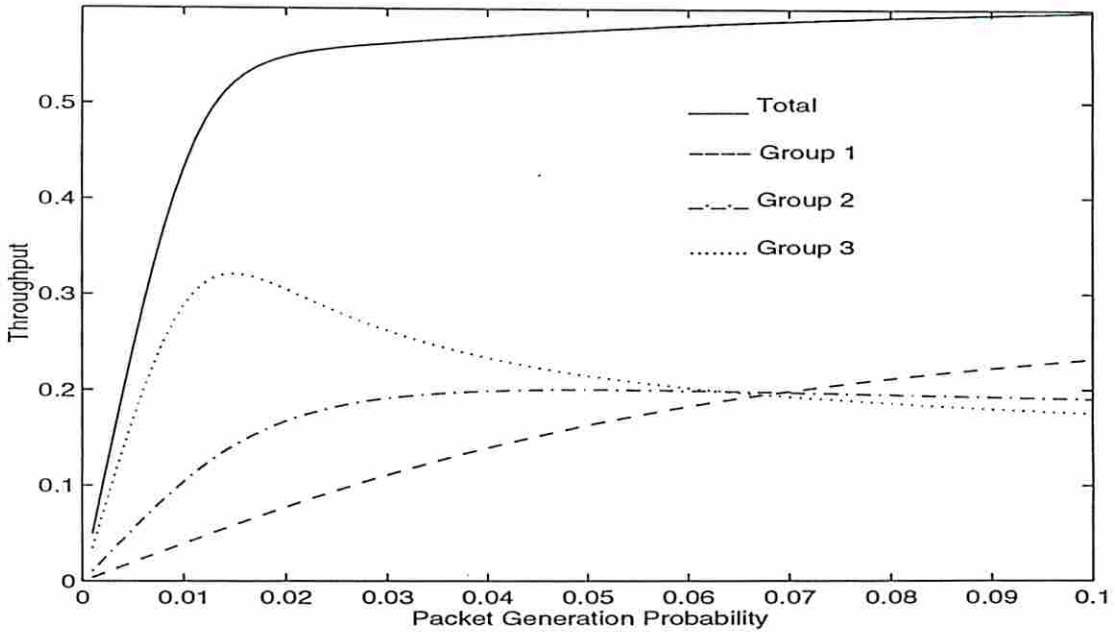


Fig. 2.5. Total/group throughputs versus σ when $q = 0.04$ under the Uniform Policy.

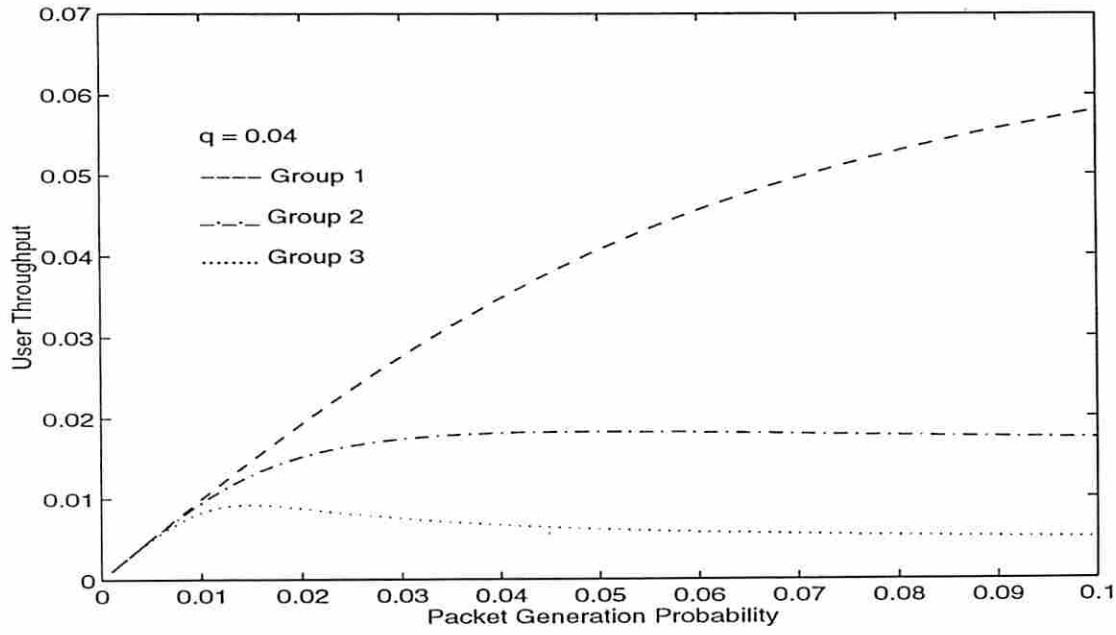


Fig. 2.6. Fairness of the user throughput under the Uniform Policy in normal traffic where $q = 0.04$.

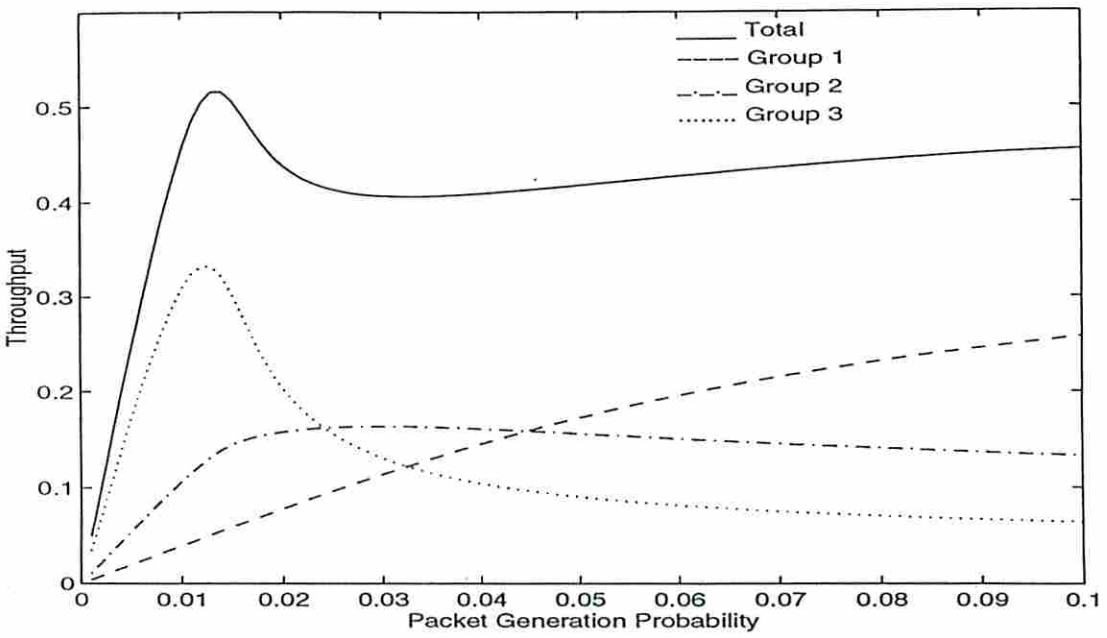


Fig. 2.7. Total/group throughputs versus σ when $q = 0.08$ under the Uniform Policy.

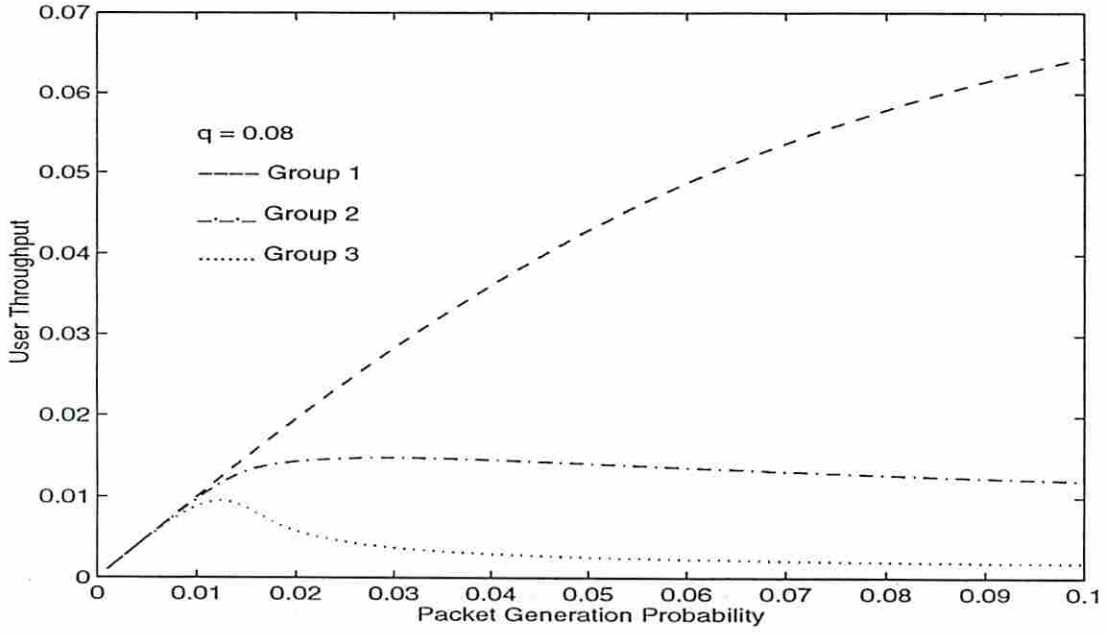


Fig. 2.8. Fairness of the user throughput under the Uniform Policy in normal traffic where $q = 0.08$.

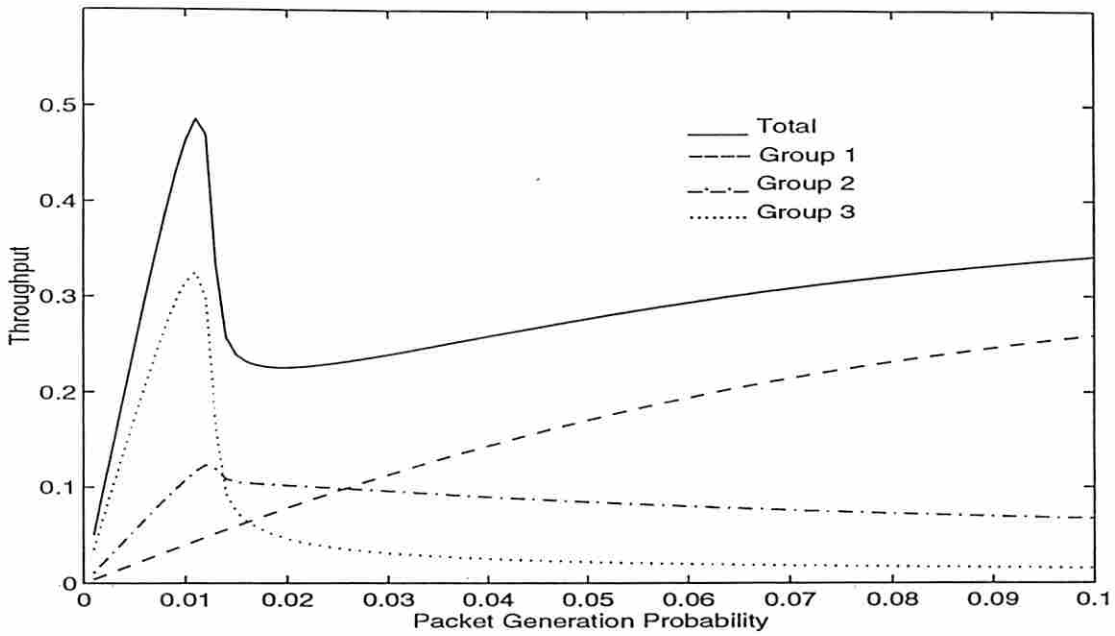


Fig. 2.9. Total/group throughputs versus σ when $q = 0.12$ under the Uniform Policy.

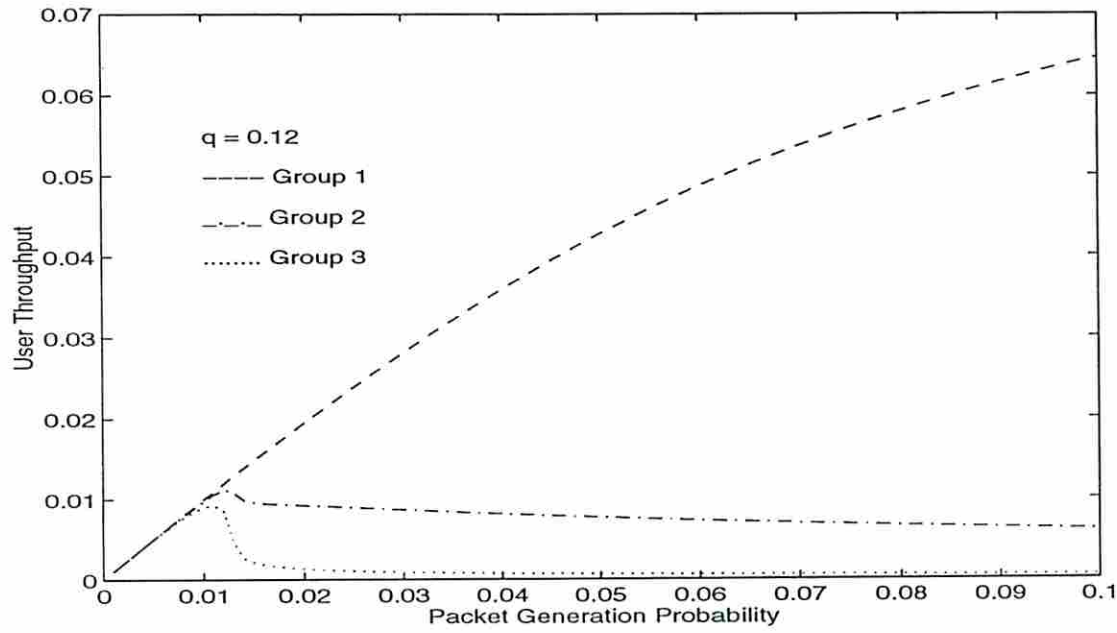


Fig. 2.10. Fairness of the user throughput under the Uniform Policy in normal traffic where $q = 0.12$.

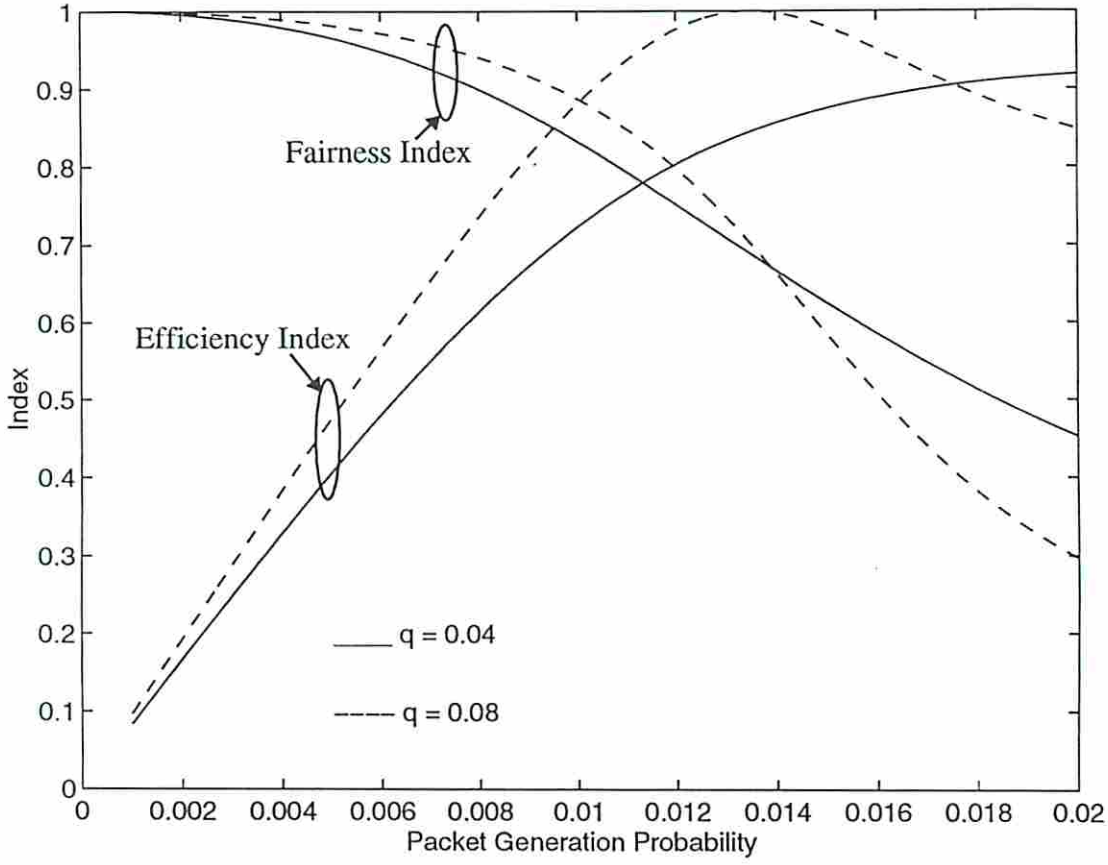


Fig. 2.11. Trade-off between user fairness and network efficiency.

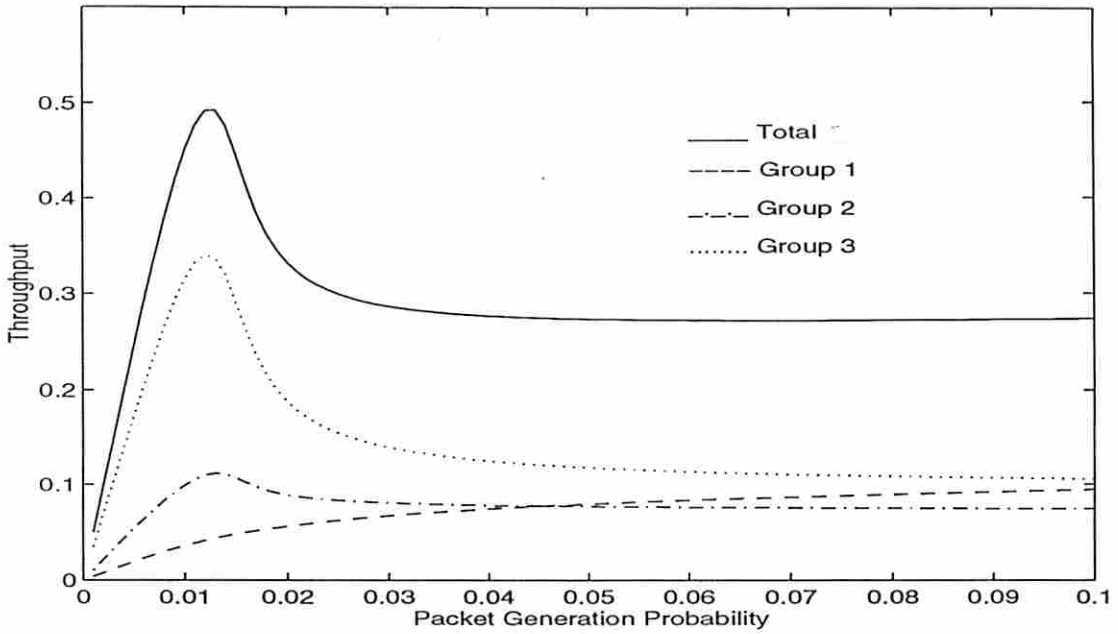


Fig. 2.12. Total/group throughputs versus σ under the Nonuniform Policy. $\hat{q}$ = (2.8E-3, 2.3E-2, 0.1) is the solution that achieves the maximum balanced throughput when $\sigma = 0.01$.

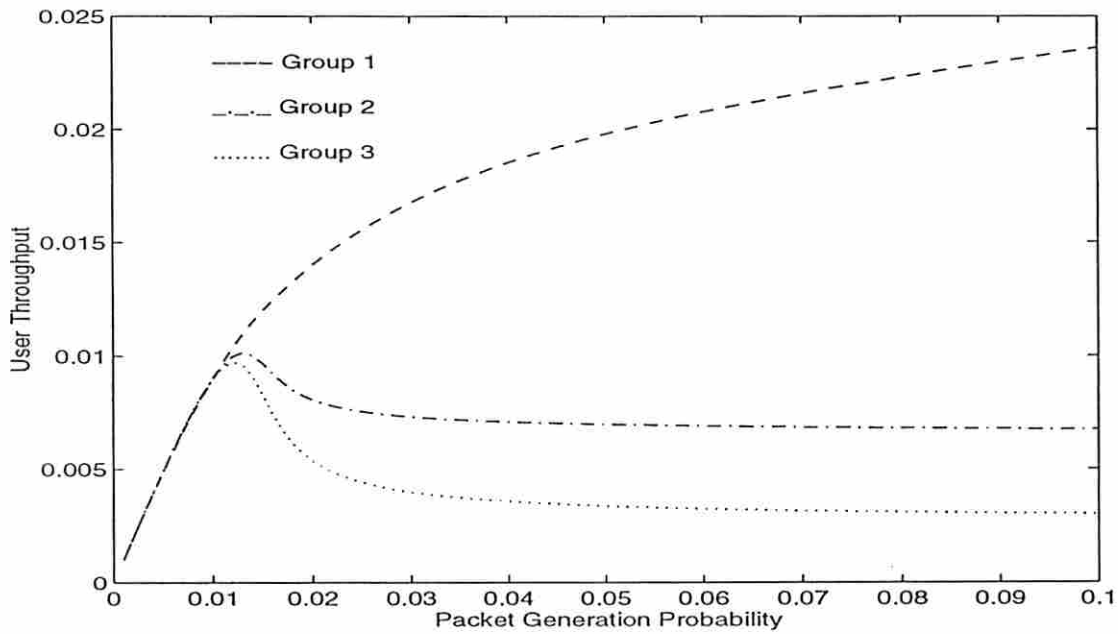


Fig. 2.13. Fairness of the user throughput under the Nonuniform Policy in normal traffic. $\hat{q}$ is the same as in Fig. 2.12.

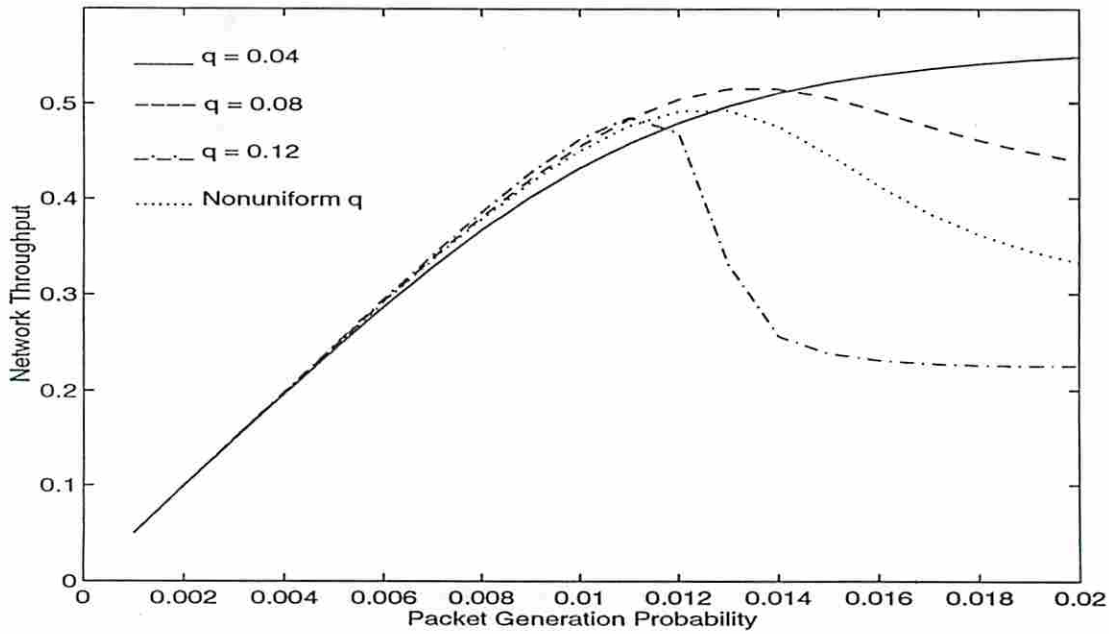


Fig. 2.14. Network throughput versus packet generation probability for various retransmission control policies.

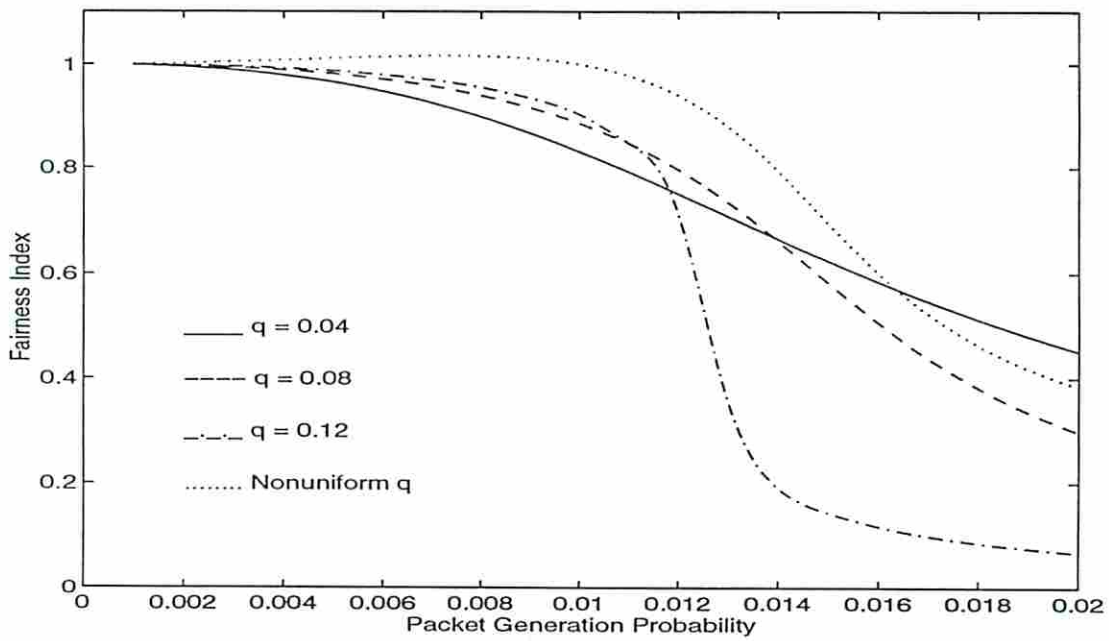


Fig. 2.15. Fairness comparison for various retransmission control policies.

