

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the original text directly from the copy submitted. Thus, some dissertation copies are in typewriter face, while others may be from a computer printer.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyrighted material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is available as one exposure on a standard 35 mm slide or as a 17" × 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. 35 mm slides or 6" × 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.



Accessing the World's Information since 1938

300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA



Order Number 8801678

Circuit-switched routing in integrated services digital network environment

Aminjafari, Ayat, Ph.D.

City University of New York, 1987

Copyright ©1987 by Aminjafari, Ayat. All rights reserved.

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106



PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark .

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received
16. Other _____



CIRCUIT-SWITCHED ROUTING IN INTEGRATED SERVICES

DIGITAL NETWORK ENVIRONMENT

by

AYAT AMINJAFARI

A dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1987

© 1987

AYAT AMINJAFARI

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

September 22, 1987
Date

Faach
Chair of Examining Committee

9/22/87
Date

Jacques E. Benveniste
Executive Officer

Professor Sorin Davidovici

Professor Donald Schilling

Professor Richard Tolimieri

Supervisory Committee

The City University of New York

Abstract

CIRCUIT-SWITCHED ROUTING IN INTEGRATED SERVICES
DIGITAL NETWORK ENVIRONMENT

by

Ayat Aminjafari

Adviser: Professor Tarek Saadawi

In circuit-switched networks, routing is a procedure in which a path is assigned to a call between the source node and the destination node. In this research, the routing problem in a circuit-switched ISDN environment is addressed. This research is felt to be necessary to provide some insight into the problems associated with the performance evaluation of multi-user communication networks such as CATV and, most importantly, ISDN. This work consists of three parts. In the first part, we present an n -dimensional birth-death steady-state traffic model to evaluate the blocking probability experienced by n different classes of users in a network. With n equal to two (i.e., two groups of users), the model is then used to determine the circuit-switched capability of CATV systems with a tree-like topology.

In the second part, the circuit-switched routing in an ISDN environment is introduced and a mathematical model for

alternate routing is developed. In this model we consider two types of users - home users and business users. The performances of two different routing procedures in fully connected networks are compared. These procedures include alternate routing and non-alternate routing. The performance criteria used in the comparison are the average network blocking rate and end-to-end blocking probability. For light to moderate traffic loads, the results show that alternate routing performs better than non-alternate routing. Under heavy traffic conditions, the performance of alternate routing deteriorates as compared to that of non-alternate routing.

In the last part, the heavy load performance of alternate routing is discussed and the trunk reservation policy for the case of two groups of users is introduced. To implement this policy, two cases are considered : 1) non-uniform and 2) uniform trunk reservation schemes. The results show that the trunk reservation scheme can be used as an effective way to overcome the performance degradation caused by uncontrolled alternate routing.

ACKNOWLEDGEMENTS

I wish to express sincere thanks to my thesis advisor, Prof. Tarek Saadawi, for his support, guidance, encouragement and friendship during the course of this research.

I am deeply indebted to my dearest wife, Heidi, for her devotion, patience and help.

Table of Contents

| | | |
|--------------|---|----|
| Chapter I: | Introduction..... | 1 |
| Chapter II: | Blocking Probability in Multi-User Environment | |
| | 1. Introduction..... | 17 |
| | 2. Uniform Traffic Distribution With One Group of Users..... | 22 |
| | 2.1 Numerical Analysis..... | 25 |
| | 3. Uniform Traffic Distribution With Two Groups of Users..... | 26 |
| | 3.1 A Recursive Formula..... | 31 |
| | 3.2 Average End-To-End Blocking Probability..... | 31 |
| | 3.3 Numerical Example..... | 33 |
| | 4. Uniform Traffic Distribution With more than Two groups of Users..... | 36 |
| | 5. Conclusions..... | 37 |
| | Figures..... | 39 |
| Chapter III: | Circuit-Switched Routing in ISDN Environments | |
| | 1. Introduction..... | 47 |
| | 2. General Alternate Routing Network Model... | 53 |
| | 3. Mathematical Analysis | 56 |
| | 3.1 The End-to-End Blocking Probability... | 60 |
| | 3.2 The Average Network Blocking Rate..... | 61 |