Table 2.1: Comparison between the maximum throughput and the maximum balanced throughput of the K-group slotted ALOHA systems

	Maximum Throughput			Maximum Balanced Throughput		
	Transmission Probability $\vec{q}$	User Throughput	Network Throughput	Transmission Probability $\vec{q}$	User Throughput	Network Throughput
$K=1$ $\vec{M}=(50)$	.0204	.0076	.3786	.0204	.0076	.3786
$K=2$ $\vec{M}=(8, 42)$	.0831 .0239	.0456 .0045	.5542	.0090 .0242	.0084	.4202
$K=3$ $\vec{M}=(4, 11, 35)$	.1409 .0464 .0226	.0871 .0141 .0034	.6224	.0094 .0125 .0280	.0091	.4549
$K=4$ $\vec{M}=(3, 5, 12, 30)$	.1767 .0748 .0153 .0278	.1115 .0281 .0036 .0039	.6353	.0096 .0112 .0175 .0300	.0094	.4682
$K=5$ $\vec{M}=(2,3,7,12,26)$	.3195 .0000 .0857 .0000 .0339	.2166 .0000 .0223 .0000 .0036	.6826	.0097 .0105 .0132 .0222 .0311	.0095	.4735

Chapter 3

Mobile Slotted ALOHA Data Networks Under Different Fading Channels

A simulation model is developed to compare the performance of mobile slotted ALOHA networks with linear topology (such as the communications from vehicles to base stations in the future Intelligent Vehicle Highway Systems) under different channel models. Typical mobile radio channels such as Rayleigh, Rician, Rayleigh with lognormal shadowing, and Rician with lognormal shadowing are considered and their throughput-delay performance are compared against each other. It is found that different channels have similar performance when data traffic load is light. The difference becomes more significant when load is heavy. Simulation results show that none of the 4 channels always has the best performance. An interesting phenomenon was found for the network configuration considered – when the network is overloaded, network throughput may first decrease and then increase as the load keeps increasing.

3.1 Introduction

Slotted ALOHA is attractive due to its simplicity. It has been considered as a viable multiple access protocol in various data communication networks, e.g., the communication from vehicles to base station in the future Intelligent Vehicles Highway Systems. In the mobile radio environment, transmission signals suffer propagation loss, shadowing, and multipath fading. Hence signals transmitted by different terminals arrive at the base station with different power levels. This gives the strongest signal a chance to be successfully received by the base station in the presence of other simultaneous transmission. This phenomenon, called the capture effect [31], can improve network performance. The performance of slotted ALOHA with capture in the mobile radio environment has been studied by many researchers, e.g., [4], [18], [49], [68]. Most of them consider a particular channel model. In the present study, a simulation model is developed to compare the performance of slotted ALOHA under different channel models. This effort will help us to understand how the propagation effects in the mobile radio channel impact system performance.

Section 3.2 introduces the 4 channel models to be studied and the capture criterion. Section 3.3 describes the network model, including models for terminals and other-cell interference. The simulation experiments and results are discussed in Section 3.4.

3.2 Channel Models and Capture Criterion

3.2.1 Characteristics of the Mobile Radio Channel

Signal propagation in a mobile radio channel is characterized by propagation power loss, shadowing, and multipath fading. When the direct path component (main LOS ray) exists, the variation of signal strength with distance was found to show distinct near and far regions separated by a break point. The power loss factor is found to be less than 2 before the break point, while it is greater than 2 after the break point [65]. In the present study, the power loss factor is chosen to be 2 before the break point and 3 after the break point. So the area mean power is related to the transmission power by

$$P_{area} = r^{-\alpha} P_{transmit}, \text{ where } \alpha = 2 \text{ or } 3.$$

With lognormal shadowing, the local mean power has a probability density function given by

$$f(p) = \frac{1}{\sqrt{2\pi}\sigma_s p} exp\left[-\frac{1}{2}\left(\frac{ln\,(p) - ln\,(P_{area})}{\sigma_s}\right)^2\right],$$

where σ_s is the standard deviation of the logarithm of the local mean power. The value of σ_s depends on the severity of shadowing. Here σ_s is chosen to be 1 neper (or 4.34 decibels).

With the main LOS ray, multipath fading is usually modeled as Rician fading with the Rician factor K defined as the ratio of the power of the LOS component to the scattered power. In the microcellular environment, K is measured to be 7 dB to 12 dB [6]. In our

study, K is chosen to be 10 dB. Note that the channel becomes Rayleigh when $K = 0$.

3.2.2 The 4 Channel Models

Four channel models are considered in this chapter, namely, Rayleigh ($K = 0$, $\sigma_s = 0$), Rician ($K = 10$, $\sigma_s = 0$), Rayleigh with lognormal shadowing ($K = 0$, $\sigma_s = 1$), and Rician with lognormal shadowing ($K = 10$, $\sigma_s = 1$). Though some of them are less likely to exist in microcellular environment, the performance of these channels can help us understand how K and σ_s impact system performance.

3.2.3 The Capture Criterion

It is assumed that the packet with the strongest signal can capture the receiver at the base station if its power is at least R times greater than the sum of all other transmission, including other-cell interference. R is called the capture threshold whose value depends on the modulation and coding schemes. Typical values of R range from 3 dB to 20 dB [34]. In the present study, R is chosen to be 4 (or 6 dB).

3.3 Network Model

3.3.1 Terminal Model

Assume there are M independent single-buffered terminals (or users) in a cell, i.e., u_1, u_2,...,u_M. The distance between u_i and base station is d_i for $i = 1, 2,..., M$. A terminal is either in the idle state or in the backlogged state. When in the idle state, a packet will be

generated and transmitted in the next slot with probability p_0. If the transmission is successful, the terminal will receive a positive feedback right after the transmission and remain in the idle state. Whereas if the transmission is not successful, it will enter the backlogged state and retransmit the packet with probability q in the following slots until it succeeds. We can consider (as an approximation) that terminal i will transmit with probability b_i (the active probability), including new transmissions and retransmissions. The b_i's can be measured by simulation and will be used to calculate the other-cell interference.

3.3.2 Other-cell Interference

To reduce the complexity of the simulation model, the effect of transmissions from users in the other cells is not simulated exactly. Instead, the interference power from other cells is approximated as Gaussian with mean m_I and variance v_I given by

$$m_I = \sum_{\text{all interferes}} b_k \,(\text{area mean power of interferer} k) \quad \text{and}$$

$$v_I = \sum_{\text{all interferes}} b_k \,[\,(\text{area mean power of interferer} k) - m_I]^2 .$$

C. Performance Measure

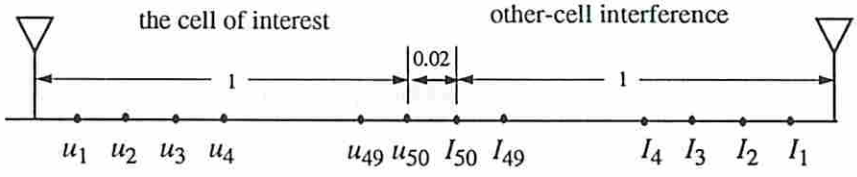
After a simulation run, the throughput and average packet delay are collected for each terminal. This gives the user performance. The network throughput can be obtained by summing the throughputs of all users in a cell. The network average delay can be obtained by

$$D = \sum_{i=1}^{M} D_i \frac{S_i}{S},$$

where S_i, S and D_i are the throughput of user i, the total network throughput, and the delay of user i, respectively.

3.4 Simulation Experiments and Results

The simulation experiments are based on the network topology shown below. $u_1, u_2,..., u_{50}$ are in the cell of interest, and $I_1, I_2,..., I_{50}$ are interferers in other cells. The cell radius is normalized to 1. It is assumed that the break point is on the cell boundary.



With equal spacing between consecutive users, we have $d_1 = 1/50$ and $d_i = i * d_1$. Define d_i' to be the distance between I_i and the base station of interest. d_i' is given by $d_i' = 2 + d_1 - d_i$. We assume that u_i and I_i have the same active probability b_i. To get estimates of m_I and v_I, they are initially set to 0. After the first simulation run, the active probabilities of $u_1, u_2,..., u_{50}$ are used to compute m_I and v_I. Then we execute 10 independent runs to collect performance statistics of each user and compute the network throughput and average delay. Ninety-five percent confidence intervals are provided as long as they will not clutter the plot.

Fig. 3.1 and 3.2 show the throughput and delay of each user for a lightly loaded system ($p_0 = 0.005$, $q = 0.05$). It is observed that the 4 channels have similar performance. Fig. 3.3 and 3.4 show the user performance for a system with higher load ($p_0 = 0.006$, $q = 0.12$). It is found that channels (Rayleigh or Rician) with shadowing have higher network throughput than channels without shadowing.

Fig. 3.5, 3.7 and 3.9 compare network throughput of the 4 channels as a function of p_0 and q. One interesting phenomenon was observed. In Fig. 3.7 and 3.9, throughput increases as we increase load until the combined load of near and far users saturate the channel and throughput drops. As we further increase the load, the far users are driven into the backlog state but the near users continue to succeed on their first attempt, resulting in increased throughput. As we increase load further we could expect a drop in throughput again when the near users saturate.

Fig. 3.6, 3.8 and 3.10 compare the throughput-delay performance of the 4 channels with various retransmission probabilities (i.e., $q = 0.04$, 0.08 and 0.12). It is observed that Rician has slightly better performance than Rayleigh when load is light. When network is heavily loaded, Rayleigh has better performance than Rician. Similarly, Rician with shadowing has better performance than Rayleigh with shadowing when load is light; Rayleigh with shadowing performs better than Rician with shadowing when load is heavy. Also, it is observed that channels with shadowing have better performance when load is heavy, whereas channels without shadowing have slightly better performance when load is light. From Fig. 3.6, 3.8 and 3.10, we can see as expected that for all channels larger q has lower

delay when load is light, whereas larger q has smaller capacity when load is heavy.

The user performance in a Rician channel for 3 different packet arrival probabilities (p_0 = 0.004, 0.007 and 0.012) are shown in Fig. 3.11 and 3.12. We can see, in Fig. 3.11, that the 3 different values give approximately the same network throughput (0.2) but completely different spatial distribution. In particular, with p_0 = 0.004, all users have similar throughput. As p_0 increases to 0.007, near and far users have significantly unbalanced throughput. When p_0 = 0.016, almost all the network throughput is attributed to near users. In Fig. 3.12, it can be seen that the delay of far users increases as p_0 increases. Though the delay of far users is large, their throughputs are small. Therefore network average delay does not necessarily increase as p_0 increases. In Fig. 3.10, for example, network delay under Rician channels starts to decrease when p_0 exceeds some value. Of course, most successful packets are transmitted by near users.

3.5 Conclusion

The performance of mobile slotted ALOHA networks under 4 different channel models are compared by simulation. It is found that when the channel load is light (heavy), 1) multipath fading and shadowing effect will degrade (improve) network performance, and 2) network performance due to different channel models varies insignificantly (considerably). In addition, the same average network performance may result from very different patterns of user performance.

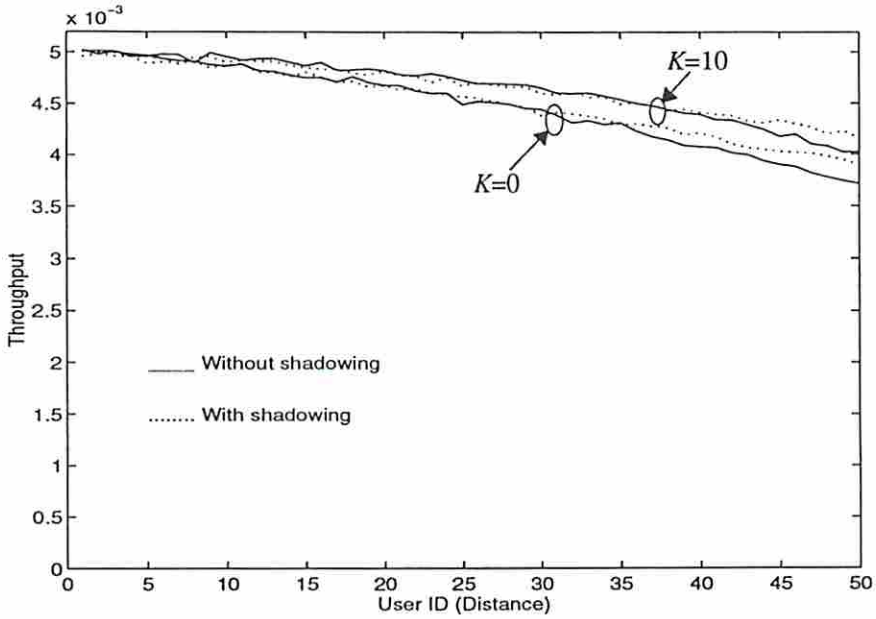


Fig. 3.1. Throughput of each user under 4 channel models. $p_0 = 0.005$, $q = 0.05$, and $\sigma_s = 1$ for shadowing.

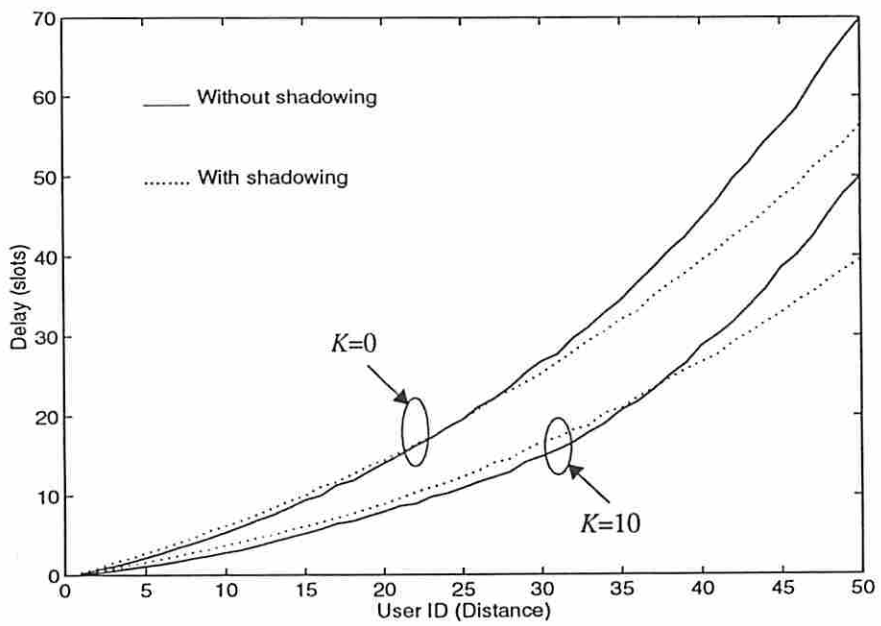


Fig. 3.2. Delay of each user under 4 channel models. $p_0 = 0.005$, $q = 0.05$, and $\sigma_s = 1$ for shadowing.

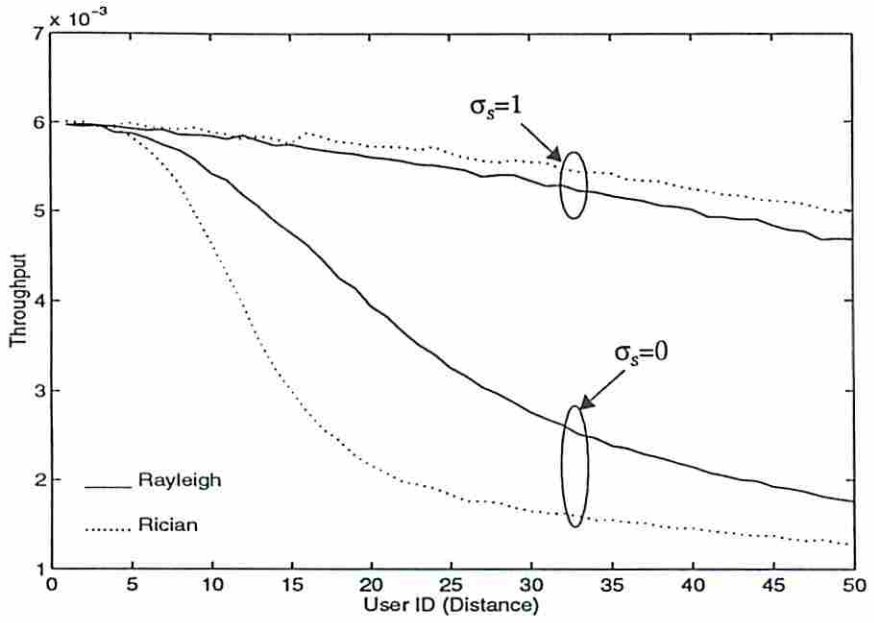


Fig. 3.3. Throughput of each user under 4 channel models. $p_0 = 0.006$, $q = 0.12$.

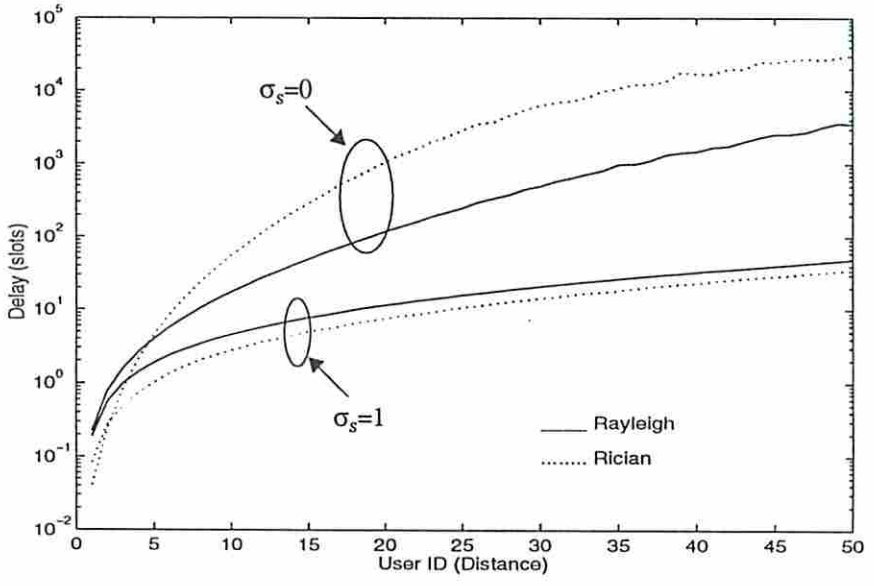


Fig. 3.4. Delay of each user under 4 channel models. $p_0 = 0.006$, $q = 0.12$.

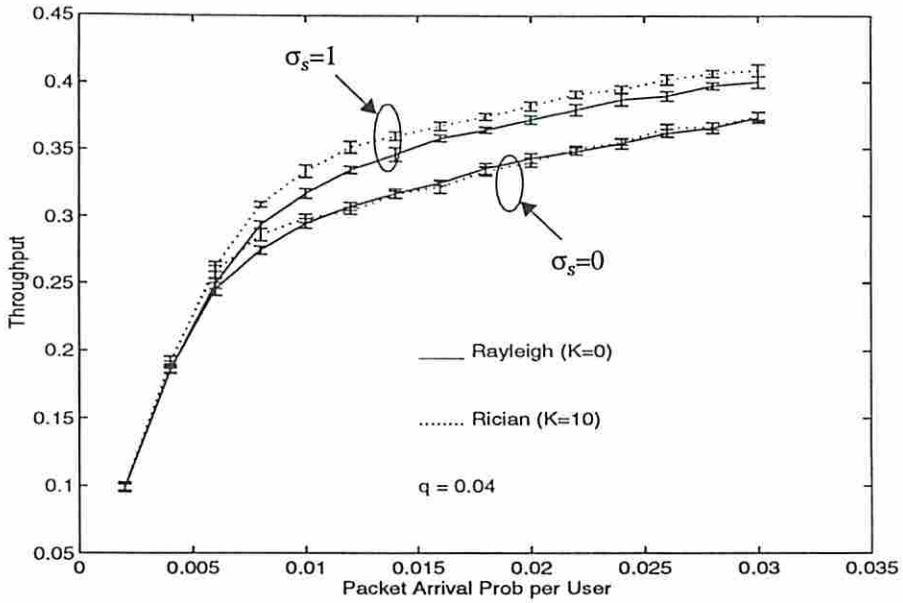


Fig. 3.5. Network throughput as a function of packet arrival probability p_0. The retransmission probability $q = 0.04$.

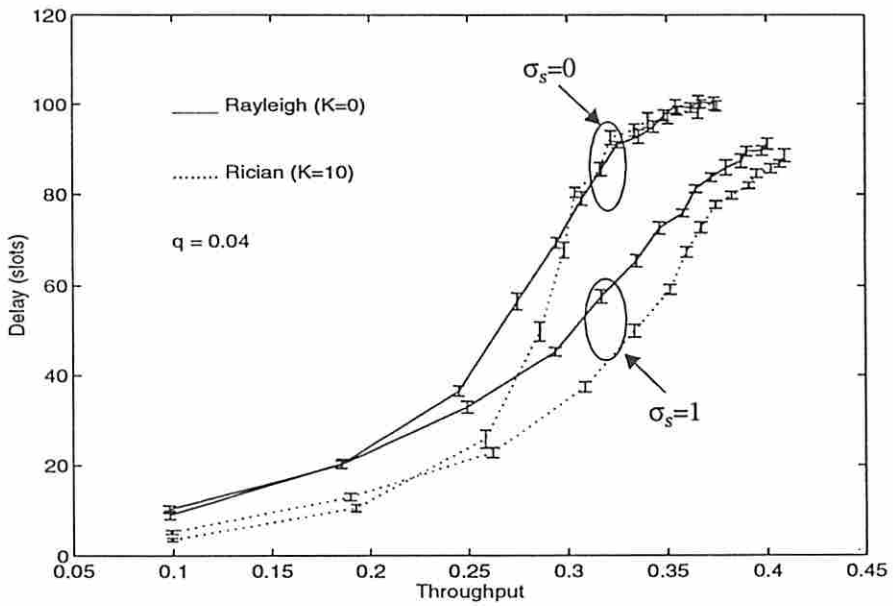


Fig. 3.6. Throughput-delay performance of 4 channels. The retransmission probability $q = 0.04$.

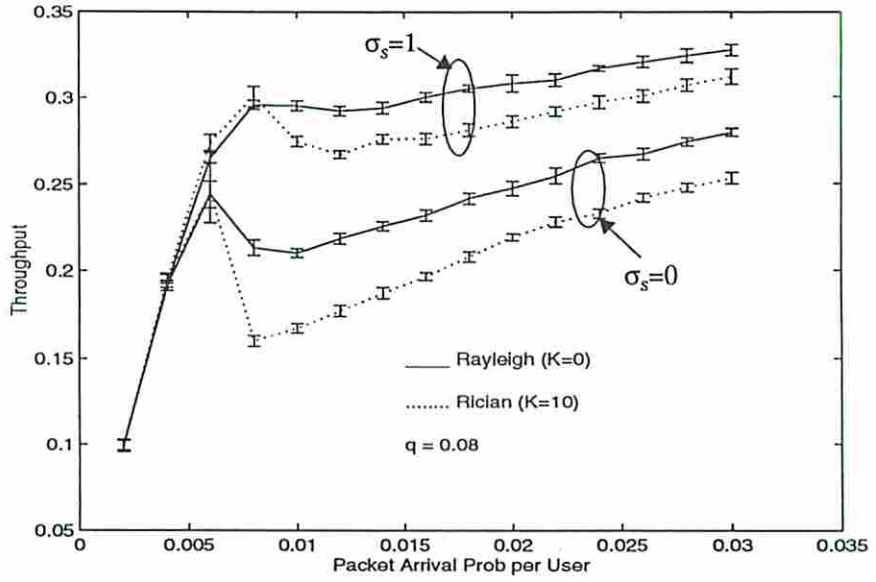


Fig. 3.7. Network throughput as a function of packet arrival probability p_0. The retransmission probability $q = 0.08$.

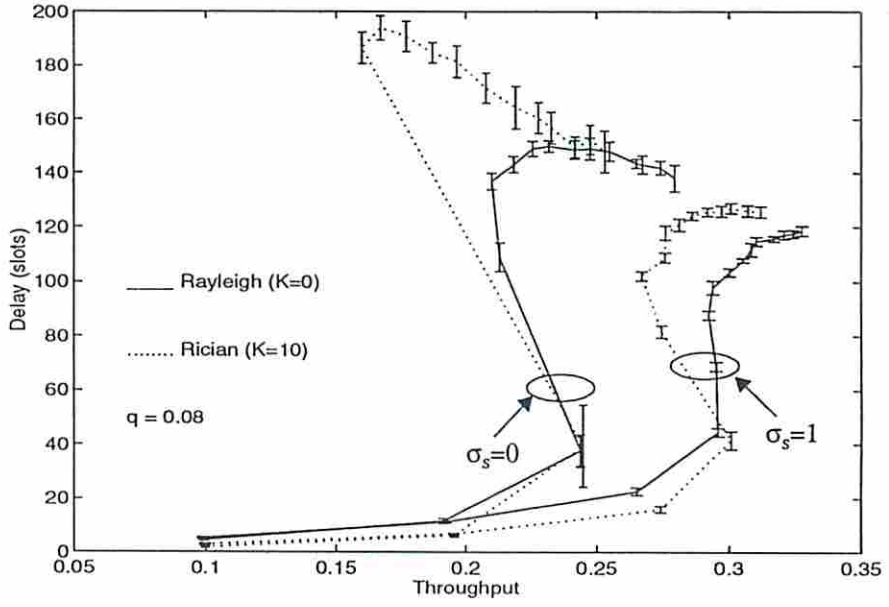


Fig. 3.8. Throughput-delay performance of 4 channels. The retransmission probability $q = 0.08$.