| | |
|---|-----|
| 4. Alternate Routing in Fully Connected Networks..... | 61 |
| 5. Non-Alternate Routing in Fully Connected networks..... | 63 |
| 6. Two-Link Alternate Routing..... | 64 |
| 6.1 End-to-End Blocking Probability..... | 66 |
| 7. Numerical Analysis..... | 69 |
| 8. Conclusion..... | 71 |
| Figures..... | 73 |
| CHAPTER IV: Alternate Routing With Trunk Reservation in ISDN | |
| 1. Introduction..... | 77 |
| 2. Non-Uniform Trunk Reservation..... | 84 |
| 3. Uniform Trunk Reservation..... | 89 |
| 4. Alternate Routing With Trunk Reservation.. | 96 |
| 5. Numerical Examples..... | 102 |
| Conclusions and Suggestions for Future Research..... | 108 |
| Figures..... | 112 |
| Appendix | 124 |
| References | 125 |

List of Tables

Chapter II:

- Table I. Mean Number of Video frames per Second, λ_1 , for a given Blocking Probability, BP, and Number of Channels per Trunk, c45
- Table II. Mean Number of Video frames per Second, λ_1 , for a given Blocking Probability, BP, and Number of Channels per Trunk, c46

Chapter IV:

- Table I. End-to-End Blocking Probability.....119
- Table II. End-to-End Blocking Probability.....120
- Table III. Network Parameters used in Figure 9.....123

List of Figures

Chapter II:

- Figure 1. Two-Way Distributed Circuit-Switched CATV System.....39
- Figure 2. Blocking Probability vs. offered load for $n=1000$ and $m=5$40
- Figure 3. Blocking Probability vs. offered load for $n=2000$ and $m=5$41
- Figure 4. State diagram for two groups of users (case 2).....42
- Figure 5. Overall blocking probability vs. a_1 for $m=5$, $n_1=1580$, and $n_2=20$43
- Figure 6. Overall blocking probability vs. a_1 for $m=5$, $n_1=3160$, and $n_2=40$44

Chapter III:

- Figure 1. A distributed network with 5 nodes and 6 duplex links.....73
- Figure 2. Route tree for source-destination node pair BD.....74
- Figure 3. A 5-node fully connected network.....75
- Figure 4. End-to-end blocking probability vs. a_176

Chapter IV:

- Figure 1. Performance comparison of non-alternate and alternate routing procedures.....112

| | |
|--|-----|
| Figure 2. Non-Uniform Trunk Reservation..... | 113 |
| Figure 3. Uniform Trunk Reservation..... | 114 |
| figure 4. End-to-End Blocking Probability vs. a_2 | 115 |
| figure 5. End-to-End Blocking Probability vs. a_2 | 116 |
| figure 6. End-to-End Blocking Probability vs. a_2 (non-uniform trunk reservation with $t_1=1$ and $t_2=3$)..... | 117 |
| figure 7. End-to-End Blocking Probability vs. a_2 (non-uniform trunk reservation with $t_1=2$ and $t_2=3$)..... | 118 |
| figure 8. End-to-End Blocking Probability vs. a_2 for $c=10$ and 15 | 121 |
| figure 9. Effect of c on alternate routing..... | 122 |

CHAPTER I

INTRODUCTION

With the growing demand for a communication network with the potential capability of offering more services to the users sharing the network resources, a great deal of effort is being made to design and develop new communication facilities.

Existing telecommunication systems which are designed for different types of services and their duplicity can noticeably be reduced if these systems are interconnected into a single network. In a uniform network with the potential capability of supporting more services, any user can have access to a wide range of information which is otherwise available only by a number of separate networks. Having a single network to carry out all types of information will obviously reduce the time and energy expended by the separate networks providing the same services.

The concept of having only a single network capable of transmitting all types of information has been attractive for many years; however, due to the only existing analog technology tailored specifically for voice communications, differences between voice and non-voice information characteristics and other limitations, the solution to this problem was less feasible twenty years ago.

Since the introduction of digital technology two decades ago, research in this area has been intensified and has received new attention.

For the past one hundred years, telephone networks based on analog technology have mainly been used for voice communication. With the widespread use of telephone networks and the increasing use of computers and digital technology on one hand, and the desirability of having a single network supporting not only voice but also non-voice information on the other hand, telecommunication suppliers have already started to design and test new information carriers. Moreover, the implementation of digital transmission and digital switching in the existing systems has already been used for many years.

In pursuit of this goal, the integration of all types of communication services has become one of the most promising subjects of the telecommunication research being done today by many vendors. The ultimate goal of this extensive research is to develop an integrated network

called the Integrated Services Digital Network, ISDN.