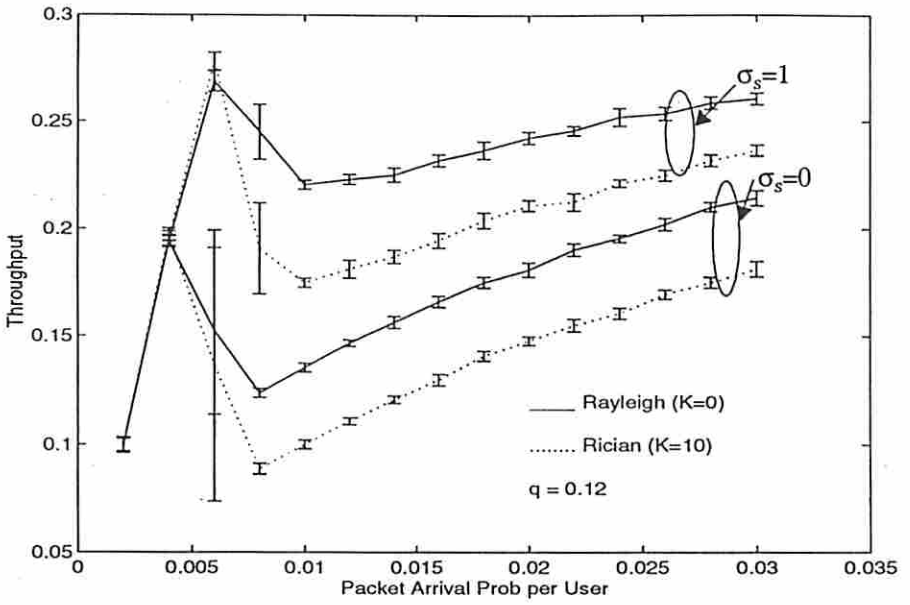


Fig. 3.9. Network throughput as a function of packet arrival probability p_0. The retransmission probability $q = 0.12$.

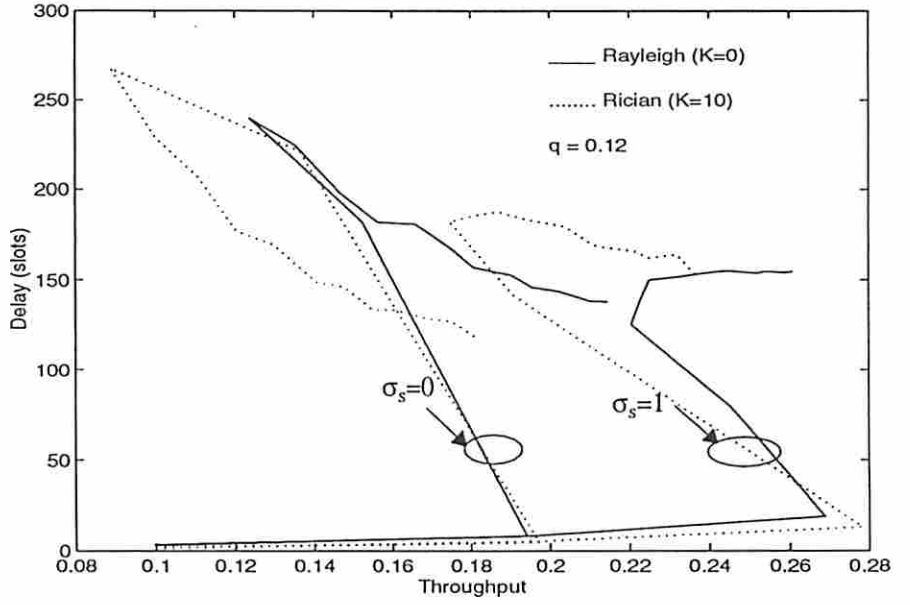


Fig. 3.10.Throughput-delay performance of 4 channels. The retransmission probability $q = 0.12$.

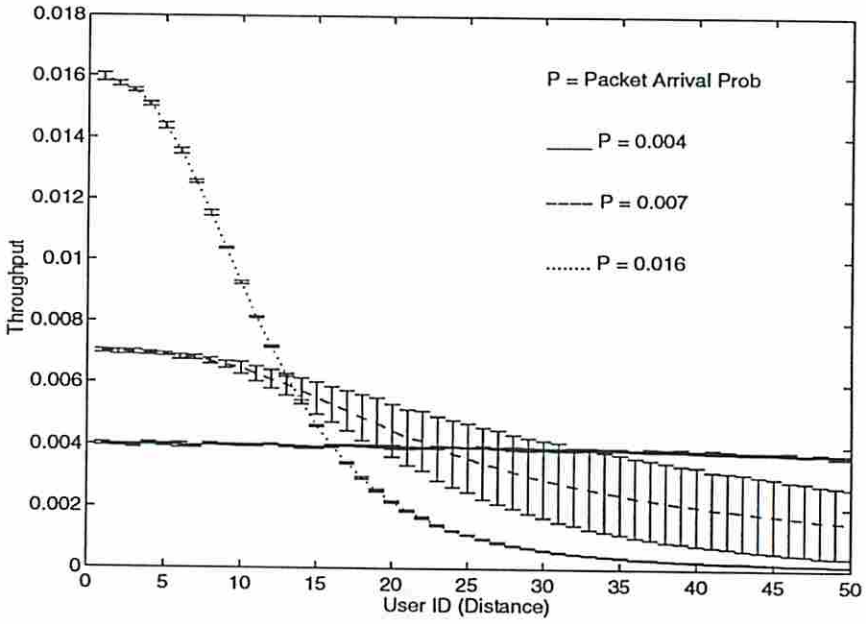


Fig. 3.11.Throughput of each user in Rician channel with $q = 0.08$. Network throughput is 0.2 approximately.

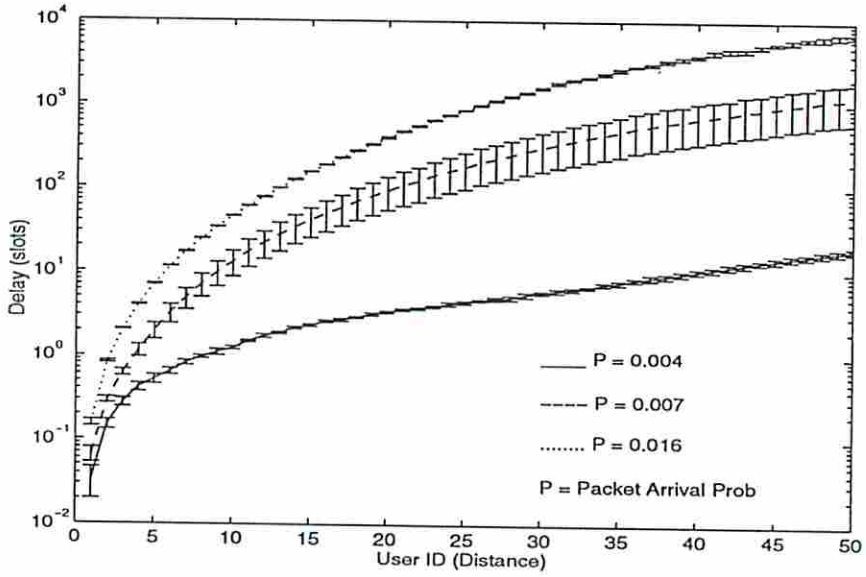


Fig. 3.12.Delay of each user in Rician channel with $q = 0.08$. Network average delay for $p_0 = 0.004$, 0.007, and 0.016 are 6.1, 125, and 191, respectively.

Chapter 4

Distributed Packet Radio Networks with Hard Real-time Communication Requirement

The performance of reservation ALOHA (R-ALOHA) to support hard real-time communications in a distributed packet radio network is evaluated analytically and by simulation. The network requirements are that each node send an update packet to its communication partner at least every T seconds (called the deadline period) with high reliability. Two scenarios are considered: the worst case and the steady-state case. For the worst-case scenario, two Markovian models are used to investigate how fast the network nodes can recover the communications after a short period of catastrophic failure of the network (i.e., recover reservation). For the steady-state scenario, a Markov chain is formulated to study the trade-off between the key system parameters such as the channel data rate, the packet error probability, the deadline period, and the reliability requirement. One application of this network is for vehicle-to-vehicle communications in intelligent vehicle highway systems where a vehicle periodically sends its updated status (velocity, acceleration, etc.) to the following vehicle to improve safety and the vehicular traffic capacity.

4.1 Introduction

Packet radio networks (PRN's) have been used for wireless data communications for two decades since the Defense Advanced Research Projects Agency (DARPA) initiated a research effort to develop a packet radio network in 1972 [26]. The goal of most existing PRN's is to provide an efficient and reliable packet transportation system in a typically noisy radio environment. Among the large number of issues involved in the design of PRN's, the efficient sharing of the common radio channel is one of the most important issues. Many channel access protocols have been considered, e.g., ALOHA [1], CSMA [29], etc. The typical performance measures are system throughput and average packet delay.

Recently there has been increased interest in applying advanced control and packet radio technology to highway systems in order to increase highway capacity. Karaasian et al. propose that vehicles in highways cruise in platoons, trains of vehicles with close spacing between consecutive vehicles (see Fig. 4.1) [24]. In a platoon, a vehicle needs to send its status information (velocity, acceleration, etc.) to its follower periodically so that its follower can keep a safe constant distance from it. Meanwhile, the follower sends an acknowledgment (ACK) back to the vehicle in front of it so that the front one may know that the follower has correctly received the status messages. This type of system requires that the communication between adjacent vehicles not suffer persistent interruption for more than a specified time interval (called the deadline period). For this kind of

applications, throughput-delay is not an appropriate performance measure for a channel access protocol. Instead, it is the reliability that determines the performance.

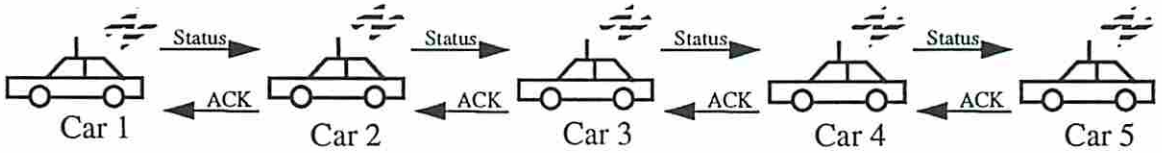


Fig. 4.1. Real-time vehicle-to-vehicle communications.

In this chapter, we evaluate the performance of R-ALOHA [11] in a distributed packet radio network (with hard real-time communications) such as the vehicle-to-vehicle communication system in highways. Deadline failure probability (*DFP*), which will be introduced in the next section, is used as the reliability measure. Two scenarios are considered: the worst case and the steady-state case. Section 4.3 considers the worst-case scenario. Two Markovian models are proposed to investigate how fast the network nodes (e.g. vehicles) can recover the communications after a short period of catastrophic failure of the network (i.e. recover reservation). The difference between the two models is in the assumption about the capability of distinguishing collision from successful transmission. The steady-state scenario is considered in Section 4.4. A Markov chain model is proposed to study the trade-off between the key system parameters including the channel data rate, the packet error probability, the deadline period, and the reliability requirement. Numerical results of the models for both of the worst and steady-state cases are presented in Section 4.5.

4.2 R-ALOHA and Deadline Failure Probability

In R-ALOHA, channel time is slotted with N consecutive slots constituting a frame. Each user transmits once in a frame. By continuously monitoring the activity of slots, a user randomly chooses an empty slot to transmit its status packet and it checks whether the transmission was successful by the ACK from the receiver. If successful, the sender will become the owner of this slot in the next and subsequent frames. If no ACK is received, however, it assumes a collision and switches to another (empty) slot for the next frame. The ACK's are assumed to be carried on a different channel either by TDM or FDM.

The performance measure of interest is the probability that a user does not receive the status of its partner for more than a specified number of frames. This probability is designated as the deadline failure probability (*DFP*). Most of the time, a user will stick to a certain slot in a frame to transmit its status packets. However, packet errors may occur due to channel fading or collision. Although R-ALOHA is a reservation-based protocol, a collision may still occur when users using the same slot move into the interfering range of one another. When an error is detected, the users involved will switch to other slots that are currently recorded as empty in the next frame. If the number of users that switch simultaneously is small, the probability of an error occurrence is low. The *DFP* is the probability that more than F consecutive packet errors occurs.

4.3 The Worst Case Scenario

The worst case is when all of the active slots in a frame are corrupted at the same time. All of the users will switch to other slots in the following frames until everyone successfully locks onto a slot. For this case, we study how fast the users in the network can recover reservation.

Assume that there are M users in the system and each is within the interference region of one another so that any simultaneous transmissions will cause mutual collision. In addition, each user must transmit once per N-slot frame ($N > M$). Initially each user randomly chooses one of the N slots. After the first trial, some of the slots are used by exactly one user (success slots), some are left empty (empty slots), and the others are used by more than one user (collision slots). Success slots will be owned in succeeding frames, whereas empty slots are open for contention by users involved in a collision. In reality, a collision slot is usually not distinguishable from a success slot for the users who did not transmit in this slot. For simplicity, the first model assumes that collision slots *are* known to all users and can therefore be used in the next frame. The case where the collision slots are not distinguishable from success slots and hence cannot be used immediately in the next frame is considered in the second model. We define the system stabilization time $SST(M, N)$ to be the number of the frames that elapse until everyone in the system has successfully locked onto a slot (for M users in the system and N slots in a frame). Furthermore, we define the first success time $FST(M, N)$ to be the number of the frames

that elapse until a particular user succeeds. By analyzing the underlying combinatorial problem, the distributions and mean values of these two random variables will be derived for both models. It is worth mentioning that Mathar and Mann have studied *SST(M, N)*– they found the mean but not the distribution [40].

4.3.1 Model 1

Let $p[k \mid m, n]$ be the probability of k success slots given that each of the m users randomly selects one of the n slots, and $b[m, i, x]$ be the binomial probability that i out of m users randomly with probability x choose a particular slot. $b[m, i, x]$ is given by

$$b[m, i, x] = \binom{m}{i} x^i (1-x)^{m-i}, \text{ where } \binom{m}{i} = \frac{m!}{i!\,(m-i)!}.$$

By conditioning on the number of transmissions in the first of the n available slots, $p[k \mid m, n]$ is given by the recursive formula as follows.

$$\begin{aligned} p[k|m,n] &= b[m,0,1/n]\,p[k|m,n-1] + b[m,1,1/n]\,p[k-1|m-1,n-1] \\ &\quad + \sum_{i=2}^{m} b[m,i,1/n]\,p[k|m-i,n-1], \end{aligned} \tag{4.1}$$

where $n \geq 1$, $m \geq 0$, $m \geq k \geq 0$.

The boundary conditions of (4.1) are

$$p[k|m,n] = \begin{cases} 1, & \text{if } n = 1, m = 0, k = 0 \\ 0, & \text{if } n = 1, m = 1, k = 0 \\ 1, & \text{if } n = 1, m = 1, k = 1 \\ 1, & \text{if } n = 1, m \geq 2, k = 0 \\ 0, & \text{if } n = 1, m \geq 2, k > 0 \\ 1, & \text{if } n \geq 2, m = 0, k = 0 \\ 0, & \text{if } n \geq 2, m = 1, k = 0 \\ 1, & \text{if } n \geq 2, m = 1, k = 1. \end{cases}$$

Define X_f to be the number of the success slots after the fth frame, $f = 1, 2, 3, \ldots$. We have the probability of k success after the first frame's contention given by

$$Pr\{X_1 = k\} = p[k|M, N], \qquad 0 \leq k \leq M.$$

With the distribution of X_1, we can compute the distribution of X_f $(f \geq 2)$ iteratively. That is,

$$Pr\{X_f = k | X_{f-1} = j\} = p[k-j|M-j, N-j],$$

$0 \leq j \leq k \leq M$.

By averaging over the distribution of X_{f-1}, we get the probability of k successes after the fth frame's contention by

$$Pr\{X_f = k\} = \sum_{j=0}^{k} Pr\{X_{f-1} = j\} p[k-j|M-j, N-j], \qquad 0 \leq k \leq M.$$

With the distribution of X_f, we can compute the cumulative distribution functions of SST and FST in the sequel.

$$Pr\{SST \leq f\} = Pr\{X_f = M\}, \qquad f \geq 1.$$

$$Pr\{FST \le f\} = \sum_{j=1}^{M} Pr\{X_f = j\} \frac{j}{M} = \frac{1}{M} E\{X_f\}, \qquad f \ge 1.$$

With the distribution of *SST* and *FST*, the mean values of *SST* an *FST* can be computed numerically. Nevertheless, the following recursive relations provide a faster way to get the mean values of *SST* and *FST*. The mean of *SST*, $E\{SST|M, N\}$, can be obtained by computing $E\{SST|0, N-M\}$, $E\{SST|1, N-M+1\}$, $E\{SST|2, N-M+2\}$, and so on, iteratively. The recursive formula to compute the above is

$$E\{SST|m, n\} = \sum_{k=0}^{m} p[k|m, n] \, (1 + E\{SST|m-k, n-k\}),$$

$0 \le m \le n$,

with the boundary conditions given by $E\{SST \mid 0, n\} = 0$ for $n \ge 0$. Similarly, the mean of *FST*, $E\{FST|M, N\}$, can be obtained by using the recursive form

$$E\{FST|m, n\} = \sum_{k=0}^{m} p[k|m, n] \left(\frac{k}{m} + \frac{m-k}{m} (1 + E\{FST|m-k, n-k\}) \right),$$

$0 \le m \le n$,

with the boundary condition as $E\{FST \mid 0, n\} = 0$ for $n \ge 0$. The worst-case *DFP* given by model 1 can be computed by

$$Pr[FST \ge \text{Deadline Period}], \qquad (4.2)$$

where the value of the deadline period is provided by the communications requirements of the system.

4.3.2 Model 2

In this model, collision slots cannot be used in the next frame but will become available one frame later. Let $p[k, j \mid m, n]$ be the probability of k successful slots and j empty slots given that each of the m users randomly selects one of the n slots, and $b[m, i, x]$ be the same as defined before. By looking at what happens in the first of the n available slots, we can get $p[k, j \mid m, n]$ by the recursive formula as follows.

$$\begin{aligned} p[k,j|m,n] &= b[m,0,1/n]\,p[k,j-1|m,n-1] \\ &\quad + b[m,1,1/n]\,p[k-1,j|m-1,n-1] \\ &\quad + \sum_{i=2}^{m} b[m,i,1/n]\,p[k,j|m-i,n-1], \end{aligned} \tag{4.3}$$

for $n \geq 1, m \geq 0, m \geq k \geq 0, n - \delta(m) \geq j \geq max(n-m, 0)$,

where $\delta(m) = \begin{cases} 0, & \text{for } m = 0, \\ 1, & \text{for } m > 0, \end{cases}$ and $max(a,b) = \begin{cases} a, & \text{if } a \geq b, \\ b, & \text{otherwise.} \end{cases}$

The boundary conditions of (4.3) are

$$p[k,j|m,n] = \begin{cases} 1, & \text{if } n = 1, m = 0, k = 0, j = 1 \\ 0, & \text{if } n = 1, m = 1, k = 0, j = 0 \\ 1, & \text{if } n = 1, m = 1, k = 1, j = 0 \\ 1, & \text{if } n = 1, m \geq 2, k = 0, j = 0 \\ 0, & \text{if } n = 1, m \geq 2, k > 0, j = 0 \\ 1, & \text{if } n \geq 2, m = 0, k = 0, j = n \\ 0, & \text{if } n \geq 2, m = 1, k = 0, j = n-1 \\ 1, & \text{if } n \geq 2, m = 1, k = 1, j = n-1. \end{cases}$$

Define X_f and E_f to be the number of successful slots and the number of empty slots after the fth frame, respectively. We have

$$Pr\{X_1 = k, E_1 = j\} = p[k, j|M, N],$$

$0 \le k \le M, 0 \le j \le N$.

Then the joint density of X_f and E_f can be computed by conditioning on X_{f-1} and E_{f-1} as follows.

$$Pr\{X_f = k, E_f = j\} = \sum_{u=0}^{k} \sum_{v=N-u-j}^{N-u} Pr\{X_{f-1} = u, E_{f-1} = v\}$$

$$\cdot p[k-u, j-(N-u-v)|M-u, v], \quad f \ge 2, 0 \le k \le M, 0 \le j \le N,$$

Note that $(N\text{-}u\text{-}v)$ is the number of collision slots in frame $(f-1)$, which will become empty in frame f. Then the marginal distribution of X_f is given by

$$Pr\{X_f = k\} = \sum_{j=0}^{N} Pr\{X_f = k, E_f = j\}, \qquad 0 \le k \le M.$$

With the distribution of X_f, the cumulative distribution functions of SST and FST are given by

$$Pr\{SST \le f\} = Pr\{X_f = M\}, \qquad f \ge 1.$$

$$Pr\{FST \le f\} = \sum_{j=1}^{M} Pr\{X_f = j\} \frac{j}{M} = \frac{1}{M} E\{X_f\}, \qquad f \ge 1.$$

With the distribution of SST and FST, the mean values can be computed numerically. Alternatively, the following recursive formulae provide another way to get the mean values. For the mean of SST, $E\{SST|M, N\}$ can be found by solving a linear equations

system of size MN as follows.

$$E\{SST|m,n\} = \sum_{k=0}^{m} \sum_{j=n-m}^{n} p[k,j|m,n] \cdot (1 + E\{SST|m-k, j+N-n-M+m\}), \quad (4.4)$$

where $0 \le m \le n, 0 \le m \le M, 0 \le n \le N$.

The boundary condition of (4.4) is $E\{SST|0,n\} = 0$ for $n \ge 0$. For the mean of FST, $E\{FST|M,N\}$ can be obtained by solving a similar set of linear equations system shown below.

$$E\{FST|m,n\} = \sum_{k=0}^{m} \sum_{j=n-m}^{n} p[k,j|m,n] \cdot \left(\frac{k}{m} + \frac{m-k}{m}(1 + E\{FST|m-k, j+N-n-M+m\})\right), \quad (4.5)$$

where $0 \le m \le n, 0 \le m \le M, 0 \le n \le N$. The boundary condition of (4.5) is $E\{FST|0,n\} = 0$ for $n \ge 0$. As by model 1, the DFP of the worst case given by model 2 can be computed by (4.2).

4.4 The Steady-state Scenario

In the worst-case analysis just presented, it is assumed that a packet error occurs only if two or more users simultaneously switch to the same slot. Once a user makes a successful transmission in a slot, the following transmissions in that slot will always be successful. In this section, we assume that an error occurs in a reserved slot with probability P_e which

accounts for the impact of channel fading and collision (due to user mobility) on the packet transmission. We define the recovery time (RT) to be the number of frames it takes for a user to lock onto a slot, once the user has a packet error. Suppose we know the distribution of RT, the DFP is

$$P_e \cdot Pr[RT \geq \text{Deadline Period}]. \tag{4.6}$$

In the following, we formulate a Markov chain to find the mean of RT. Then the distribution of RT is obtained by approximating RT as a geometrically distributed random variable.

4.4.1 The Markov Chain Analysis

Similar to the worst-case analysis, it is assumed that there are M users in the system, each one is within the interference region of each other, and every user transmits once per frame of N slots ($N > M$). Nevertheless, errors will occur in a reserved slot with probability P_e. In the steady state, a user will be in one of the two states: the good state and the bad state. A user is in the good state if it gets a reserved slot and a packet error does not occur. If a packet error does occur due to fading or collision, the user will switch to an empty slot in the next frame. The switch will be successful if there is no other users who switch to the same slot simultaneously, i.e., we neglect the possibility of getting an error due to channel fading while a user is switching.

Let X_f be the number of users who need to switch at the end of frame f, and Y_f be the number of empty slots in frame f. There are (M - X_f) users in the good state and [N - (M -

X_f) - Y_f] collision slots in frame f. The collision slots in frame f cannot be used in frame (f + 1) because they are indistinguishable from success slots. But they will become available in frame (f + 2).

In frame (f + 1), the X_f users will compete for the Y_f empty slots. In addition, for the (M - X_f) users in good state, each one have a packet error in frame (f + 1) with probability P_e. Let A be the number of successful slots, B be the number of empty slots after the competition, and C be the number of users who have a packet error in frame (f + 1). Thus X_{f+1} and Y_{f+1} can be expressed as (X_f - A) + C and B + [N - (M - X_f) - Y_f], respectively. The random variables A and B have the joint conditional probability distribution

$$Pr[A = k, B = j | X_f = m, Y_f = n] = p[k, j \mid m, n],$$

and the random variable C has the conditional distribution

$$Pr[C = i | (M - X_f) = m] = \binom{m}{i} P_e^i (1 - P_e)^{m-i}.$$

Although the transition probabilities of the two dimensional chain (X, Y) can be computed and the equilibrium state probabilities can be solved accordingly, we do not try to solve the above 2-dimensional Markov chain. Instead, we propose an approximate one-dimensional Markov chain with state variable X_f being the number of users who need to switch in frame (f + 1). For the one-dimensional Markov chain, we assume that there are $\lfloor X_f/2 \rfloor$ collision slots in frame f. Therefore, Y_f can be approximated by $Y_f' = N - (M - X_f) - \lfloor X_f/2 \rfloor$, where $\lfloor \ \rfloor$ is the floor function which returns the largest integer that is smaller

than or equal to the argument. Then we have $X_{f+1} = X_f - A' + C$, where A' is the number of success slots given X_f users competing for Y_f' slots. The transition probabilities is given by

$$Pr[X_{f+1} = j | X_f = i] = Pr[(A' - C) = (i - j) | X_f = i]$$

$$= \sum_{k=0}^{i} Pr[A' = k | X_f = i] \cdot Pr[C = k - i + j | A' = k, X_f = i]$$

$$= \sum_{k=0}^{i} p[k | i, Y'_f] \cdot b[M - i, k - i + j, P_e],$$

where $Y_f' = [N - (M - i) - \lfloor i/2 \rfloor]$.

Now we can solve the equilibrium state probabilities, $Pr[X = k]$, $k = 0, 1, \ldots, M$, and compute the mean of X, denoted by $E\{X\}$. By Little's formula, the expected time in the bad state, $E\{RT\}$, is given by

$$E\{RT\} = E\{X\} / S,$$

where S is the expected number of users leaving the bad state per frame. Define S_k to be the expected number of users leaving the bad state given $X = k$. We have

$$S = \sum_{k=0}^{M} S_k Pr[X = k],$$

and

$$S_k = \sum_{j=0}^{k} j \cdot p[j | k, N - (M - k) - \lfloor k/2 \rfloor].$$

By approximating RT as a geometrically distributed random variable with the mean equal to $1/\mu$, the distribution of RT is given by $Pr[RT = k] = \mu(1-\mu)^{k-1}$, $k = 1, 2, \ldots, \infty$. Suppose the communications requirement specifies that deadline period is F frames. Then

the *DFP* can be computed by

$$DFP = P_e \cdot Pr[RT \geq \text{Deadline Period}]$$

$$= P_e \cdot (1-\mu)^{F-1}.$$

This gives the relations among the key system parameters such as channel data rate (proportional to N), the packet error probability (P_e), the deadline period (F frames), and the reliability requirement (*DFP*).

4.5 Numerical Results and Discussions

In this section, two networks are considered: network A and network B. Network A shown in Fig. 4.2 consists of 9 cars ($M = 9$) where every car is within the interfering range of any other, i.e., any two or more simultaneous transmissions in a slot will result in a collision.

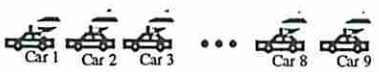


Fig. 4.2. Network A: Each and every of the 9 cars interferes and is interfered by all other cars.

Network B shown in Fig. 4.3 consists of a hundred cars in a linear highway where only the closest eight cars can interfere with a particular car. For example, cars 11-14 and 16-19 are the potential interfering sources of car 15; cars 12-15 and 17-20 are the potential interfering sources of car 16.

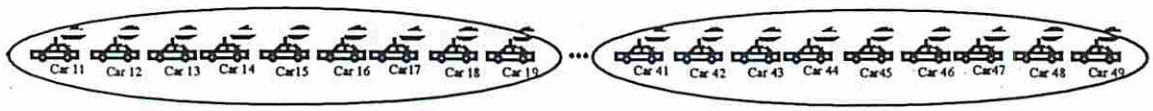


Fig. 4.3. Network B: The interference range of a particular car covers the closest 4 cars in front and the closest 4 cars behind it.

Network A and network B are similar in that every car can only detect the existence of 8 interfering users. We use computer simulation to check if the performance of network B can be well approximated by the analytical model of network A. Note that both networks use R-ALOHA with N-slot frames.

Figs. 4.4-4.7 concern the worst-case scenario; Figs. 4.8-4.10 consider the steady-state scenario. Fig. 4.4 shows the probability distribution of the system stabilization time (SST) for network A in the worst case. It can be seen that the distribution of SST for model 1 is very close to that for model 2. Simulation results also verify this observation. It implies that when N (= 16 in this case) is large compared with M (= 9), the assumption that collision slots can be used immediately in the next frame does not significantly affect the accuracy of model 1, but the computation is greatly simplified. Fig. 4.5 shows the mean values of SST of model 1 as a function of M and N. As expected, the mean of SST decreases as N increases or as M decreases.

Fig. 4.6 shows the probability distribution of the first success time (FST) for network B in the worst case. The good match between analytical and simulation results confirms that network A is a good approximation of network B as far as FST is concerned. Fig. 4.7

shows the mean of *FST* of model 1 of network A as a function of M and N. Similar to the mean of *SST*, the mean of *FST* decreases as N increases or as M decreases.

For the steady-state scenario, the probability distribution of the recovery time (*RT*) is shown in Fig. 4.8. The almost perfect match between the simulation results for networks A and B implies that network A is indeed a good approximation of network B. Good agreement between the analytical and simulation results is also observed. Fig. 4.9 shows the analytical results for the mean of *RT* as a function of N, which is proportional to the channel data rate given that the packet size (in bits) is fixed. As expected, mean *RT* decreases as N increases.

Fig. 4.10 shows the analytical results for deadline failure probability (*DFP*) as a function of N and P_e. It can be seen that *DFP* decreases as N increases or as P_e decreases. One interesting trade-off arises by increasing the channel data rate, we can either have larger N by fixing the packet size or achieve smaller P_e by using more powerful error correction codes. Which way is more beneficial is subject to more study.

4.6 Conclusion

A new performance measure, deadline failure probability (*DFP*), is used to evaluate the performance of R-ALOHA as applied to packet radio networks to support hard real-time communications. Both the worst and the steady-state cases are considered. In the worst-case analysis, two Markovian models are used to investigate how fast the network nodes

can recover the communications after a short period of catastrophic failure of the network. Whereas model 1 of the worst case is simpler, model 2 is more realistic. In the steady-state analysis, a Markov chain is formulated to study the time spent to regain reservation once a packet error occurs. The proposed models can be applied as tools to trade-off the system reliability with the important system parameters such as the channel data rate and the packet error probability.

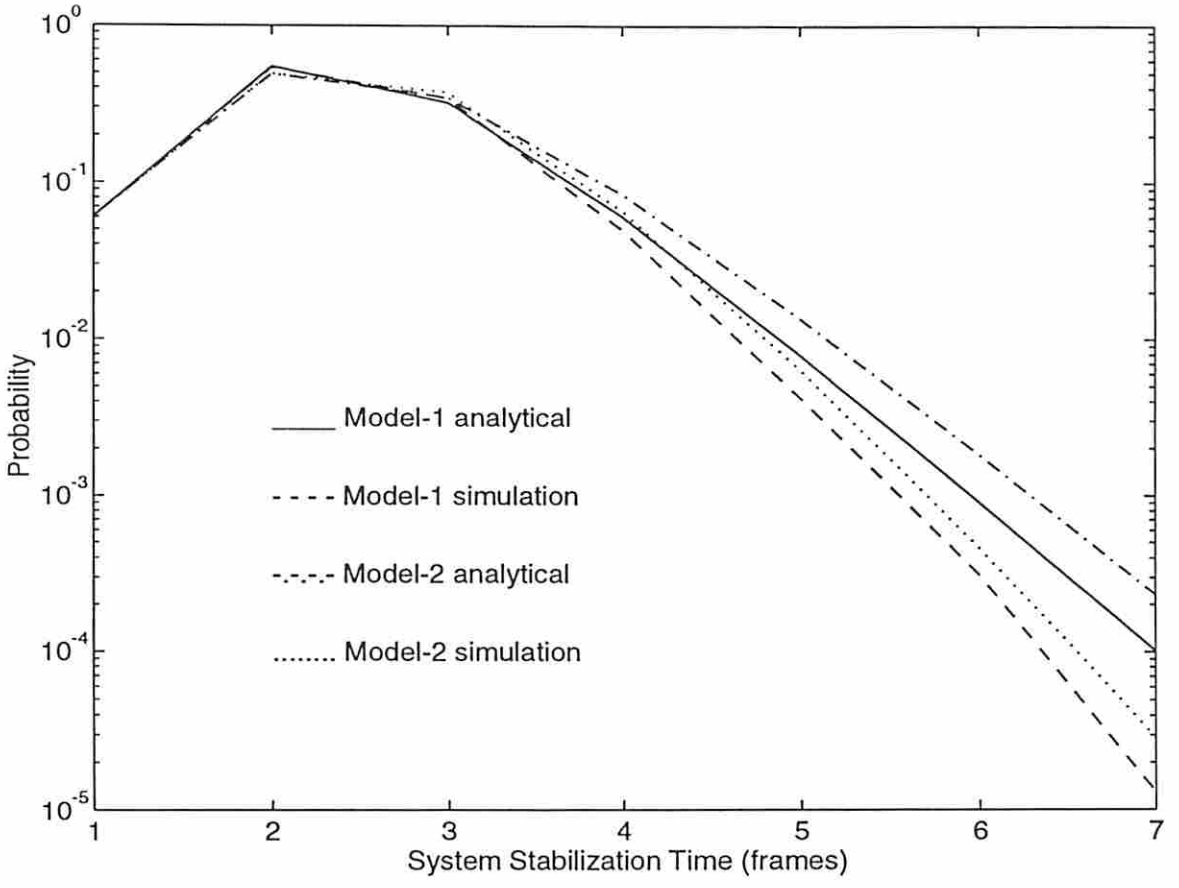


Fig. 4.4. The probability distribution of the system stabilization time for network A in the worst case. The number of users in the system is 9; the number of slots in a frame is 16.

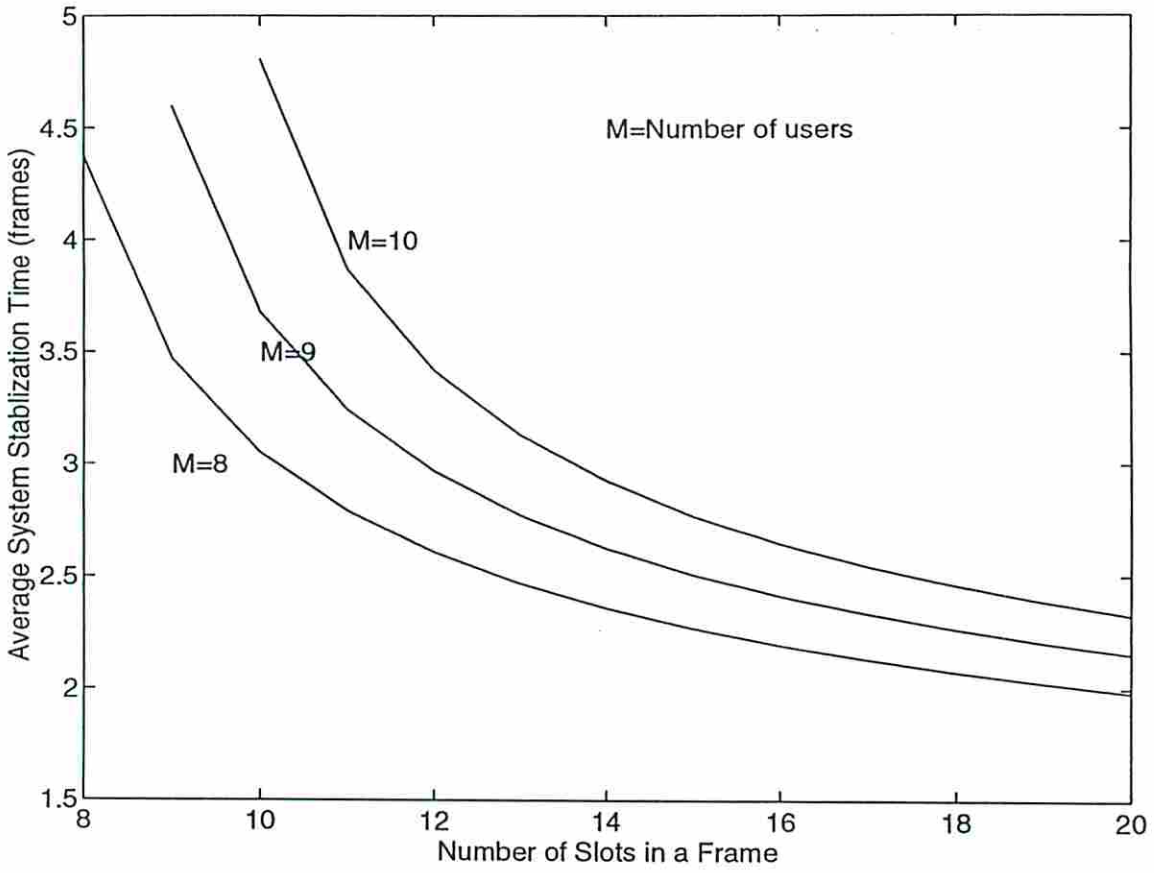


Fig. 4.5. The average system stabilization time as a function of M (number of users) and N (number of slots in a frame) for network A in the worst case.

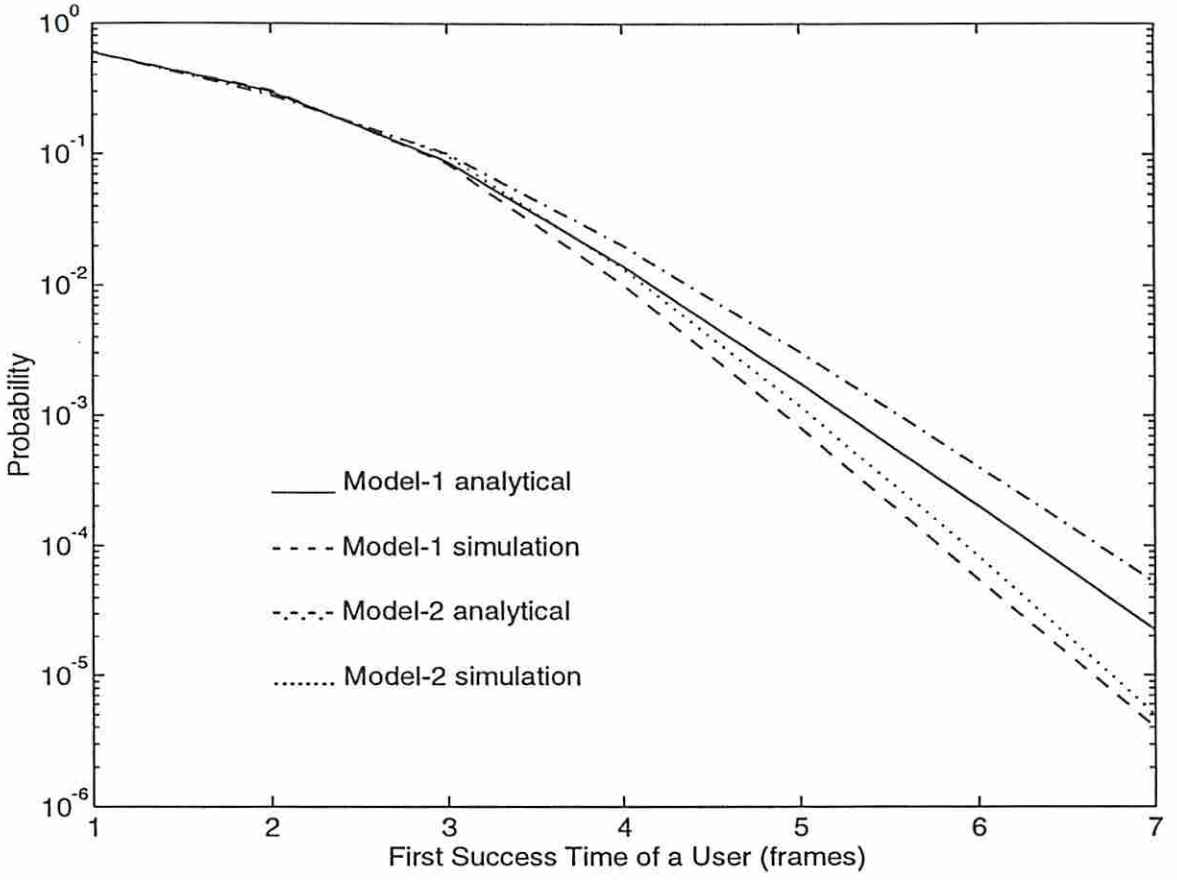


Fig. 4.6. The probability distribution of the first success time for network B in the worst case. The number of users in the interference range is 9; the number of slots in a frame is 16.

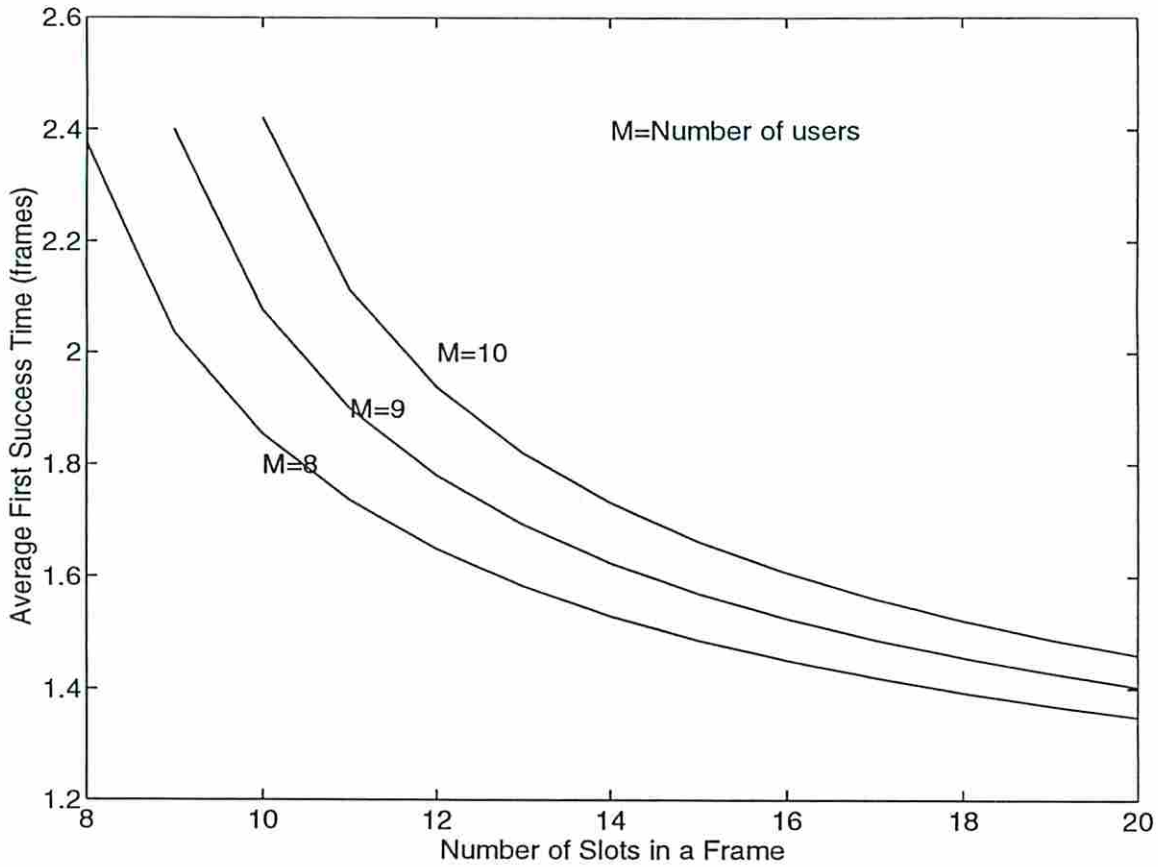


Fig. 4.7. The average first success time for various values of M (number of users within an interference range) and N (number of slots in a frame) for model 1 of network A in the worst case.

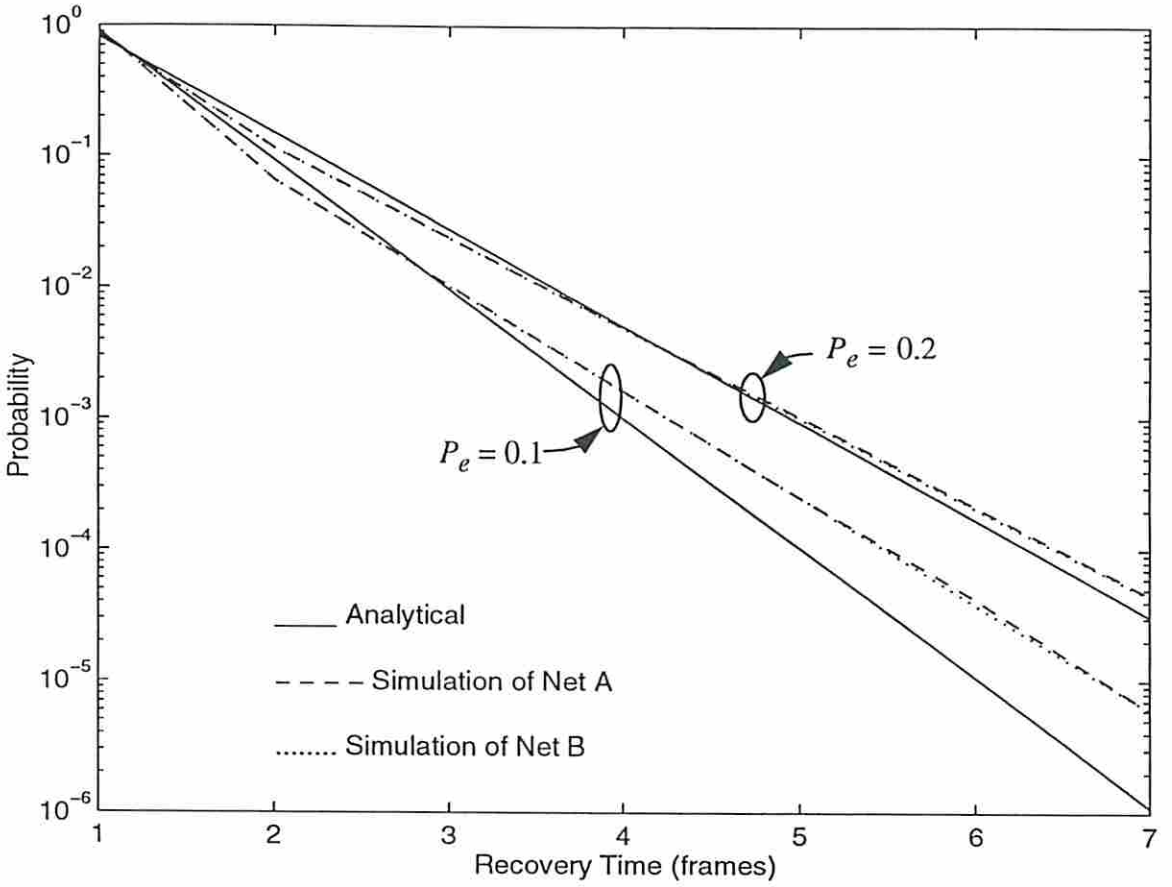


Fig. 4.8. The probability distribution of the recovery time with the packet error probability $P_e = 0.1$ and 0.2 for networks A and B in the steady state.

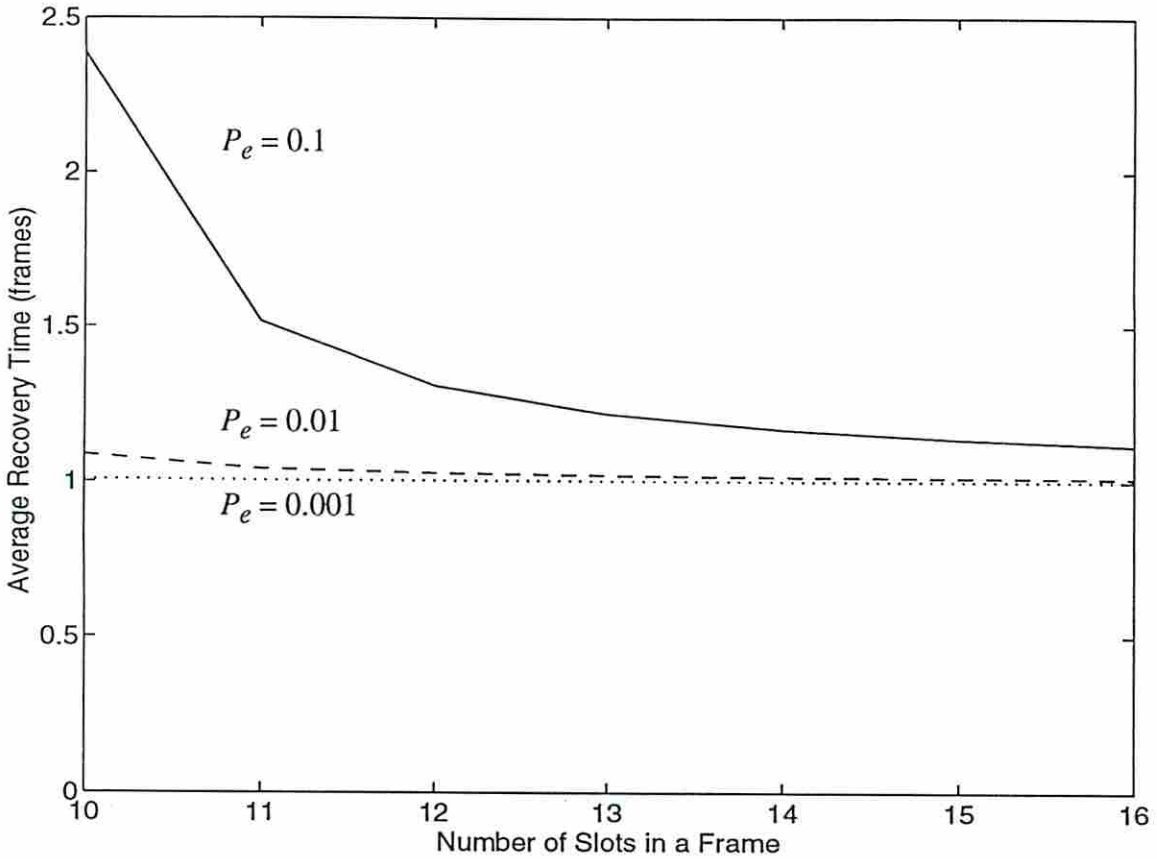


Fig. 4.9. Average recovery time as a function of the number of slots in a frame and the packet error probability (P_e) for network A in the steady state.