The ISDN, based on digital transmission and digital switching, is expected to be available to the public network users in the near future. Once these networks are equipped with the ISDN's multipurpose interfaces, a wide variety of services will be integrated into a common digital transmission media and will satisfy any user demands. From the user's point of view, supporting voice and non-voice applications by the ISDN at a price lower than if provided separately is very attractive. Some examples of applications supported by the ISDN are: 1) digital voice 2) data (circuit-switched and packet-switched data) 3) text (telex, teletex, and videotex) and 4) image (facsimile). All these services can be provided with a transmission capacity of 64 kb/s or less. Other services such as music, high-speed computer communication, and videophone will be available at a higher transmission capacity.

Cost effective provisioning of these services and the excellent flexibility provided by the ISDN will drastically change user expectations as well as their productions, and it will consequently create a mixed users' environment. Based on traffic rates and application sizes, there will be several groups of users within this system.

In a telephone network, for example, the uniform users assumption that all customers are considered identical is

justified. This is because the major service that is provided by this system is voice communication; however, in an integrated network such as the ISDN, the availability of different services, as well as the capability of the simultaneous transmission of these services on the same transmission link, will have a great effect on user diversity. Taking advantage of the features of an integrated network, there are users whose needs require a large portion of the network capacity. On the other hand, there exist some users whose demands may be far less than those of the first group of users and still some users with traffic requirements fall somewhere in between.

Considering the recent evolution in the area of communication technology and the emergence of new communication networks, it is obviously of great interest to introduce new models for measuring the relative performances of these networks. It is very important for network designers to take into consideration the wide disparities in transaction sizes, traffic rates, and holding times associated with each group of users. Evaluation of switching performances for circuit-switched or packet-switched networks with several groups of users requires multi-dimensional queueing models for analysis.

The purpose of this research is to investigate the performances of such a network, that is, a network with more than one group of users. Considering the circuit switching

mode of operation, major emphasis is given to the grade of service offered by this kind of network. The grade of service in a circuit switched network is defined as the end-to-end blocking rate imposed on a user. This blocking occurs whenever a user wishing to communicate with another user fails to establish a path to its destination due to the lack of network resources.

This research is mainly motivated by the application of ISDN multi-purpose interfaces which are expected to play a major role in the market of telecommunication industries in the future. The initiative towards this research was originally taken when we were studying the problems of data transmission over the Community Antenna Television (CATV). The concept of employing CATV based on a distributed switching model had been considered first by [1]. In addition to the installation of so called "intelligent switches" at various points within the CATV network proposed by [1], several issues associated with packet communication networks designed for transmitting bursty or low duty cycle traffic were also addressed. It was shown that using these switches would prevent the congestion problems, lower average message delay, and increase traffic throughput. On the other hand, the transfer of different types of information might require different switching techniques. In addition to packet switching, emphasis is given to circuit

switching and hybrid switching.

Shortly after the evaluation of related circuit switching performances of the proposed CATV system was begun, we realized that the traditional model used in the evaluation of performance functions in circuit-switched networks with only one group of users needed to be modified. CATV networks, supplemented by the telephone network, can provide access to every household and business. Hence, in the analysis of this sort of network, one has to deal with a communication medium in which various classes of users could be present simultaneously.

From the analysis standpoint, in order to understand the status and the behaviour of a network with more than one group of users, it is very important to adopt a mathematical model which describes the statistical behaviour of the network as precisely as possible. The accuracy of measurement of the network performances is directly influenced by the model used in the analysis of such a network.

In Chapter II, we first consider a system consisting of a trunk with c -channels connecting two adjacent nodes of a circuit-switched network. We then assume that each node contains two classes of users with different traffic rates and holding times. Using the Markov-Chain queueing model together with the Poisson arrival process and exponential holding times assumptions, we set up a loss queueing system

with two finite numbers of two groups of users. We then use this model to find a closed form solution for the link blocking probability. Next, the link blocking rates are applied to a CATV network with a tree-like topology to determine the circuit-switched capacity of the CATV network. We analyze the performance in terms of end-to-end blocking probability. It is assumed that the switches are nonblocking and, hence, we focus on the lack of channel availabilities. Assuming, furthermore, that the link blocking probabilities are statistically independent from each other, an exact equation for the end-to-end blocking probabilities between source-destination node pairs is given.