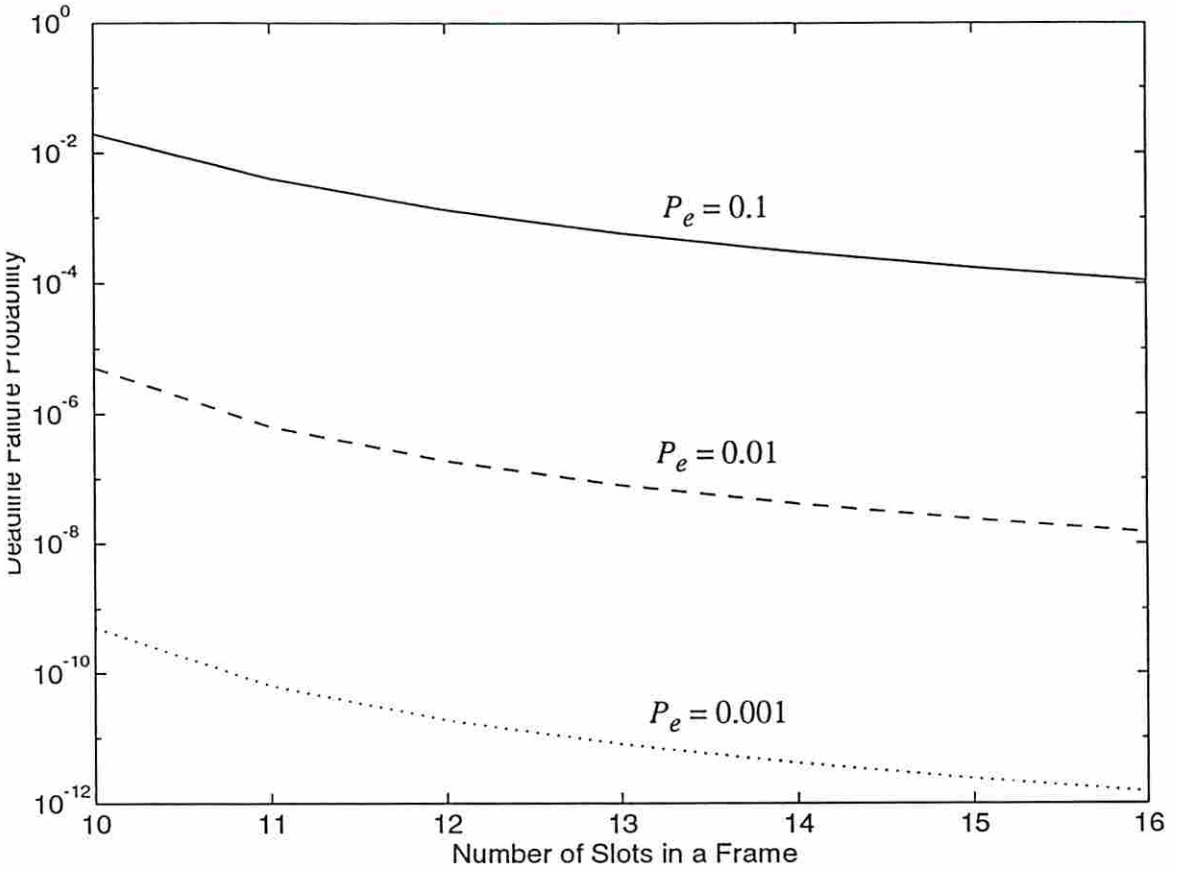


Fig. 4.10. The deadline failure probability as a function of the number of slots in a frame and the packet error probability (P_e) for network A in the steady state. The deadline period is assumed to be 4 frames.

Chapter 5

Congestion Control Policies for Quality of Service Guarantee in Wireless CDMA Networks Supporting Integrated Voice/Video/Data Services

In this chapter, congestion control policies are introduced and analyzed for the integrated voice, video and data traffic in the uplink (mobile to base station) of packet-switched wireless CDMA networks supporting personal communication services. These schemes allows data users to fully utilize the residual channel capacity left by real-time traffic (i.e., voice and video), while guaranteeing the quality of real-time services which are subject to admission control only. A Markovian model is developed for the integrated traffic and the performance measure of the system such as the bit error rate and the throughput are evaluated. This model requires numerical solution of a large Markov chain. Based on this model, two congestion control policies are derived according to the available feedback information from base stations to data users. The analysis of the congestion control policies helps in determination of optimal control parameters and in engineering the admission control policy.

5.1 Introduction

The research and development of personal communication systems (PCS) has been a very active area recently. The objective of PCS is to provide integrated services (i.e., voice, video, and data) to users on the move. One of the major challenges is the design of efficient network control schemes which can accommodate services of heterogeneous bit rates and quality of service (QoS) requirements for the channel from mobile users to base stations (the uplink). In particular, real- time services such as interactive voice/video need guaranteed delay but can tolerate certain loss of packets, whereas data services usually do not impose delay constraints and thus are queueable at mobile users. Moreover, the bit rate requirement of a video user is much higher than that of a voice user.

To provide QoS guarantees for real-time services such as voice and video, admission control is required to make sure that the system has sufficient resources (e.g., bandwidth) for the new and existing calls before the system can accept a new incoming call. Since the real-time traffic sources will not always transmit information at their peak rates, there will be some bandwidth unused from time to time. This residual bandwidth can be used to service data users provided that the bandwidth can be quickly released if real-time services need it again. The dynamic bandwidth allocation for data users is called congestion control.

The integration of voice and data services have been investigated in some depth, e.g., [14], [39] and [58]. Call level admission control for voice/data services has been

considered in [58]. Our focus is on the short-term traffic control in the uplink, i.e., burst level congestion control. A multi-code CDMA system (MC-CDMA) [16] for the integration of voice, data, and video services is considered. In MC-CDMA, a code can be used to transmit information at a basic bit rate. Users that need higher transmission rates can use multiple codes simultaneously. The higher data rate generated by multiple codes manifests itself as higher transmission power in order to keep the energy per bit (E_b) constant. Therefore, higher rate sources such as video users will generate more interference power compared to lower rate sources such as voice and data users.

In this chapter, we have identified a set of parameters, called the conditional link quality, which summarize the effect of the physical layer parameters (spreading/ modulation/coding) on the performance of congestion control policies. A Markovian model is developed for the integrated voice/video/data traffic under 2 congestion control policies with complete or limited feedback information about the system state. Two performance measures are used for the evaluation of these control policies, namely the data throughput and the average bit error rate, which could be traffic-type dependent. resulting model is a large Markov model requiring numerical solution.

Section 5.2 introduces the system model for the integrated traffic CDMA network. Component models including the traffic models for voice, video and data users, and the model for CDMA link quality are described. A Markovian model is presented and the performance measures of the CDMA system are evaluated. In Section 5.3, we develop congestion control policies for 2 types of feedback information, i.e., complete and limited

feedback. The approach to optimize the control policies is also described. Numerical results are presented in Section 5.4 and conclusions are given in Section 5.5.

5.2 System Model

The wireless CDMA network consists of M_{vo} voice users, M_{vi} video users, and M_d data users. These users have been admitted to the system by a given admission control policy. The time axis of the CDMA system is fully slotted, with a slot duration equal to the transmission time of a packet which is the same size for all types of users. All users are synchronized at the packet level, i.e., they have to transmit at the beginning of time slots.

5.2.1. Traffic Model for a Voice User

The model for a single voice source is the On-Off model proposed by Brady [5]. In the On-Off model, a voice source alternates between the On-state and the Off-state as shown in Fig. 5.1. The time durations that the source stays in the On- and Off-states are geometrically distributed with means $1/p_{on\_off}$ and $1/p_{off\_on}$, respectively. The On-state corresponds to the talkspurt and the Off-state corresponds to the silence period in human speech. In the On-state, a voice source generates voice packets at some constant rate which is equal to the transmission rate of a single CDMA code., whereas a voice source will not generate packets when it is in the Off-state. The parameters p_{on_off} and p_{off_on} are determined by the reciprocals of the expected durations of a talksput and a silence period, respectively. The probability that a voice will stay in the On- and Off-state are

$$\pi_{on} = \frac{p_{off\_on}}{p_{on\_off} + p_{off\_on}}$$

and

$$\pi_{off} = \frac{p_{on\_off}}{p_{on\_off} + p_{off\_on}},$$

respectively.

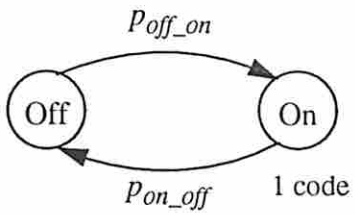


Fig. 5.1. The on-off model for a voice source.

5.2.2. Traffic Model for a Video User

With compression, the output rate of a video codec is variable. Most of the time, the codec generates packets with low rates that corresponds to no scene change in the video being compressed. Occasionally, the output rate will increase abruptly due to scene change in the video.

Various models have been proposed for variable bit rate video sources in ATM networks, e.g., [47], [56], and [67]. Most of them model the output rate of a video codec as an autoregressive process [56], [67]. To simplify the channel access control of a video

source, we assume that the output of a video codec is further buffered such that the buffered video source is either generating packets at a high rate or at a low rate. So the model for a video source is a two-state high-low model shown in Fig. 5.2. A source will generate video packets at the rates that can be accommodated by C_{hi} and C_{lo} codes, when it is in the High- and Low-state, respectively. Similar to the voice model, the time durations that a video source stays in the High- and Low-states are geometrically distributed with means equal to $1/p_{hi\_low}$ and $1/p_{low\_hi}$, respectively. Thus, the probabilities that a video user is in the High- and Low-state are

$$\pi_{hi} = \frac{p_{low\_hi}}{p_{hi\_low} + p_{low\_hi}}$$

and

$$\pi_{lo} = \frac{p_{hi\_low}}{p_{hi\_low} + p_{low\_hi}},$$

respectively. The parameters, p_{hi_low}, p_{low_hi}, C_{hi}, and C_{lo}, are functions of the compression scheme, the buffering delay, and the characteristics of the video itself.

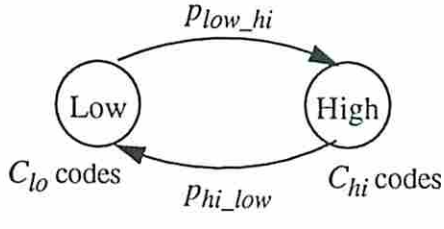


Fig. 5.2. The high-low model for a video source.

5.2.3. Traffic Model for a Data User

The underlying admission control policy is assumed to admit only M_d data users (with outstanding messages to be transmitted) into the system. Under the admission control, the M_d data users are either transmitting or have data messages waiting to be transmitted. As soon as a data user finishes transmitting its message, the base station will admit another data user and therefore keep the number of data user in the system constant.

The admitted data users are further subject to the congestion control imposed by the base station, which broadcasts feedback[1] regarding the status of the system on the downlink (from the base station to mobile users.) A data user under congestion control is modeled by a 2-state Markov chain illustrated in Fig. 5.3. When the user is in the Idle-state, it will transmit with probability p_{try} and enter the Active-state. When in the Active-state, the user will transmit C_d packets in parallel using C_d spreading codes. The data message is geometrically distributed and has a mean of L packets. The transition probability from the Active-state to the Idle-state is not exactly $1/L$ because the data user may be forced to stop transmission when the system gets congested. The selection of p_{try} and p_{quit} based on the feedback is the subject of congestion control which is elaborated in Section 5.3.

1. The feedback information about the system state can sometimes be obtained by data users themselves, e.g. [19].

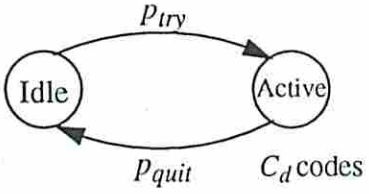


Fig. 5.3. The state transition diagram of a data source.

5.2.4. Model of CDMA Channel Quality

In a CDMA channel, the quality of the reception of a particular packet is usually defined in terms of bit error rate (BER) or packet error rate (PER), which depends on many factors. First, the fading characteristics of a wireless link will impact the bit error rate (BER) and thus the packet error rate (PER). However there exist many countermeasures for the impairment of fading channels, including forward error correction (FEC) coding, coded modulation, diversity combining, and power control. Second, the multiuser interference due to the simultaneous transmission of other users sharing the same frequency band will also cause the degradation of BER and PER. There are ongoing research on CDMA multiuser detection that tries to detect multiuser signals by advanced signal processing techniques.

With the complexity of the aforementioned physical layer parameters and the diverse transmission rates of different services, the conditional link quality in a slot from the perspective of a particular user can be defined as a function of the number of active users in each of the service types[1]. In other words, the conditional link quality of a voice user

that is in the On-state in a particular slot can be expressed as $Q_{vo}(n_1, n_2, n_3)$, where n_1, n_2, and n_3 are the number of voice users in the On-state, the number of video users in the High-state, and the number of data users in transmission, respectively. The number of video users in the Low-state is simply $(M_{vi} - n_2)$. Similarly the conditional link quality of a video user in the High- and Low-state can be expressed as $Q_{hi}(n_1, n_2, n_3)$ and $Q_{lo}(n_1, n_2, n_3)$; the conditional link quality of a data user is given by $Q_d(n_1, n_2, n_3)$.

The exact computation of these link quality functions is not trivial, especially when multipath fading is serious and the dependency between simultaneous transmissions are taken into account, e.g., [15], [25]. These functions that describe the conditional link quality will be used in the next section to compute the long-term average link quality.

5.2.5. Markovian Model for the Integrated Traffic

With the Markovian traffic models defined in Sections 5.2.1-5.2.3, the CDMA system can be modeled by a 3-dimensional Markov chain with the state variables being the number of voice users in the On-state (N_t^{vo}), the number of video users in the Hi-state (N_t^{vi}), and the number of data users in the Active-state (N_t^d), where the subscript t denotes time slot t. In time slot t+1, the state variables are given by

$$N_{t+1}^{vo} = N_t^{vo} + B(M_{vo} - N_t^{vo}, p_{off\_on}) - B(N_t^{vo}, p_{on\_off}),$$

1. This relies on the symmetry assumption that users with the same traffic model are indistinguishable from each other as far as the base station is concerned.

$$N_{t+1}^{vi} = N_t^{vi} + B\left(M_{vi} - N_t^{vi}, p_{low\_hi}\right) - B\left(N_t^{vi}, p_{hi\_low}\right),$$

and

$$N_{t+1}^{d} = N_t^{d} + B\left(M_d - N_t^{d}, p_{try}\right) - B\left(N_t^{d}, p_{quit}\right),$$

where $B(N, p)$ is a binomial random variable with the parameters N and p, the probability of success in an independent Bernoulli trial. Note that the state evolution regarding voice and video users is independent of other types of traffic sources, whereas the state evolution of data users depends on all of the 3 types of traffic through p_{try} and p_{quit}. This means that real-time services (voice and video) are subject to admission control only. Once they are admitted to the system, their behavior depends completely on their traffic models, not on anything else. The 1-step transition probability of the system Markov chain then can be simplified as

$$Pr\left[N_{t+1}^{vo}, N_{t+1}^{vi}, N_{t+1}^{d} \middle| N_t^{vo}, N_t^{vi}, N_t^{d}\right]$$

$$= Pr\left[N_{t+1}^{vo} \middle| N_t^{vo}\right] Pr\left[N_{t+1}^{vi} \middle| N_t^{vi}\right] Pr\left[N_{t+1}^{d} \middle| N_t^{vo}, N_t^{vi}, N_t^{d}\right]$$

Depending on the congestion control policy used, the transition probabilities of this 3-dimensional Markov chain can be computed and the equilibrium state probabilities, $Pr\left[\left(N^{vo}, N^{vi}, N^{d}\right) = (n_1, n_2, n_3)\right]$, $0 \le n_1 \le M_{vo}$, $0 \le n_2 \le M_{vi}$, $0 \le n_3 \le M_d$, can be obtained by solving a system of linear equations [27].

5.2.6. Performance Measures

Two performance measures are of interest for the CDMA system with integrated services, namely, the long-term link quality experienced by a user, which may be service-dependent, and the data throughput.

With the equilibrium state probabilities and the conditional link quality functions given in Section 5.2.4, the long-term link quality seen by a particular voice user in the On-state is given by

$$\overline{Q_{vo}} = \sum_{n_1=1}^{M_{vo}} \sum_{n_2=0}^{M_{vi}} \sum_{n_3=0}^{M_d} Q_{vo}(n_1, n_2, n_3) \frac{\frac{n_1}{M_{vo}} Pr[(N^{vo}, N^{vi}, N^d) = (n_1, n_2, n_3)]}{\pi_{on}},$$

where $\frac{\frac{n_1}{M_{vo}} Pr[(N^{vo}, N^{vi}, N^d) = (n_1, n_2, n_3)]}{\pi_{on}}$ is the conditional probability that the system is in state (n_1, n_2, n_3) given that a particular voice user is in the On-state.

Similarly, the long-term link quality of a data user in the Active-state is

$$\overline{Q_d} = \sum_{n_1=0}^{M_{vo}} \sum_{n_2=0}^{M_{vi}} \sum_{n_3=1}^{M_d} Q_d(n_1, n_2, n_3) \frac{\frac{n_3}{M_d} Pr[(N^{vo}, N^{vi}, N^d) = (n_1, n_2, n_3)]}{\sum_{n_1}\sum_{n_2}\sum_{n_3} \frac{n_3}{M_d} Pr[(N^{vo}, N^{vi}, N^d) = (n_1, n_2, n_3)]},$$

where the triple summation in the denominator of the above equation is the probability that a particular data user is in transmission mode.

The long-term link quality of a video user can be computed by properly averaging the long-term link quality of a video user given that it is in the Hi-state ($\overline{Q_{hi}}$) and the long-term link quality of a video user given that it is in the Low-state ($\overline{Q_{lo}}$). That is,

$$\overline{Q_{vi}} = \overline{Q_{hi}}\pi_{hi} + \overline{Q_{lo}}\pi_{lo},$$

where $$\overline{Q_{hi}} = \sum_{n_1=0}^{M_{vo}} \sum_{n_2=1}^{M_{vi}} \sum_{n_3=0}^{M_d} Q_{hi}(n_1, n_2, n_3) \frac{\frac{n_2}{M_{vi}} Pr\left[(N^{vo}, N^{vi}, N^d) = (n_1, n_2, n_3) \right]}{\pi_{hi}}$$

and

$$\overline{Q_{lo}} = \sum_{n_1=0}^{M_{vo}} \sum_{n_2=0}^{M_{vi}-1} \sum_{n_3=0}^{M_d} Q_{lo}(n_1, n_2, n_3) \frac{\frac{(M_{vi}-n_2)}{M_{vi}} Pr\left[(N^{vo}, N^{vi}, N^d) = (n_1, n_2, n_3) \right]}{\pi_{lo}}.$$

The data throughput (packets/slot) of the system is given by

$$S = \sum_{n_1} \sum_{n_2} \sum_{n_3} n_3 C_d \cdot \left(1 - PER_d(n_1, n_2, n_3)\right) \cdot Pr\left[(N^{vo}, N^{vi}, N^d) = (n_1, n_2, n_3) \right],$$

where $PER_d(n_1, n_2, n_3)$ is the conditional probability that a particular data packet will be in error, which could be negligibly small if the required link quality is high and FEC coding is applied on data packets.

5.3 Congestion Control Policies

All of the congestion control policies considered in this section are Markovian in the sense that all of the data users will receive some form of feedback information about the system state at the end of a slot, which will be used to determine the transmission probabilities of data users in the next slot.

The feedback could be either complete (i.e., the number of voice users in the On-state, the number of video users in the Hi-state, and the number of data users in the Active-state are all known,) or limited (e.g., only some function of the system state is known.) In this section, two policies are considered for both types of feedback and optimization of the control parameters is described.

5.3.1. Complete Feedback

In the case of complete feedback, it is assumed that in the beginning of time slot t+1, the system state in time slot t, $(N_t^{vo}, N_t^{vi}, N_t^d)$, is known. Two policies are considered, namely the threshold policy and the direct policy.

A. The Threshold Policy

Define $f(N_t^{vo}, N_t^{vi}, N_t^d) = N_t^{vo} + N_t^{vi} \times C_{hi} + (M_{vi} - N_t^{vi}) \times C_{lo} + N_t^d \times C_d$. The function $f(.)$ gives the number of active codes in the CDMA channel at time slot t, which can be used as a congestion index. The basic idea of the threshold policy is to shut down

data transmission when the system is congested, i.e., when the number of active CDMA codes exceeds the threshold T_{max}. Whereas if the system is not congested, transmitting data users continue their transmission while waiting data users transmit with some appropriate probability to avoid congestion. Given the complete feedback in time slot t, the transmission probabilities of waiting data users (p_{try}) and active data users (p_{quit}) in time slot $t+1$ are given by

$$p_{try} = \begin{cases} 0, & \text{if } f(N_t^{vo}, N_t^{vi}, N_t^d) > T_{max} \\ min\left(1, \dfrac{T_{max} - T_s - f(N_t^{vo}, N_t^{vi}, N_t^d)}{C_d (M_d - N_t^d)}\right), & \text{otherwise,} \end{cases}$$

and $$p_{quit} = \begin{cases} 1, & \text{if } f(N_t^{vo}, N_t^{vi}, N_t^d) > T_{max} \\ \dfrac{C_d}{L}, & \text{otherwise,} \end{cases}$$

where T_{max} is the threshold for congestion control and T_s is the safe margin for congestion avoidance. The rationale that p_{try} is chosen in the particular form is that [T_{max} - $f(N_t^{vo}, N_t^{vi}, N_t^d)$]/C_d can be thought of as the residual channel capacity, and (M_d - N_t^d) is the number of data users awaiting for transmission. By choosing p_{try} equal to the ratio of the above two factors, the expected number of data users that will start to transmit will be equal to the leftover capacity. The inclusion of T_s in p_{try} gives us the degree of freedom to find the best trade-off between congestion control (T_{max}) and congestion avoidance (T_s).

This policy can be optimized by finding the T_{max} and T_s that maximize the data throughput while guaranteeing some link quality measure, e.g., the long-term BER of

voice users.

B. The Direct Policy

Instead of choosing the optimal threshold parameters that guarantee the long-term link quality like the threshold policy, the direct policy tries to find the maximum p_{try} that satisfies the average link quality in the next slot. Given the complete feedback in time slot t and a particular value of p_{try}, the average link quality of voice users in time slot (t+1) is given by

$$E\left[Q_{t+1}^{vo} \middle| N_t^{vo}, N_t^{vi}, N_t^{d}\right]$$

$$= \sum_{n_1}\sum_{n_2}\sum_{n_3} Q_{vo}(n_1, n_2, n_3)\, Pr\left[(N_{t+1}^{vo}, N_{t+1}^{vi}, N_{t+1}^{d}) = (n_1, n_2, n_3) \middle| N_t^{vo}, N_t^{vi}, N_t^{d}\right]$$

$$= \sum_{n_1}\sum_{n_2}\sum_{n_3} Q_{vo}(n_1, n_2, n_3)\, Pr\left[N_{t+1}^{vo} = n_1 \middle| N_t^{vo}\right] Pr\left[N_{t+1}^{vi} = n_2 \middle| N_t^{vi}\right]$$

$$\cdot\, Pr\left[N_{t+1}^{d} = n_3 \middle| N_t^{vo}, N_t^{vi}, N_t^{d}\right].$$

It can be seen from the above expression that the average short-term link quality increases as p_{try} increases monotonically. If p_{try} is chosen such that the average short-term link quality is equal to the required link quality for QoS guarantee, the data throughput in slot t+1 can be maximized.

Based on this observation, the direct policy sets the p_{try} in each slot that will result in

the average short-term link quality no worse than[1] the required link quality. This policy therefore maximizes the data throughput in each slot, while maintaining the long-term link quality by guaranteeing the short-term link quality.

The direct policy requires the computation of the optimal p_{try} for every system state, which is not required for the threshold policy. Although the computation is intensive, the set of optimal p_{try} can be stored in a look-up table in the base station.

5.3.2. Limited Feedback

In the case of limited feedback, it is assumed that in the beginning of time slot t+1, the number of active codes in the CDMA channel at time slot t, $f(N_t^{vo}, N_t^{vi}, N_t^d)$, is known. Both the threshold policy and the direct policy are considered as follows.

A. The Threshold Policy

Given the limited feedback, a conservative choice of p_{try} and p_{quit} at time slot t+1 is

$$p_{try} = \begin{cases} 0, & \text{if } f(N_t^{vo}, N_t^{vi}, N_t^d) > T_{max} \\ min\left(1, \dfrac{T_{max} - T_s - f(N_t^{vo}, N_t^{vi}, N_t^d)}{M_d C_d}\right), & \text{otherwise,} \end{cases}$$

1. Sometimes the real-time traffic may be so light that even p_{try} = 1 cannot make BER in the next slot equal to the required BER. So the long-term BER may be lower than the required BER, and thus leaving some room for improving data throughput.

and $$p_{quit} = \begin{cases} 1, & \text{if } f(N_t^{vo}, N_t^{vi}, N_t^d) > T_{max} \\ \dfrac{C_d}{L}, & \text{otherwise.} \end{cases}$$

In this case, p_{quit} is the same and p_{try} is almost the same as in the case of complete feedback except for the denominator when the system is not congested. This policy also can be optimized by the proper selection of T_{max} and T_s.

B. The Direct Policy

The average link quality of voice users in time slot (t+1) given the limited feedback in time slot t is given by

$$E\left[Q_{t+1}^{vo} \middle| f(N_t^{vo}, N_t^{vi}, N_t^d) = f_0\right]$$

$$= \sum_{n_1}\sum_{n_2}\sum_{n_3} Q_{vo}(n_1, n_2, n_3)\, Pr\left[(N_{t+1}^{vo}, N_{t+1}^{vi}, N_{t+1}^d) = (n_1, n_2, n_3) \middle| f(N_t^{vo}, N_t^{vi}, N_t^d) = f_0\right]$$

The conditional probability can be written as

$$Pr\left[(N_{t+1}^{vo}, N_{t+1}^{vi}, N_{t+1}^d) = (n_1, n_2, n_3) \middle| f(N_t^{vo}, N_t^{vi}, N_t^d) = f_0\right]$$

$$= \sum_{m_1}\sum_{m_2}\sum_{m_3} Pr\left[(N_{t+1}^{vo}, N_{t+1}^{vi}, N_{t+1}^d) = (n_1, n_2, n_3) \middle| (N_t^{vo}, N_t^{vi}, N_t^d) = (m_1, m_2, m_3)\right]$$

$$\cdot Pr\left[(N_t^{vo}, N_t^{vi}, N_t^d) = (m_1, m_2, m_3) \middle| f(N_t^{vo}, N_t^{vi}, N_t^d) = f_0\right].$$

It can be seen that the optimal p_{try} for a given limited feedback, $f(N_t^{vo}, N_t^{vi}, N_t^d)$, cannot be solved in a similar way as in the case of complete feedback since the equation involves some 1-step transition probabilities which depend on the optimal p_{try} of other

system states. In fact, the optimal p_{try} for every system state is the solution of a system of nonlinear equations with size $(M_{vo}+1)\times(M_{vi}+1)\times(M_d+1)$.

5.4 Numerical Results

The numerical results of the performance of the congestion control policies are presented in this section. Both complete and limited feedback are considered for the threshold policy, whereas only complete feedback is investigated for the direct policy.

To focus on the issue of congestion control, we consider a direct sequence CDMA system employing perfect power control and DPSK demodulation with hard decision. The BER is used as the measure of link quality. With the standard Gaussian assumption for other-user interference, the BER given k active interfering[1] CDMA codes is

$$BER(k) = \frac{1}{2}\exp\left(-\frac{E_b}{N_{eff}}\right), \text{ where } \frac{E_b}{N_{eff}} = \left[\left(\frac{E_b}{N_0}\right)^{-1} + \left(\frac{G}{k}\right)^{-1}\right]^{-1},$$

where E_b/N_0 is the signal to noise ratio due to the thermal noise and other-cell interference, and G is the processing gain. Therefore the conditional link quality (BER) of the traffic sources given that the system is in state (n, n_2, n_3) is given by

$$BER_{vo}(n_1, n_2, n_3) = BER\left(f(n_1, n_2, n_3) - 1\right), \; n_1 \geq 1.$$

1. The multiple codes used by a traffic source are assumed to have zero cross-correlation and hence do not interfere with each other.

$$BER_{hi}(n_1, n_2, n_3) = BER\left(f(n_1, n_2, n_3) - C_{hi}\right),\ n_2 \geq 1.$$

$$BER_{lo}(n_1, n_2, n_3) = BER\left(f(n_1, n_2, n_3) - C_{lo}\right),\ n_2 \leq M_{vi} - 1.$$

$$BER_{d}(n_1, n_2, n_3) = BER\left(f(n_1, n_2, n_3) - C_{d}\right),\ n_3 \geq 1.$$

Given the number of voice, video and data users that are transmitting in a slot, voice users usually have the worst link quality (or the highest BER), since there are more spreading codes that will interfere with a voice user. Therefore in the following numerical examples, the BER performance of the system will be based on the BER experienced by a voice user if not stated otherwise. With the BER requirement of real-time services, the packet error probability of data users is assumed to be negligible throughout this section.

The parameters for the CDMA system are given in Table 5.1. Note that the average bit rate of a video user is 10 times as high as that of a voice user. From the numerical values of the traffic source parameters given in Table 5.1, π_{on}, π_{off}, π_{hi} and π_{lo} are equal to 0.4359, 0.5641, 0.0897 and 0.9103, respectively.

In Fig. 5.4, the data throughput of a system with (M_{vo}, M_{vi}, M_d) = (10, 2, 10) and (C_d, L) = (1, 10) under the threshold policy with complete feedback is plotted as a function of the 2 threshold parameters, i.e., T_{max} and T_s. It can be seen that data throughput first increases as T_{max} or T_s increases, and then levels off. The corresponding BER is shown in Fig. 5.5. Both performance measures have similar behavior as T_{max} and T_s change. To fully utilize the system capacity, it is desirable to have data throughput (S) as high as

Table 5.1: System parameters of the CDMA network

Item	Symbol	Value
Transition probability of a voice source from On- to Off-state	p_{on_off}	1/17
Transition probability of a voice source from Off- to On-state	p_{off_on}	1/22
Transition probability of a video source from High- to Low-state	p_{hi_low}	1/3
Transition probability of a video source from Low- to High-state	p_{low_hi}	1/30.43
Number of codes active while a video source is in High-state	C_{hi}	8
Number of codes active while a video source is in Low-state	C_{lo}	4
Number of codes active while a data source is in transmission	C_d	1, 2
Average length of data message (packets)	L	3, 10
Bit energy to (1-sided) noise spectral density (dB)	E_b/N_0	11
Processing gain	G	256
Slot duration (ms)	*	20
Number of traffic sources admitted to the system	(M_{vo}, M_{vi}, M_d)	(10, 2, 10), (10, 2, 15), (30, 0, 10), (30, 0, 15)

possible, while the BER does not exceed some specified value, say 10^{-3}, for the QoS guarantee of real-time services. To show the trade-off between data throughput and BER, we define the cost function (the objective function for system optimization) to be

$$\text{Cost}(S, BER) = w \cdot g\left(BER - 10^{-3} \right) - S,$$

where w is a weighting factor and $g(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$, and S is the data throughput. The larger the weighting factor is, the more penalty is imposed when the system BER

exceeds 10^{-3}. For a given number of voice and video users admitted to the system, it is preferable to have the cost function as small as possible. The cost function with various T_{max} and T_s for the same system are plotted in Fig. 5.6. It can be seen that there are a set of (T_{max}, T_s) which give low cost, e.g., (27, 4), (28, 5), (29, 6), (30, 7), etc.

The data throughput and BER of the system with (M_{vo}, M_{vi}, M_d) = (10, 2, 10) under various values of C_d and L are compared in Fig. 5.7 and 5.8. From the cost function shown in Fig. 5.9, we see that there exist a best T_{max} which results in the lowest cost for each pair of C_d and L.

The effect of different mixes of traffic sources is explored in Fig. 5.10-5.12. It can be seen that the system performance is not sensitive to the exact numbers of voice and video users admitted, whereas the system performance is quite different as the number of data users varies. However, if the threshold parameters are chosen to minimize the cost function, the cases with M_d = 10 give lower cost than the cases with M_d = 15. This shows that the admission control policy plays a role in optimizing system performance.

The BER performance of different traffic sources is compared in Fig. 5.13. It can be seen that the BER performance of voice and data users (C_d = 2) is very close, whereas the video users have much lower BER due to the fact that more spreading codes are used by video users.

The performance of the threshold policy with limited feedback is shown in Fig. 5.14-5.16. Although these curves are very similar to those with complete feedback, the set of

(T_{max}, T_s) that give low cost has smaller T_s compared with the case of complete feedback., e.g., (27, 0), (28, 1), (29, 2), (30, 3), etc.

The performance of the direct policy with complete feedback under different numbers of admitted traffic sources are compared in Fig. 5.17-5.19. As expected, the data throughput decreases as more voice/video users are admitted to the system, as shown in Fig. 5.17. From Fig. 5.18, it is observed that $BER \leq 10^{-3}$ cannot always be guaranteed if too many voice/video users are accepted. To be more specific, T_{max} needs to be chosen below 24 to maintain $BER \leq 10^{-3}$ when (M_{vo}, M_{vi}, M_d) = (10, 3, 10), whereas the BER will always exceed 10^{-3} when (M_{vo}, M_{vi}, M_d) = (9, 4, 10). The best values for T_{max} that result in the lowest cost can be found in Fig. 5.19.

The performance of the threshold policy with complete and limited feedback and the direct policy with complete feedback is compared in Fig. 5.20-5.21. For the threshold policy, the performance curves are obtained by choosing the "best"[1] T_s for each value of T_{max}. It can be seen that the threshold policy under complete and limited feedback works almost equally well if the threshold parameters are carefully chosen. The direct policy has slightly lower throughput since its BER is slightly lower than 10^{-3}. A fair comparison between these policies requires the optimal[2] values of the threshold parameters to be found first.

---

1. The best in the finite set of parameter space that we have evaluated.
2. The optimal threshold parameters should result in BER exactly equal to the BER requirement.

5.5 Conclusion

An analytical model is developed to evaluate the performance of 2 congestion control policies for an integrated voice/video/data CDMA packet radio network. The model can be used for any signalling method at the physical layer since the effect of the physical layer parameters are summarized in a set of conditional link quality functions.

For each of the congestion control policies, we have evaluated its performance under 2 types of feedback information. It is found that the performance of the threshold policy is not sensitive to the type of feedback if the threshold parameters are chosen optimally.

The numerical results show that the system performance is not sensitive to the exact number of voice and video users as long as their aggregate bit rate is constant. On the other hand, the traffic parameters of data users, such as average length of data messages and the number of spreading codes allocated for each data user, play a more important role.

Based on the performance prediction of the proposed model, a base station can perform admission control and reject a call when it finds that the required QoS cannot be guaranteed. The model can be incorporated into admission control to trade-off between the call blocking probability of voice/video calls and the delay of data messages.

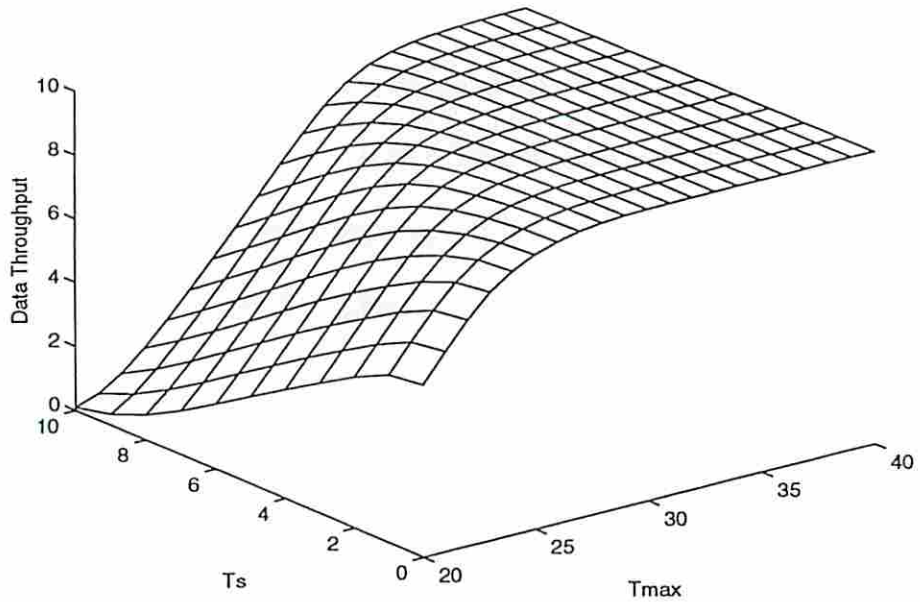


Fig. 5.4. Data throughput vs. the threshold parameters under the threshold policy with complete feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$ and $(C_d, L) = (1, 10)$.

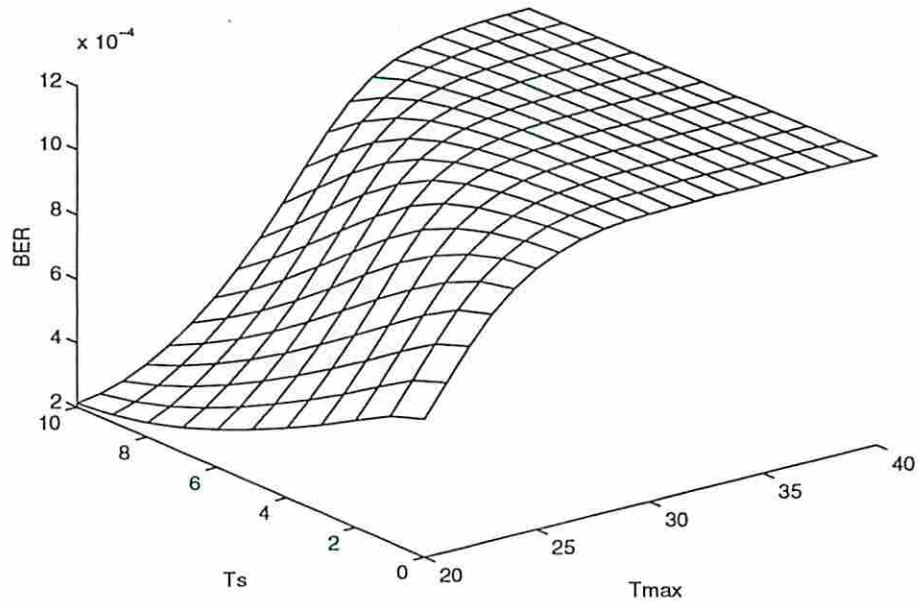


Fig. 5.5. Bit error rate vs. the threshold parameters under the threshold policy with complete feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$ and $(C_d, L) = (1, 10)$.

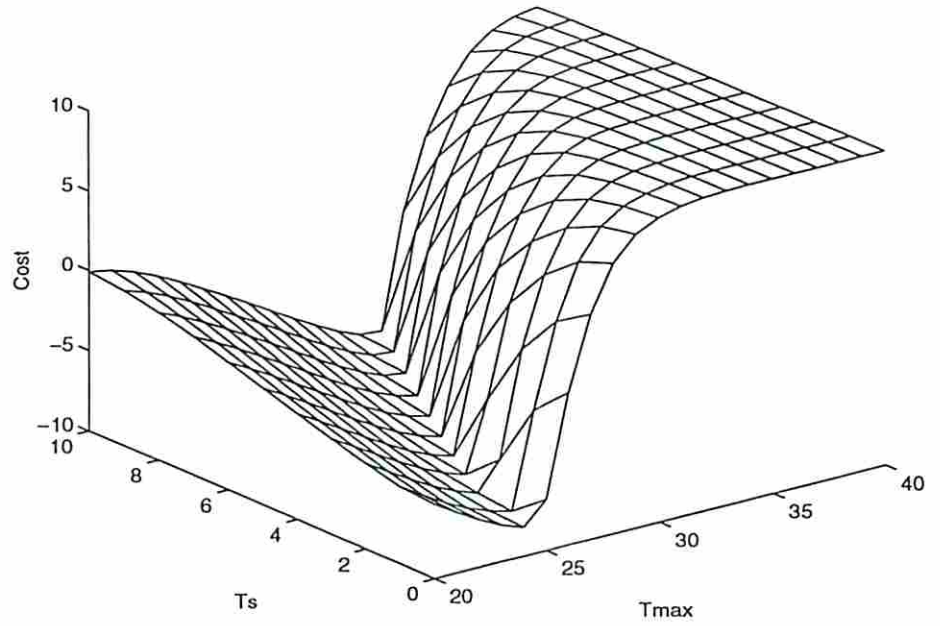


Fig. 5.6. Cost function vs. the threshold parameters under the threshold policy with complete feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$, $(C_d, L) = (1, 10)$ and $w = 2*10^5$.

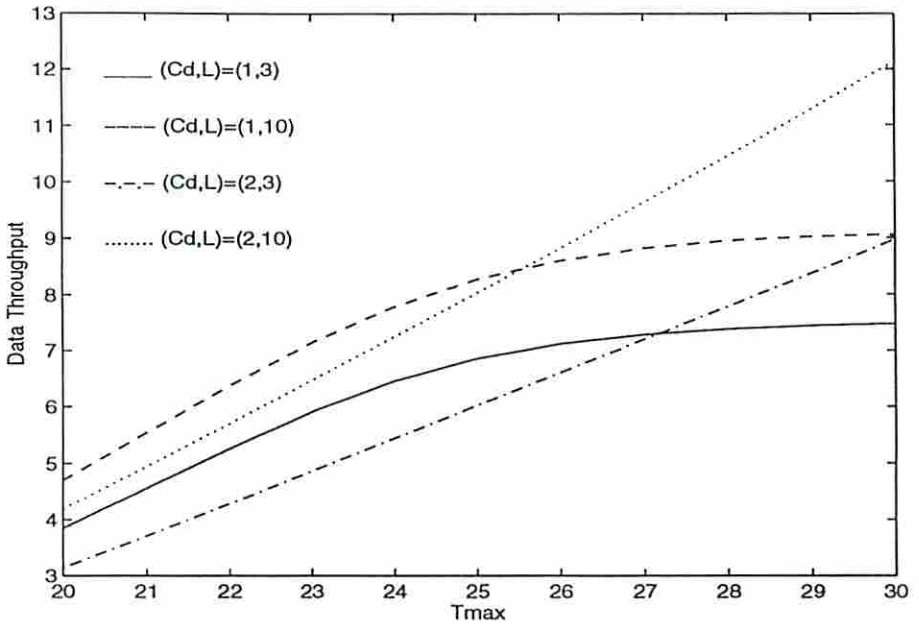


Fig. 5.7. Data throughput vs. T_{max} under the threshold policy with complete feedback for various (C_d, L). $T_s = 2$ and $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$.

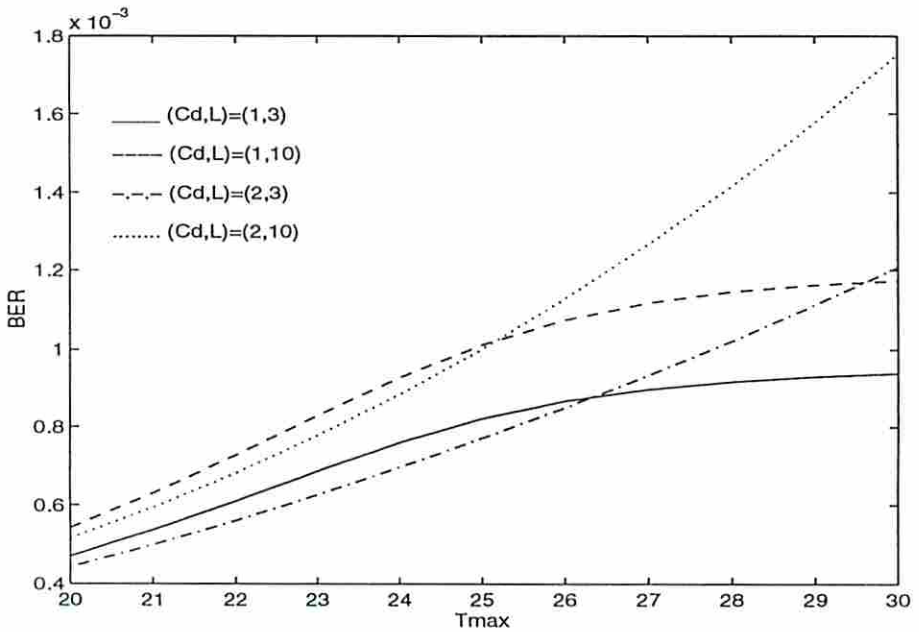


Fig. 5.8. Bit error rate vs. T_{max} under the threshold policy with complete feedback for various (C_d, L). $T_s = 2$ and $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$.

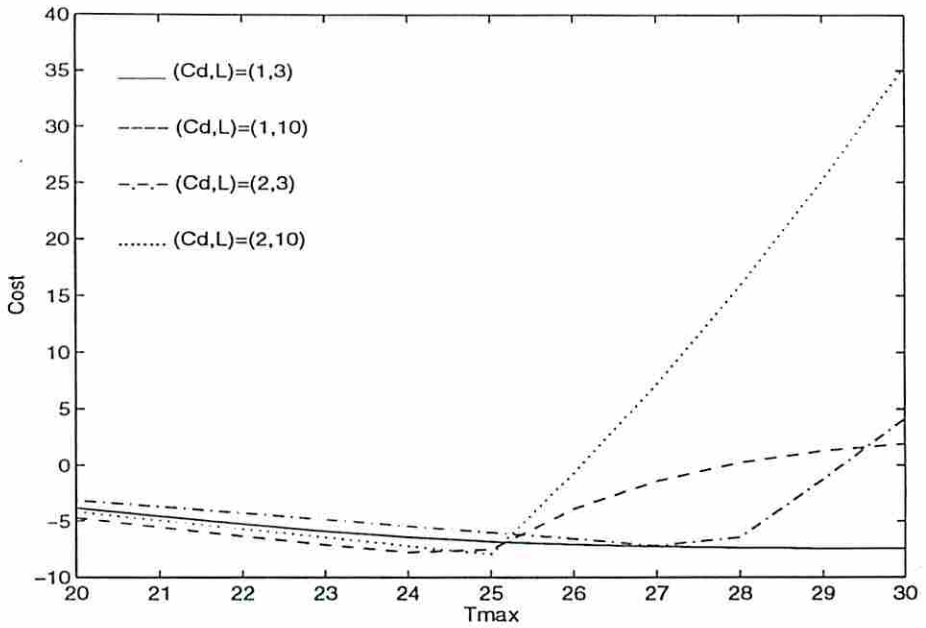


Fig. 5.9. Cost function vs. T_{max} under the threshold policy with complete feedback for various (C_d, L). $T_s = 2$, $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$ and $w = 6.3*10^4$.

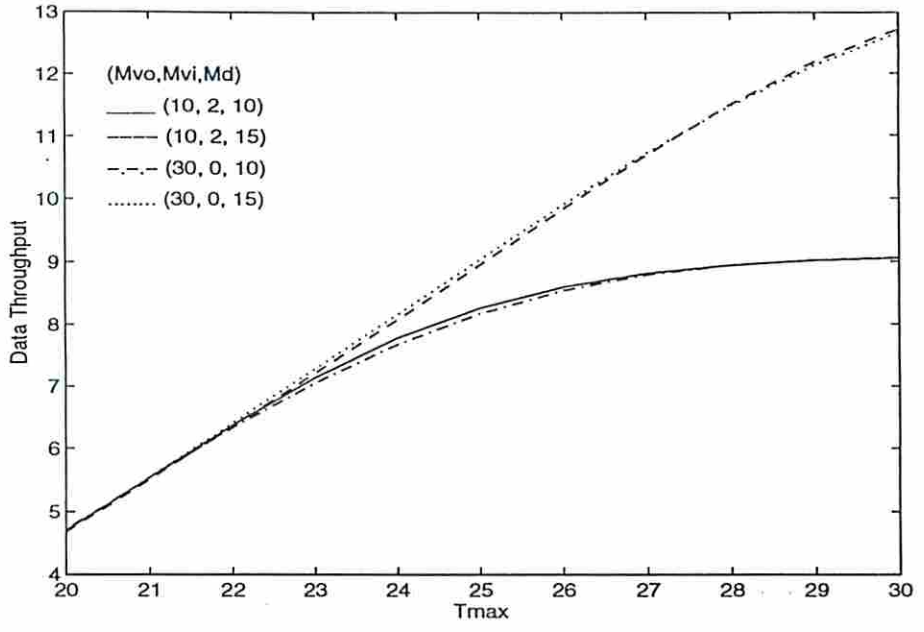


Fig. 5.10. Data throughput vs. T_{max} under the threshold policy with complete feedback for various (M_{vo}, M_{vi}, M_d). $T_s = 2$ and $(C_d, L) = (1, 10)$.

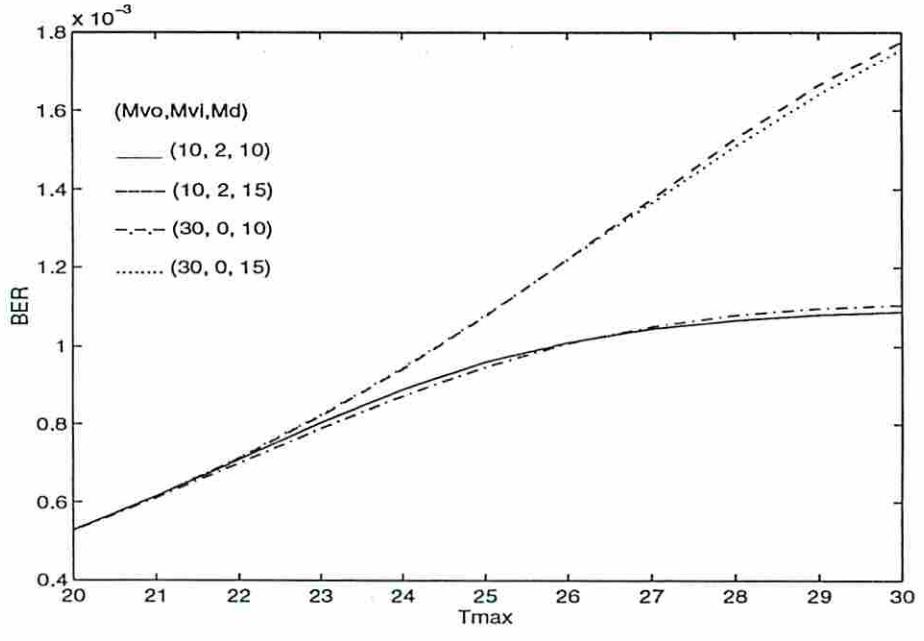


Fig. 5.11. Bit error rate vs. T_{max} under the threshold policy with complete feedback for various (M_{vo}, M_{vi}, M_d). $T_s = 2$ and $(C_d, L) = (1, 10)$.