Following the presentation of a recursive formula to calculate the link blocking rates more effectively, the average end-to-end blocking probability is defined and some typical examples are given. Plots and tables are provided for comparison purposes. It is shown that introducing intelligent switches in distributed CATV networks will significantly reduce the average blocking rates. Finally, in Chapter II, we shall extend our analysis to the case of more than two groups of users, and a steady-state equation for representing the state probability of the system will be derived.

Link blocking probabilities obtained in Chapter two provide a basis for the network objective function in order

to address and discuss the impact of the related design issues on the network performance. During the design period of any circuit-switched communication network and the time that it is operating there exist several issues which have to be studied at various stages. In general, these issues are concerned with the total cost allowed to construct the system, the network topology, protocol, routing, channel capacity assignment, traffic flow control, etc.

The next part of this research focuses on the routing problem of circuit-switched networks with two groups of users.

In a communication network, one of the most important tasks in transmitting a message from a source node to its destination node is to search for and select the best possible path from among a set of feasible paths with respect to some network performance. In a network with more than one path between any source-destination node pair, routing procedure plays an important role in the utilization of network resources.

However, before we address the routing problem, it is necessary to define and establish a set of network performance criteria. Network performance criteria are used in the comparison of different routing procedures. For example, in packet-switched networks time delays and network throughputs are the most important performance criteria used in the analysis. On the other hand, a circuit-switched

network's designer is usually concerned with the end-to-end blocking probability and the average system blocking. The first criterion is selected in the user's optimization problems to minimize the individual source-destination path blocking rate. The second performance function, the average system blocking, represents the global performance of the network and gives overall information about the network capability. Furthermore, the average system blocking rate provides a measure of the system response to the input traffic and is, hence, the most common performance function used in the assessment of circuit-switched network capacity.

In a circuit-switched network with a fixed cost, the end-to-end blocking probability and/or the average system blocking probability could be generally minimized over the design variables such as the link capacity assignment, the network topology, queueing discipline, and the routing technique. As mentioned earlier, however, we have assumed that the network topology and link capacities are fixed and main attention is given to the routing doctrine.

The routing doctrine in a communication network is defined as a set of rules that is used to determine the best possible paths between a group of source-destination pairs on the basis of some optimal criteria. In addition to assigning proper paths to all users, the routing doctrine must be able to respond quickly to link and node failures.

Therefore, the assurance of the rapid delivery of information between users is the responsibility of the routing doctrine.

In a communication network, during the call set up phase, a call arriving at a node will be routed from its incoming link to one of the outgoing links according to the routing procedure. This process is repeated at each intermediate node until either the call reaches its destination or is blocked at a node along the path after failing to find a free link. Depending on the routing doctrine employed, the blocked call may be rerouted to another path or it may be dropped out and considered a lost call.

Using an improper routing procedure could severely damage the network performance and cause high blocking probabilities. Therefore, for a given network, it is extremely important to adopt a routing algorithm which is able to manage and control the network resources shared by the network customers. Because of the close relationship between the network performance and routing strategies, the efficiency of a circuit-switched network is heavily influenced by the various routing procedures.

From the network's status point of view (such as the average traffic load and network topology), routing algorithms in circuit-switched networks are classified into two categories. The first category consists of fixed

(deterministic) algorithms. A fixed routing algorithm selects routes based upon a given deterministic decision rule. Fixed routing algorithms are designed on the basis of the average forecasted load and fixed network topology. Routing rules are time-fixed and so cannot respond to the load and topology variations.

The second category includes adaptive (dynamic) routing algorithms in which routes are assigned to calls on the basis of network topology and/or traffic requirements. Therefore, adaptive routing rules may vary from time to time as the network's status changes and consequently the path between a source-destination node pair may vary with the network's traffic condition.

In real life, an operating network must be capable of adapting to the changes in the network. Thus, an adaptive routing strategy which is able to accommodate some or all changes in the network is desirable. An ideal adaptive (dynamic) routing scheme will optimize the routing of all calls in the network and will improve network efficiency and performance.

Based on the state of the network available for routing decisions, the frequency of updating the information, the strategy used, and, most importantly, the availability of the tools needed, adaptive routing algorithms can furthermore be divided into different groups.