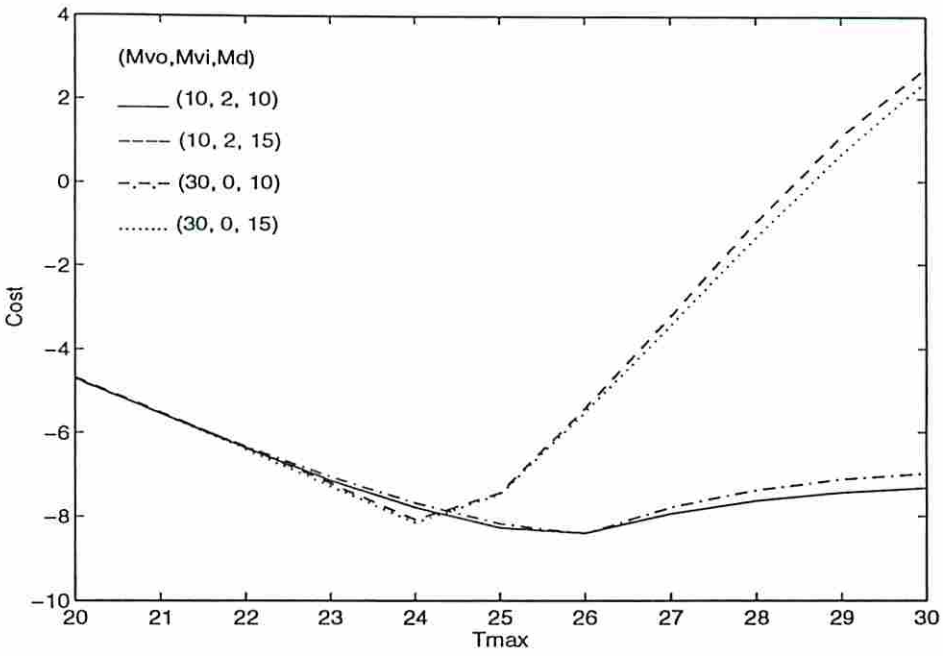


Fig. 5.12. Cost function vs. T_{max} under the threshold policy with complete feedback for various (M_{vo}, M_{vi}, M_d). $T_s = 2$, (C_d, L) = (1, 10) and $w = 2*10^4$.

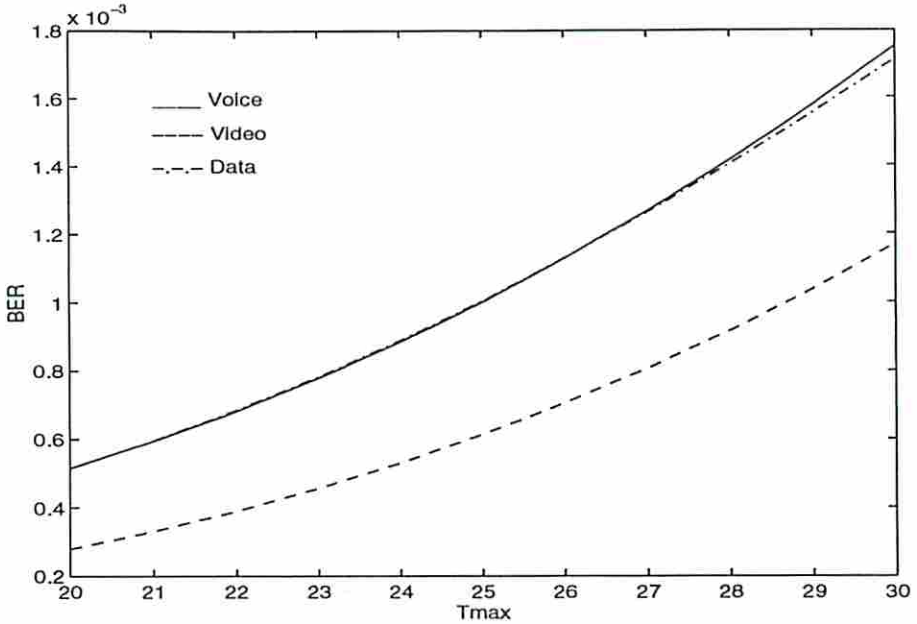


Fig. 5.13. The bit error rate performance vs. T_{max} for different types of traffic sources under the threshold policy with complete feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$, $(C_d, L) = (2, 10)$ and $T_s = 2$.

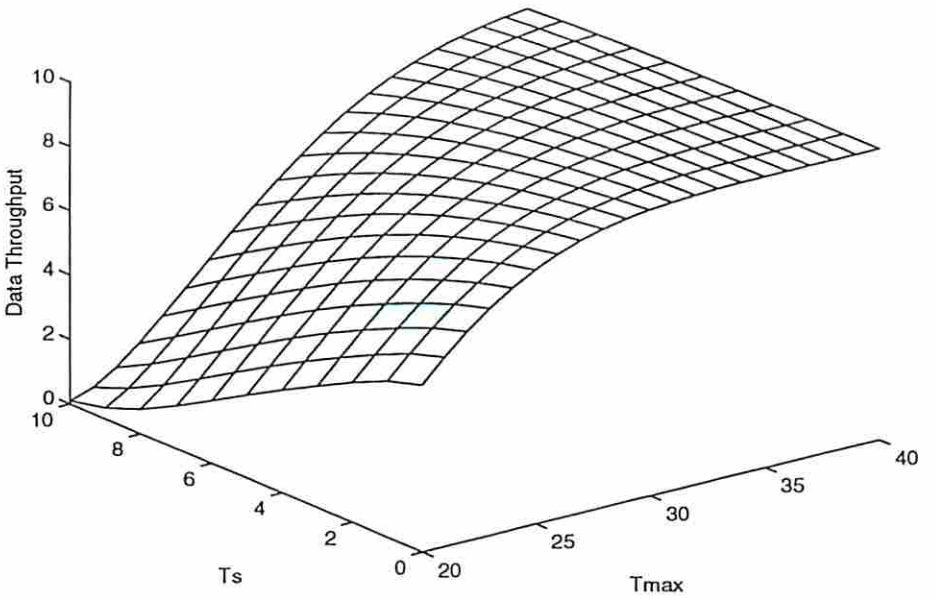


Fig. 5.14. Data throughput vs. the threshold parameters under the threshold policy with *limited* feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$ and $(C_d, L) = (1, 10)$.

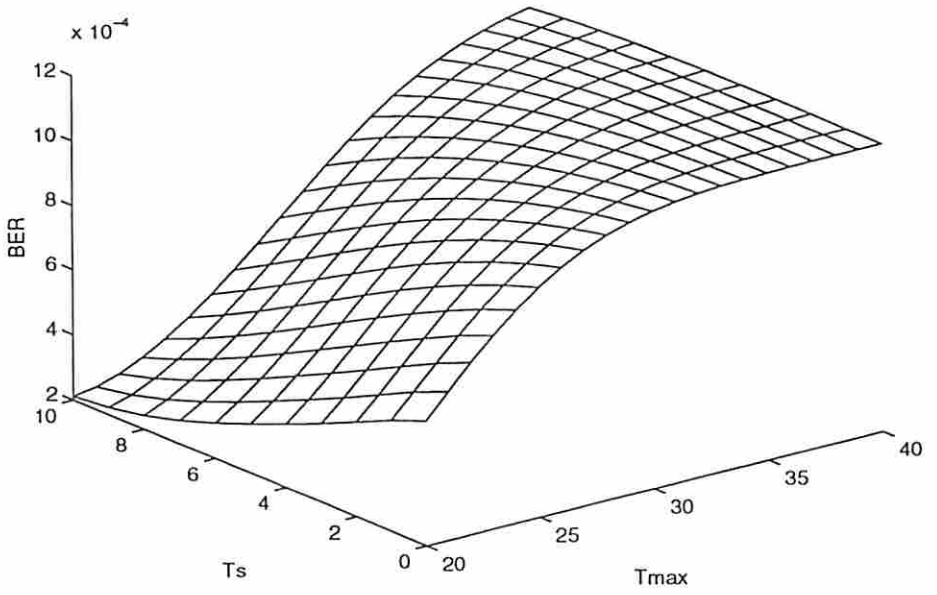


Fig. 5.15. Bit error rate vs. the threshold parameters under the threshold policy with *limited* feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$ and $(C_d, L) = (1, 10)$.

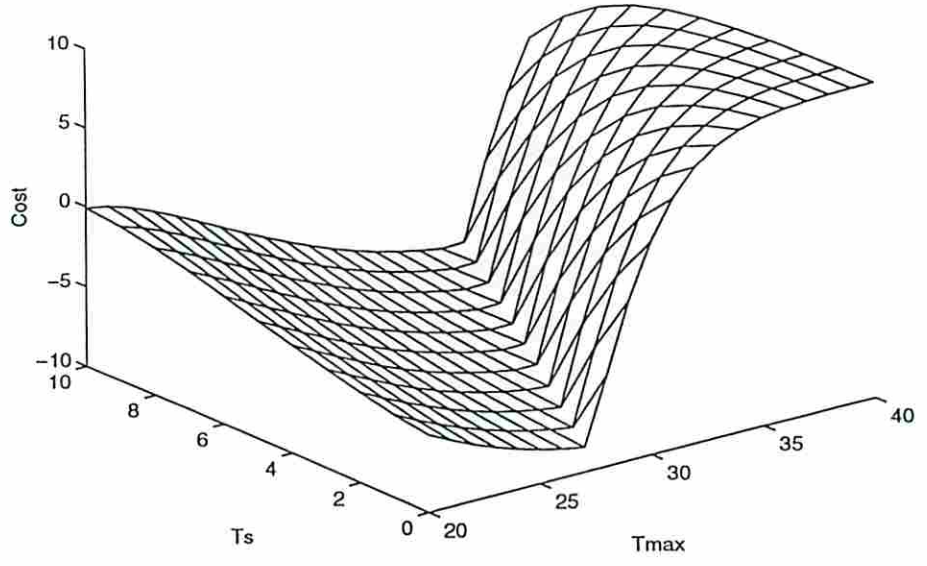


Fig. 5.16. Cost function vs. the threshold parameters under the threshold policy with *limited* feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$, $(C_d, L) = (1, 10)$ and $w = 2*10^5$.

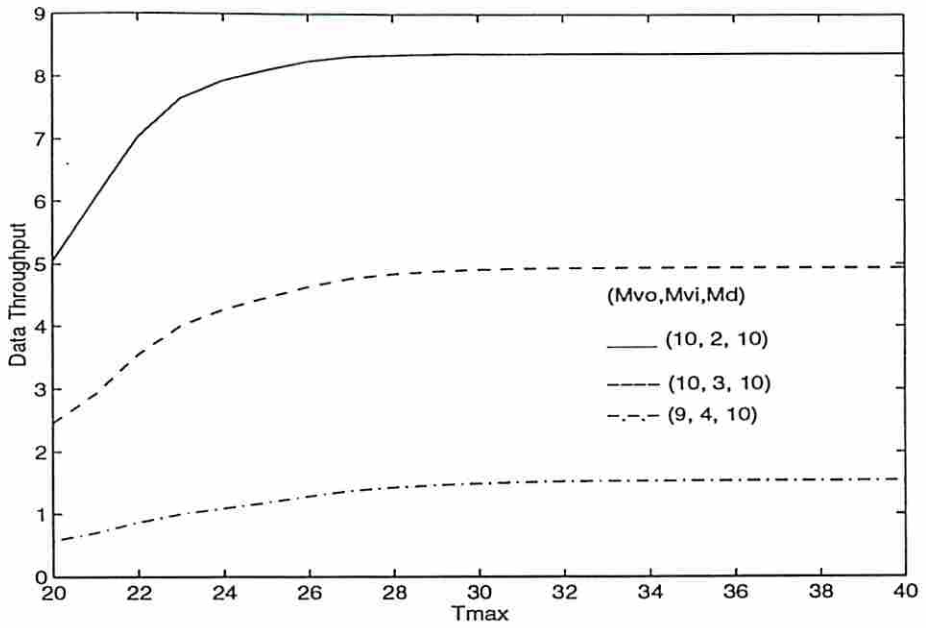


Fig. 5.17. Throughput vs. T_{max} under the *direct* policy with complete feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$ and $(C_d, L) = (1, 10)$.

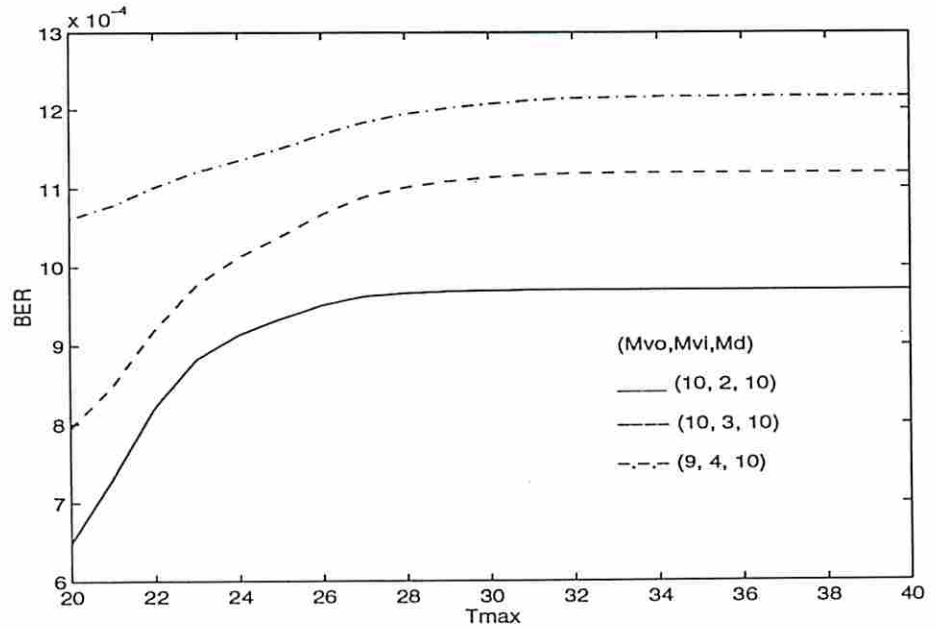


Fig. 5.18. Bit error rate vs. T_{max} under the *direct* policy with complete feedback. $(M_{vo}, M_{vi}, M_d) = (10, 2, 10)$ and $(C_d, L) = (1, 10)$.

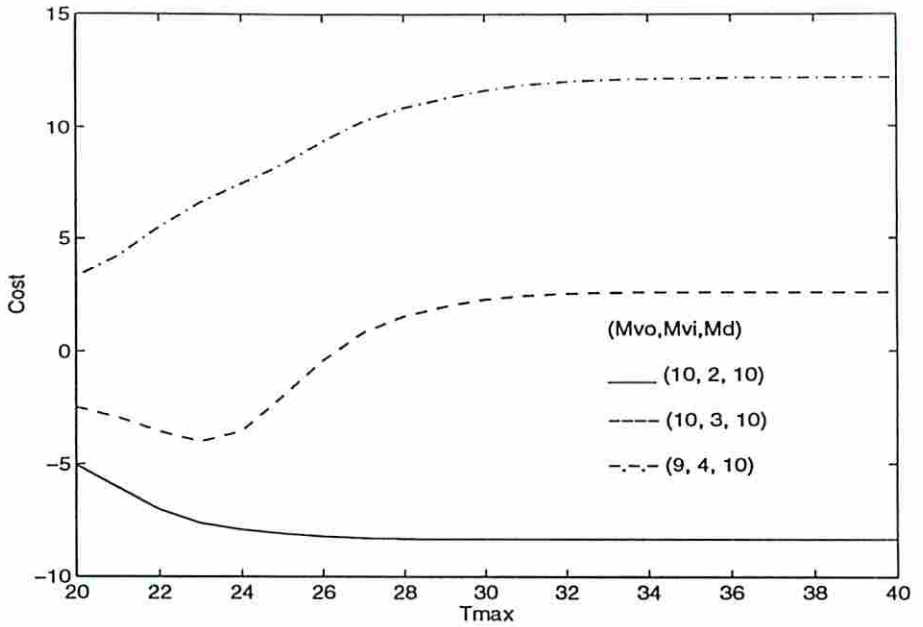


Fig. 5.19. Cost function vs. T_{max} under the *direct* policy with complete feedback. (M_{vo}, M_{vi}, M_d) = (10, 2, 10), (C_d, L) = (1, 10) and $w = 6.3*10^4$.

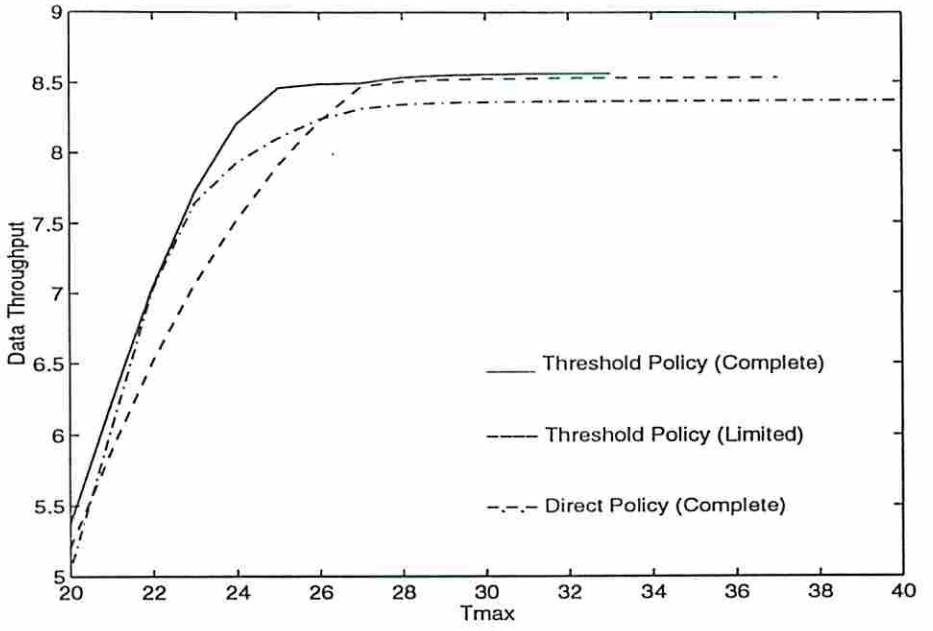


Fig. 5.20. Comparison of data throughput between the threshold policy (with complete and limited feedback) and the direct policy (with complete feedback). (M_{vo}, M_{vi}, M_d) = (10, 2, 10) and (C_d, L) = (1, 10).

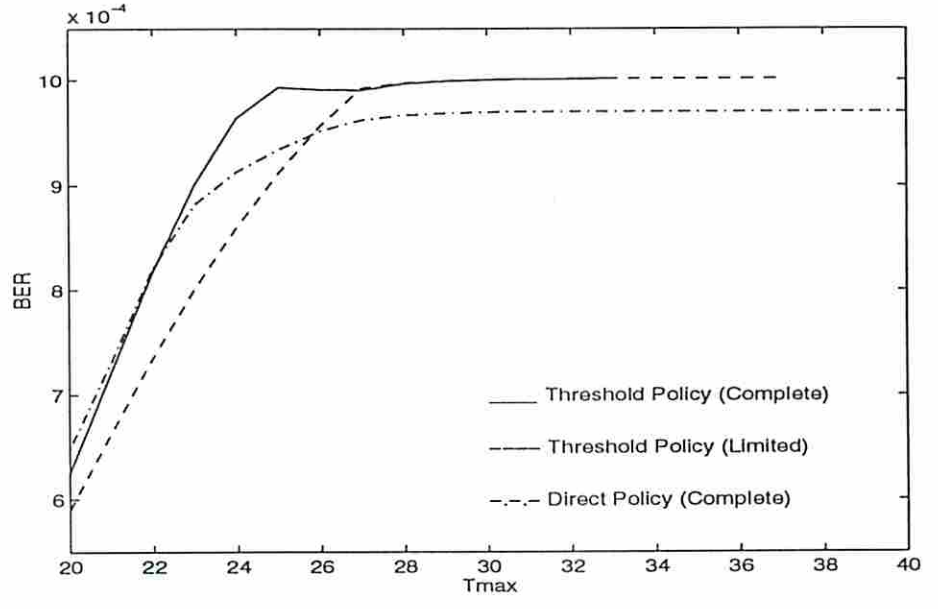


Fig. 5.21. BER comparison between the 2 threshold policies and the direct policy. (M_{vo}, M_{vi}, M_d) = (10, 2, 10) and (C_d, L) = (1, 10).

Chapter 6

Conclusions and Future Research

6.1 Conclusions

Wireless networks can be modeled at various levels of detail. The appropriate modeling detail depends on the aspect of interest. In this dissertation, the focus is on the access layer, which interacts with the physical layer (i.e., channels, modulation, coding, and spreading) and the link layer (i.e., traffic sources) of the network. Through a set of interface parameters which summarize the effect of the physical layer parameters, the network can be analyzed by Markovian models which take into account the dynamic behavior of different types of traffic sources with possibly different service requirements. If the resulting Markovian model has a state space exceeding the capability of computers, the system performance evaluation has to resort to simulation. This modeling approach has been used to study several wireless communication networks with different service requirements.

For mobile slotted ALOHA data networks (Chapter 2 and 3), we found that the near-far effect has a strong impact on the fairness of the users in the network. The unfairness of

user performance is solved by introducing a Non-uniform retransmission control policy, which allows distant users to retransmit with higher probability. The improvement on the fairness of the network allows the wireless data network to operate at a higher traffic load, while still guaranteeing the QoS of data users at the edge of service area. In addition to the fairness issue, we have studied the impact of multipath fading and shadowing effects on the system performance. Fading and shadowing are found to slightly degrade system performance (for both single-user and the whole network) when the channel load is light. However, they will improve the throughput performance of the whole network at the cost of unfair user performance when the channel load is heavy.

For the distributed mobile data network with hard real-time communication requirement (Chapter 4), the reliability of the network (the deadline failure probability) is a trade-off between various system parameters such as channel data rate, packet error probability, and the duration of the deadline period. Markovian models are used to analyze the network behavior in both the worst-case and the steady-state scenarios. The worst-case analysis tells us how fast the system is able to recover from a short period of catastrophic failure, whereas the analysis of the steady-state shows the impact of the packet error due to fading and collision on the system reliability.

For wireless multimedia (voice/video/data) CDMA networks (Chapter 6), the trade-off between the data throughput and the link quality (e.g., bit error rate) is analyzed by a Markovian model, where the transition probability is driven by the underlying congestion control policy. Two congestion control policies are developed and their performance under

both the complete and limited feedback are evaluated. The link quality of different traffic sources could be different because of their difference in the number of CDMA codes used for the transmission of different bit rates, e.g., voice and video. In the numerical results, it is observed that both policies have comparable performance as long as their control parameters are chosen optimally.

6.2 Future Research

The grouping strategy used for the multi-group model in Chapter 2 is an ad-hoc one. More study on other strategies that can improve the accuracy of the performance prediction is desirable. Besides, the transmission power of the mobile users are assumed to be constant (i.e., with no power control.) Study on the joint effect of transmission power and retransmission probability is a possible extension.

In the study of the hard real-time mobile data network, capture effect which could improve the system performance (reliability) is not taken into account. Moreover, the system performance is expected to improve drastically if some intelligent policy is applied which tries to distinguish packet errors due to fading from due to collision. In other words, if packet errors are due to fading, there is no reason to switch to other slots in the frame. Also the number of slots that can be used to transmit whenever the reserved slot is lost is an interesting trade-off factor.

For the wireless CDMA networks studied in Chapter 6, the optimization of the

threshold parameters for both the threshold policy and the direct policy need to be done before a fair comparison between the 2 policies can be performed. Also it is interesting to see the joint effect of admission control and congestion control and study the trade-off between call blocking probability of voice/video calls and data message delay. In addition, the system performance can be improved by power control of high data-rate video sources which have experienced better link quality due to the smaller number of interfering codes in the CDMA channel.

Appendix A

Decoupled Solution for the Multi-group Model

This appendix show how a K-dimensional Markov chain is approximated by K one-dimensional chains.

A.1 The Markovian Model for a Decoupled Group

For a particular group i, the state variable of the corresponding Markov chain is the number of backlogged terminals in the group. The state transition probabilities are governed by M_i, σ_i, q_i, and $\phi_i(k)$, where $\phi_i(k)$ is the conditional probability that one of terminals in group i succeeds given that k terminals in group i transmit. Although $\phi_i(k)$ depends on the activities of all other groups, we assume the interference on group i due to other groups is statistically the same and independent in every slot. We will show how to compute $\phi_i(k)$ after the following notations are defined. Let

n_i = the state variable of group i, $i = 1, 2, \ldots, K$.

$\pi_i(k)$ = the stationary probability that group i is in state k.

$\underline{\pi}_i = (\pi_i(0), \pi_i(1), \ldots, \pi_i(M_i))$, i.e., the stationary state vector of group i.

a_i = the number of transmissions in group i, where $0 \le a_i \le M_i$.

$\min(x, y)$ = the smaller of x and y.

$A_i(j, k)$ = Pr[j unbacklogged terminals in group i transmit $\mid n_i = k$]

$$= \binom{M_i - k}{j} \sigma_i^j (1 - \sigma_i)^{M_i - k - j}, \text{ where } 0 \le j \le M_i - k.$$

$B_i(j, k)$ = Pr[j backlogged terminals in group i transmit $\mid n_i = k$]

$$= \binom{k}{j} q_i^j (1 - q_i)^{k - j}, \text{ where } 0 \le j \le k.$$

We have

$$\phi_i(k) = \sum_{j_1}^{M_1} \cdots \sum_{j_{i-1}}^{M_{i-1}} \sum_{j_{i+1}}^{M_{i+1}} \cdots \sum_{j_K}^{M_K} p_i(j_1, \ldots, j_{i-1}, k, j_{i+1}, \ldots, j_K)$$

$$\cdot \Pr[a_1 = j_1, \ldots, a_{i-1} = j_{i-1}, a_{i+1} = j_{i+1}, \ldots, a_K = j_K \mid a_i = k],$$

where $p_i(j_1, \ldots, j_{i-1}, k, j_{i+1}, \ldots, j_K)$ is the probability that one terminal in group i succeeds given the activity vector $(j_1, \ldots, j_{i-1}, k, j_{i+1}, \ldots, j_K)$. Further assuming that the activities of different groups are independent, we have

$$\phi_i(k) = \sum_{j_1}^{M_1} \cdots \sum_{j_{i-1}}^{M_{i-1}} \sum_{j_{i+1}}^{M_{i+1}} \cdots \sum_{j_K}^{M_K} p_i(j_1, \ldots, j_{i-1}, k, j_{i+1}, \ldots, j_K) \cdot \prod_{l=1, l \ne i}^{K} \Pr[a_l = j_l].$$

By the law of total probability,

$$\Pr[a_l = j_l] = \sum_{k=0}^{M_l} \pi_l^k \Pr[a_l = j_l \mid n_l = k]$$

$$= \sum_{k=0}^{M_l} \pi_l^k \sum_{b=0}^{min(k, j_l)} \Pr[\ b \text{ backlogged terminals}$$

and $(j_l - b)$ unbacklogged terminals transmit $|\ n_l = k\ ]$

$$= \sum_{k=0}^{M_l} \pi_l^k \sum_{b=0}^{min(k, j_l)} B_l(b, k) A_l(j_l - b, k) \ .$$

We have shown in the above how to compute $\phi_i(k)$ given the stationary state occupancy probabilities of all interfering groups. Namislo [43] expressed the state transition probabilities for a slotted ALOHA network with capture in terms of the number of terminals, the probability of packet generation, the retransmission probabilities, and the conditional success probability given the number of active transmissions in a slot. The state occupancy probabilities can then be obtained recursively [27]. Now we can compute $\underline{\pi}_i$ in terms of M_i, σ_i, q_i, and $\phi_i(k)$, $k = 1,2,..., M_i$. From $\underline{\pi}_i$, the average throughput and delay can be obtained. The drift analysis can be used to determine whether group i is stable or bistable.

A.2 The Interaction of the Decoupled Markov Chains

We have showed in the above how to compute the state probabilities $\underline{\pi}_i$ given the state probabilities of the other groups $\{\underline{\pi}_j\}_{j=1, j\neq i}^{j=K}$. We now introduce the iterative procedure that can be used to obtained the state probabilities of all groups. The procedure consists of

three steps.

Step 1: Set $\underline{\pi}_i = (1, 0,..., 0)$, $i = 1, 2,..., K$.

Step 2: For $i = 1, 2,..., K$, do Step 2.1-2.

Step 2.1: Compute $\phi_i(k)$, $k = 1, 2,..., M_i$.

Step 2.2: Solve for $\underline{\pi}_i$ in terms of M_i, σ_i, q_i, $\phi_i(k)$, $k = 1,2,..., M_i$, and $\{\underline{\pi}_j\}_{j=1, j\neq i}^{j=K}$.

Step 3: Stop if no significant change in $\{\underline{\pi}_j\}_{j=1}^{j=K}$ is observed. Otherwise, go to Step 2.

After the iterative procedure stops, the state occupancy probabilities can be used to determined the throughput, delay, and stability of each group.

Appendix B

Maximum and Maximum Balanced Throughput of Slotted ALOHA with Multi-level Dominating Power

In this appendix we will derive the maximum throughput and the maximum balanced throughput of a slotted ALOHA network with multi-level dominating power.

B.1 Multi-group S-ALOHA with Multi-level Dominating Power in Heavy Traffic

Consider a S-ALOHA system with K groups of terminals (or users) with group i consisting of M_i terminals that are identical and independent. The received power at the central station is assumed to be such that if there is no other users from the same or higher-power group transmit, the packet is successfully received with probability one, otherwise the packet is lost. Groups with smaller index are assumed to have stronger power at the central station. In the heavy traffic scenario considered, a terminal in group i will transmit

with probability q_i in a slot. The probability that a_i users in group i transmit in a slot is given by a binomial probability $\binom{M_i}{a_i} q_i^{a_i} (1-q_i)^{M_i - a_i}$, denoted by $B(M_i, a_i, q_i)$. The throughput of group i, S_i, is therefore

$$S_i = \sum_{a_1=0}^{M_1} \cdots \sum_{a_K=0}^{M_K} P_i(a_1, \ldots, a_K) \prod_{j=1}^{K} P_r[a_j \text{ users in group } j \text{ transmit}]$$

$$= \sum_{a_1=0}^{M_1} \cdots \sum_{a_K=0}^{M_K} P_i(a_1, \ldots, a_K) \prod_{j=1}^{K} B(M_j, a_j, q_j) .$$

The throughput of a user in group i is thus S_i / M_i due to symmetry. The sum of the throughput of all groups is the network throughput S, which is given by

$$S = \sum_{i=1}^{K} S_i$$

$$= \sum_{a_1=0}^{M_1} \cdots \sum_{a_K=0}^{M_K} \left[\sum_{i=1}^{K} P_i(a_1, \ldots, a_K) \right] \prod_{j=1}^{K} B(M_j, a_j, q_j) .$$

Let $P_s(a_1, \ldots, a_K) = \sum_{i=1}^{K} P_i(a_1, \ldots, a_K)$. S can be further simplified as

$$S = \sum_{a_1=0}^{M_1} \cdots \sum_{a_K=0}^{M_K} P_s(a_1, \ldots, a_K) \prod_{j=1}^{K} B(M_j, a_j, q_j) .$$

B.2 The Maximum Throughput Problem

For a K-group S-ALOHA network with given $\vec{M} = (M_1, M_2, ..., M_K)$ and the transmission probabilities $\vec{q} = (q_1, q_2, ..., q_K)$, the network throughput S can be maximized by an optimal choice of $\vec{q}$. This problem can be written as a multivariate nonlinear constrained optimization problem given by

$$\max_{q_1, ..., q_K} S$$

subject to $0 \le q_i \le 1$, $i = 1, ..., K$.

Since the central station can capture the transmission from a user in group i if there is no other transmission from groups 1, 2, ..., i. Therefore we have the throughput of each group and the overall network throughput given by

$$S_1 = B(M_1, 1, q_1),$$

$$S_2 = B(M_2, 1, q_2) B(M_1, 0, q_1),$$

$$\vdots$$

$$S_K = B(M_K, 1, q_K) \prod_{j=1}^{K-1} B(M_j, 0, q_j),$$

$$S = \sum_{i=1}^{K} S_i.$$

To find the (local) optimal $\vec{q}$, we need to solve the following system of K

simultaneous equations $\frac{\partial S}{\partial q_i} = 0$, $i = 1, \ldots, K$. It turns out that the K equations can be solved recursively starting from $\frac{\partial S}{\partial q_K} = 0$, which gives $q_K = 1/M_K$. Then, $\frac{\partial S}{\partial q_{K-1}} = 0$ can be used to find the optimal q_{K-1}. By repeating this process, the optimal $\vec{q}$ can be obtained as follows.

$$q_i = \begin{cases} \dfrac{1}{M_K} & \text{if } i = K \\ \dfrac{1 - B(M_K, 1, q_K)}{M_{K-1} - B(M_K, 1, q_K)} & \text{if } i = K-1 \\ \dfrac{1 - B(M_{i+1}, 1, q_{i+1}) - \sum_{j=i+2}^{K} B(M_j, 1, q_j) \prod_{l=i+1}^{j-1} B(M_l, 0, q_l)}{M_i - B(M_{i+1}, 1, q_{i+1}) - \sum_{j=i+2}^{K} B(M_j, 1, q_j) \prod_{l=i+1}^{j-1} B(M_l, 0, q_l)} & \text{if } 1 \le i \le K-2 \end{cases}$$

It is not difficult to see that as M_i approaches infinity ($i = 1, \ldots, K$), the above recursion will match the one given in [41], where the maximum throughput of a multi-group S-ALOHA networks with infinite population and multi-level dominating power was considered.

B.3 The Maximum Balanced Throughput Problem

The optimal transmission probabilities $\vec{q} = (q_1, q_2, \ldots, q_K)$ that maximize the

network throughput do not guarantee that the throughput per user is the same for all users, which is a desirable feature for applications that equally concern the throughput of individual user. To ensure that the user throughput is the same throughout the system while the network throughput is maximized, we need to add (K - 1) more constraints to the maximum throughput problem. The problem now becomes

$$\max_{q_1, \ldots, q_K} \quad S$$

subject to $0 \le q_i \le 1$, $i = 1, \ldots, K$, and

$$\frac{S_1}{M_1} = \frac{S_2}{M_2} = \ldots = \frac{S_K}{M_K},$$

where $S_1 = B(M_1, 1, q_1)$,

$$S_i = B(M_i, 1, q_i) \prod_{j=1}^{i-1} B(M_j, 0, q_j), \quad 2 \le i \le K, \text{ and}$$

$$S = \sum_{i=1}^{K} S_i.$$

By the method of Lagrange multipliers, we define

$$T = \sum_{i=1}^{K} S_i + \lambda_1 \left(\frac{S_1}{M_1} - \frac{S_2}{M_2} \right) + \lambda_2 \left(\frac{S_2}{M_2} - \frac{S_3}{M_3} \right) + \ldots + \lambda_{K-1} \left(\frac{S_{K-1}}{M_{K-1}} - \frac{S_K}{M_K} \right),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_{K-1}$ are Lagrange multipliers. The (local) optimal $\vec{q} = (q_1, q_2, \ldots, q_K)$

can be obtained by solving the following system of equations

$$\begin{cases} \dfrac{\partial T}{\partial q_i} = 0 & i = 1, ..., K \\ \dfrac{\partial T}{\partial \lambda_j} = 0 & j = 1, ..., K-1. \end{cases}$$

From $\dfrac{\partial T}{\partial \lambda_j} = 0$, we have $\dfrac{S_j}{M_j} = \dfrac{S_{j+1}}{M_{j+1}}$. After some simplification, we get

$$q_j = \frac{q_{j+1}(1-q_{j+1})^{M_{j+1}-1}}{1+q_{j+1}(1-q_{j+1})^{M_{j+1}-1}} \qquad 1 \le j \le K-1.$$

From $\dfrac{\partial T}{\partial q_K} = 0$, we have $q_K = \dfrac{1}{M_K}$. Therefore the optimal $\vec{q}$ can be obtained by the recursive formula

$$q_i = \begin{cases} \dfrac{1}{M_K} & \text{if } i = K \\ \dfrac{q_{i+1}(1-q_{i+1})^{M_{i+1}-1}}{1+q_{i+1}(1-q_{i+1})^{M_{i+1}-1}} & \text{if } 1 \le i \le K-1. \end{cases}$$

The numerical results for the maximum throughput and the maximum balanced throughput of several K-group systems with multi-level dominating power are summarized in Tables B.1. It is observed that the maximum (balanced) throughput increases as K increases with the help of the near-far effect. Nevertheless, as the total network throughput is maximized, the user throughput is extremely unbalanced. The maximum throughput in

is smaller than the maximum balanced throughput as expected.

Table B.1: The maximum throughput and the maximum balanced throughput of K-group S-ALOHA networks with multi-level dominating power

	Maximum Throughput			Maximum Balanced Throughput		
	Transmission Probability $\vec{q}$	User Throughput	Network Throughput	Transmission Probability $\vec{q}$	User Throughput	Network Throughput
$K=1$ $\vec{M}=(50)$	.0200	.0074	.3679	.0200	.0074	.3679
$K=2$ $\vec{M}=(8, 42)$	.0823 .0238	.0451 .0045	.5482	.0088 .0238	.0083	.4130
$K=3$ $\vec{M}=(4, 11, 35)$	.1318 .0590 .0286	.0863 .0182 .0031	.6544	.0094 .0106 .0286	.0091	.4569
$K=4$ $\vec{M}=(3, 5, 12, 30)$	.1492 .1023 .0538 .0333	.1080 .0409 .0105 .0023	.7239	.0101 .0106 .0123 .0333	.0099	.4942
$K=5$ $\vec{M}=(2,3,7,12,26)$	.2181 .1509 .0706 .0538 .0385	.1705 .0665 .0170 .0066 .0017	.7819	.0107 .0110 .0120 .0142 .0385	.0106	.5285

Bibliography

[1] N. Abramson, "The ALOHA system -- Another alternative for computer communications," in *AFIPS Conf. Proc. 1970 Fall Joint Comput. Conf.*, vol. 37, pp. 281-285, 1970.

[2] A. A. Abu-Dayya, and N. C. Beaulieu, "Comparison of diversity with simple block coding on correlated frequency-selective fading channels", *IEEE Trans. Commun.*, vol. 43, no. 11, pp. 2704-2713, Nov. 1995.

[3] K. Araki, K. Ohata, K. Okada, H. Mizuno *et al.*, "Satellite communication systems for multimedia services", *IEE International Conference on Digital Satellite Communications*, pp. 690-5, 1994.

[4] J. C. Arnbak and W. V. Blitterswijk, "Capacity of slotted ALOHA in Rayleigh-fading channels," *IEEE Journal on selected areas in communications*, Vol. 5, No. 2, pp. 261-9, Feb. 1987.

[5] P. T. Brady, "A model for generating On-Off speech patterns in two-way conversations," *Bell Systems Technical Journal*, pp. 2445-2472, Sep. 1969.

[6] R. J. C. Bultitude and G. K. Bedal, "Propagation characteristics on microcellular urban mobile radio channels at 910 MHz," *IEEE Journal on selected areas in communications*, Vol. 7, No. 1, pp. 31-39, Jan. 1989.

[7] A. B. Carleial and M. E. Hellman, "Bistable behavior of ALOHA-type systems," *IEEE Trans. Commun.*, Vol. 23, pp. 401-410, 1975.

[8] P. Castoldi, G. Immovilli and M. L. Merani, "Network and single user performance evaluation of a mobile data system over flat fading transmission channels," *Wireless Networks*, Vol. 1, No. 1, pp. 95-106, 1995.

[9] W. W. Chu *et al.*, "Measurement and evaluation studies of packet radio communications systems", *Final Technical Report, UCLA-ENG 8003*, March 19, 1980.

[10] D. J. Cichon, T. Zwick and J. Lahteenmaki, "Ray optical indoor modeling in multi-floored buildings: simulations and measurements", *IEEE Antennas and Propagation Society International Symposium*, 1995, vol. 1, pp. 522-5255.

[11] W. Crowther *et al.*, "A system for broadcast communication: Reservation-ALOHA," *Proc. sixth Hawaii Int. Conf. Sys. Sci.*, pp. 371-374, Jan. 1973.

[12] D. C. Cox, "Wireless network access for personal communications", *IEEE Commun. Mag.*, Dec. 1992, pp. 96-115.

[13] D. C. Cox, "910 MHz urban mobile radio propagation: multipath characteristics in New York city", *IEEE Trans. Commun.*, vol. 21, no. 11, pp. 1188-94, Nov. 1973.

[14] S. Dastangoo, "A multimedia medium access control protocol for ATM based mobile networks", *IEEE PIMRC'95*, pp. 794-798.

[15] E. Geraniotis and J. Wu, "The probability of multiple correct packet receptions in direct-sequence spread-spectrum networks," *IEEE Journal on selected areas in communications*, vol. 12, no. 5, pp. 871-884, Jun. 1994.

[16] C.-L. I, R. D. Gitlin, "Multi-code CDMA wireless personal communications networks", *IEEE ICC'95*, pp. 1060-1064.

[17] D. J. Goodman and A. A. M. Saleh, "The near-far effect in local ALOHA radio communications," *IEEE Trans. Veh. Techn.*, Vol. 36, No. 1, pp. 19-27, Feb. 1987.

[18] I. Habbab, M. Kavehrad and C.-E. Sundberg, "ALOHA with capture over slow and fast fading radio channels with coding and diversity," *IEEE Journal on selected areas in communications*, Vol. 7, No. 1, pp. 79-88, Jan. 1989.

[19] B. Hajek, "Recursive retransmission control–Application to a frequency-hopped spread-spectrum system," *Proceedings of 1982 Conference on Inform. and Systems Sciences*, 1982.

[20] P. Harley, "Short distance attenuation measurements at 900 MHz and 1.8 GHz using low antenna heights for microcells", *IEEE Journal on selected areas in communications,* vol. 7, no. 1, pp. 5-11, Jan. 1989.

[21] H. Hashemi, "Simulation of the urban radio propagation channel", *IEEE Trans. Veh. Techn.*, vol. 28, no. 3, pp. 213-25, Aug. 1979.

[22] *IMSL MATH/Library User's Manual*, edition 2.0, IMSL, Houston, 1991.

[23] D. G. Jeong and W. S. Jeon, "Performance of an exponential backoff scheme for slotted-ALOHA protocol in local wireless environment," *IEEE Trans. Veh. Techn.*, Vol. 44, No. 3, pp. 470-479, Aug. 1995.

[24] U. Karaasian, P. Varaiya, and J. Walrand, "Freeway traffic flow, platooning, and control", *research report, Institute of Transportation Studies*, University of California, Berkeley, Aug. 1990.

[25] T. J. Ketseoglou and E. Geranoitis, "Multireception probabilities for FH/SSMA communications," *IEEE Journal on selected areas in communications*, vol. 40, no. 1, Jan. 1992.

[26] R. E. Khan *et al.*, "Advances in packet radio technology," *Proc. IEEE*, vol 66, pp. 1468-1496, Nov. 1978.

[27] L. Kleinrock, *Queueing Systems*, Volume 1: Theory, John Wiley & Sons, 1975.

[28] L. Kleinrock and S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. Commun.*, Vol. 23, pp. 410-423, 1975.

[29] L. Kleinrock and F. A. Tobagi, "Packet switching in radio channels: Part I -- Carrier sense multiple access modes and their throughput-delay characteristics," *IEEE Trans. Commun.*, vol. 23, no. 12, pp. 1400-1416, Dec. 1975.

[30] Kurose, J. F. and Mouftah, H. T., "Computer-aided modeling, analysis, and design of communication networks", *IEEE Journal on selected areas in communications*, vol. 6, no. 1, pp. 130-145, Jan. 1988.

[31] C. Lau and C. Leung, "Capture models for mobile packet radio networks," *Proc. Int. Conf. Commun.*, ICC, Atlanta, Apr. 1990, pp. 1226-1230.

[32] C. L. Law and T. H. Teo "Error performance of narrowband and spread-spectrum modulation with diversity in frequency-selective channel", *International Journal of Electronics*, Oct. 1995, vol.79, no.4, pp. 421-38.

[33] V. O. K. Li, and X. Qiu "Personal communication systems (PCS)", *Proc. of the IEEE*, Sept., 1995, 83, no. 9, pp. 1210-1243.

[34] J.-P. Linnartz, *Narrowband Land-Mobile Radio Networks*, Artech House, Boston, 1993.

[35] T.-K. Liu, J. A. Silvester and A. Polydoros, "A analytical model for multi-group slotted ALOHA with capture," *IEEE PacRim Conf.*, pp. 534-537, 1995.

[36] T.-K. Liu, J. A. Silvester and A. Polydoros, "A general performance model for mobile slotted ALOHA networks with capture," *Proceedings of IEEE ICC*, Vol. 3, pp. 1582-1586, 1995.

[37] T.-K. Liu and J. Silvester, "The maximum throughput and the maximum balanced throughput of mobile slotted ALOHA networks," *Technical Report, CENG 95-24, University of Southern California*, 1995.

[38] J. H. Lodge, "Mobile satellite communications systems: Toward global personal communications", *IEEE Commun. Mag.*, Nov. 1991, pp. 24-30.

[39] N. Mandayam, J. Holtzman and S. Barberis, "Erlang capacity for an integrated voice/ data DS-CDMA wireless system with variable bit rate sources", *IEEE PIMRC'95*, pp.1078-1082.

[40] R. Mathar and A. Mann, "Analyzing a distributed slot assignment protocol by Markov chains," *Proc. 1992 IEEE Vehicluar Techn. Conf.*, pp. 715-720.

[41] J. J. Metzner, "On improving utilization in ALOHA networks," *IEEE Trans. Commun.*, April, pp. 447-448, 1976.

[42] M. Naghshineh and A. S. Acampora, "QOS provisioning in micro-cellular networks supporting multimedia traffic", *IEEE INFOCOM*, pp. 1075-84, 1995.

[43] C. Namislo, "Analysis of mobile radio slotted ALOHA networks," *IEEE Trans. Veh. Technol.*, vol. VT-83, no.3, pp. 199-204, Aug. 1984.

[44] Y. Onozato, J. Liu, and S. Noguchi, "Stability of a slotted ALOHA system with capture effect," *IEEE Trans. Veh. Techn.*, vol. 38, no. 1, pp. 31-36, Feb. 1989.

[45] K. Pahlavan and A. H. Levesque, "Wireless data communications", *Proc. of the IEEE*, Sept. 1994, Vol. 82, no. 9, pp. 1398-1430.

[46] K. Pahlavan and A. H. Levesque, *Wireless Information Networks*, Wiley, 1995.

[47] P. Pancha and M. El Zarki, "MPEG coding for variable bit rate for video transmission," IEEE Communications Magzine, vol. 32, no. 5, pp. 54-66, May 1994.

[48] J. D. Parson. *The mobile radio propagation channel.* BPCC Wheatons Ltd, Exeter. 1991.

[49] C. V. D. Plas and J.-P. M. G. Linnartz, "Stability of mobile slotted ALOHA network with Rayleigh fading, shadowing, and near-far effect," *IEEE Trans. Veh. Techn.*, Vol. 39, No. 4, pp. 359-66, Nov. 1990.

[50] A. Polydoros, A. Anastasopoulos, T.-K. Liu, P. Panagiotou, and C.-M. Sun, "An Integrated Physical/Link-Access Layer Model of Packet Radio Architectures", *California PATH Research Report UCB-ITS-PRR-94-20*, 1994.

[51] T. S. Rapapport *et al.*, "Statistical channel impulse response models for factory and open plan building radio communication system design", *IEEE Trans. Commun.*, vol. 39, no. 5, pp. 794-807, May 1991.

[52] P. D. Rasky, G. M. Chiasson, D. E. Borth, and R. L. Peterson, "Slow frequency-hop TDMA/CDMA for macrocellular personal communications", *IEEE Personal Communications*, vol.1, no.2, pp. 26-35, 1994.

[53] L. G. Robert, "ALOHA packet system with and without slots and capture," *Comput. Commun. Rev.*, No. 5, pp. 28-42, 1975.

[54] A. J. Rustako *et al.*, "Radio propagation at microwave frequencies for line-of-sight microcellular mobile and personal communications", *IEEE Trans. Veh. Techn.*, vol. 40, no. 1, pp. 203-10, Feb. 1991.

[55] S. A. Sachs and P. Varaiya, "A communication system for the control of automated vehicles", *Draft of research report*, UC Berkeley, Nov. 1992.

[56] C. Shim, I. Ryoo, J. Lee, and S. Lee, "Modeling and call admission control algorithm of variable bit rate video in ATM Networks," *IEEE Journal on selected areas in communications*, vol. 12 (no. 2):332-44, Febrary 1994.

[57] M. S. Smith and L. Neal, "A comparison of polarisation and space diversity for indoor propagation at 900 MHz", *2nd International Conference on Universal Personal Communications, IEEE*, 1993, Vol. 1, pp. 74-78.

[58] M. Soroushnejad and E. Geraniotis, "Multi-access stratehies for an integrated voice/ data CDMA packet radio network", *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, Feb./ Mar./ Apr., 1995, pp. 934-945.

[59] "Special issue on packet radio networks", *Proc. of IEEE*, Jan. 1987.

[60] G. M. Stamatelos, "Design considerations for a broadband indoor wireless system", *International Conference on Universal Personal Communications*, pp. 900-5, 1993.

[61] W. K. Tam and V. N. Tran, "Propagation modelling for indoor wireless communication", *Electronics & Communication Engineering Journal*, Oct. 1995, vol.7, no.5, 221-228.

[62] A. Valdovinos, Navas, A. Gelonch and F. J. Casadevall, "Performance of soft-output equalization and convolutional coding over frequency-selective fading radio channels", *Symposium on Personal, Indoor and Mobile Radio Communications*, Vol. 1, pp. 209-213.

[63] J. Van Rees, "Measurements of impulse response of a wideband radio channel characteristics for rural, residential, and suburban areas", *IEEE Trans. Veh. Techn.*, 1987, VT-36, no. 1, pp. 2-6.

[64] G. Wu, A. Jalali and P. Mermelstein, "On channel model parameters for microcellular CDMA systems", *IEEE Trans. Veh. Techn.,* Aug. 1995, vol.44, no.3, pp. 706-11.

[65] H. H. Xia et al., "Radio propagation measurements and modelling for line-of-sight microcellular systems," *IEEE Vehicluar Techn. Conf.*, pp. 349--54, 1992.

[66] Chen Xuan and J. Y. Chouinard, "Adaptive equalization of frequency selective indoor wireless communication channels", *IEEE Canadian Conference on Electrical and Computer Engineering*, 1994, vol. 1, pp. 226-229.

[67] F. Yegenoglu, B. Jabbari and Y.-Q Zhang, "Motion-classified autoregressive modeling of variable bit rate video," *IEEE Trans. on Circuits and Systems for Video Technology,* Vol. 3, No. 1, pp.42-53, Feb. 1993.

[68] K. Zhang and K. Pahlavan, "Relation between transmission and throughput of slotted ALOHA local packet radio networks," *IEEE Trans. Commun.*, Vol. 40, No. 3, pp. 577-583, 1992.

[69] W. Zhuang, W. A. Krzymien and P. A. Goud, "Trellis-coded CPFSK and soft-decision feedback equalization for micro-cellular wireless applications", *Wireless Personal Communications,* 1994-1995, vol.1, no.4, pp. 271-285.

[70] A. Zigic, "Adaptive equalization of the Rician indoor radio channels", *IEEE Symposium on Communications and Vehicular Technology*, 1994, pp. 166-73.