Most of the existing circuit-switched networks in the United States and abroad use some form of semi-adaptive routing policy called alternate routing [2]. The alternate routing policy in a distributed circuit-switched network provides an effective use of network resources and is intended to lower the end-to-end blocking probability. Moreover, all the operational routing policies have been practically designed according to some sort of hierarchical structure. Besides the semi-adaptive characteristic of hierarchical routing, this class of alternate routing is widely used for three basic reasons: 1) limitations on switching for rapid delivery of information signals 2) extensive computational requirements for the end-to-end blocking probabilities and 3) amount of memories required to store information concerning the network's instantaneous state.

In a network using hierarchical routing, routes are assigned to different levels to form a hierarchy [3]. In telephone networks, for example, a name is also provided for each level of the hierarchy. The lowest level is called the "direct trunk group" and the highest level is referred to as the "final trunk group". In hierarchical routing networks, the search for free path starts from the first level (the direct trunk group), and moves to the next level, if necessary, until either it finds a free path or it reaches the final trunk group. A call is then said to be blocked if

a free path is not found in the final trunk group.

Considering the development of switching technology in the last century, circuit switching technology has evolved from its earlier switching plan based on manual switching to electro-mechanical switching and finally to computer-controlled switching. The present circuit switching technology in a stored program control (SPC) network with the out-band common channel interoffice signaling (CCIS) has greatly eliminated most of the switching limitations discussed above. It has provided a switching mechanism practically free of operating irregularities. Some of the current switching systems can support as many as 50,000 simultaneous voice calls [4]. With computer-controlled electronic switching, it is now becoming possible to look ahead and extend current telephone network routing rules far beyond the conventional hierarchy.

Recently, Ash, Cardwell, and Murray in [5] looked at a set of routing strategies called "dynamic nonhierarchical routing". Taking advantage of switching facilities provided by the stored program control (SPC) network, they considered three routing methods in order to minimize the cost of the network. These methods are: progressive routing, multilink path routing, and two-link path routing. They have shown that an overall savings is possible when using these methods as compared to present hierarchical techniques. From the

savings results and implementation considerations, it is also concluded that two-link routing is preferred.

When considering the above factors, coupled with the emergence of the ISDN multi-media environment, we are being motivated towards the need for further research to study different routing plans as compared to present ones.

Using the model developed in Chapter II, different routing algorithms are discussed in Chapter III. We begin the discussion with the performance analysis of two routing procedures called non-alternate routing and alternate-routing respectively, without any traffic flow control. They differ basically in the amount of traffic allowed to attempt various alternate paths in the network and are taken as the representatives of two extreme cases.

As will be illustrated, one problem with alternate routing is that the network can become congested in the presence of certain traffic patterns. This stability problem occurs whenever the direct paths are heavily loaded. In the absence of an effective control policy, the alternate paths, which are themselves the direct paths for some other traffic, will be overloaded. When this happens, traffic in the network will back up and cause a higher end-to-end blocking probability. This stability problem in the blocking probability, using a nonhierarchical network with only one group of users, is also discussed by [6], [7], and [8].

The stability problem in alternate routing networks can

always be avoided by applying some form of control mechanism on the routing rules. It is done by blocking the alternately routed traffic so that the routing rules will not over react. For example, a control strategy called the "trunk reservation scheme" is suggested by [6], [7], and [8]. This scheme is used to block the traffic overflowed by direct paths in an attempt to overcome the stability problem which arises during heavy traffic conditions.

In Chapter IV, using the model of two groups of users, we focus on the trunk reservation scheme. In a community of heterogeneous users, the trunk reservation technique has not been discussed before. For the case of two classes of users, two possible methods of implementing the trunk reservation technique have been introduced in Chapter IV. In one approach, a certain number of channels are specifically reserved for direct calls. This is done in order to protect the direct calls and to avoid the performance degradation that is caused by alternate routing during overload conditions. We refer to this case as the "uniform-trunk reservation" method. Another approach is to combine the trunk reservation rule with a queueing discipline in which priority for using alternate paths is given to one of the two classes of users. We will later refer to this case as the "non-uniform trunk reservation" scheme.

Analytical analysis of both trunk reservation

techniques is carried out on a fully connected network basis and the numerical examples are given. The results of trunk reservation schemes have shown that an appropriate trunk reservation algorithm can prevent congestion and operate within an acceptable range of network performance.

CHAPTER II

BLOCKING PROBABILITY IN MULTI-USER ENVIRONMENT

1. INTRODUCTION

With the great demands for high capacity and megabit range communications, the communication industry has made a successful effort to develop new systems to meet the needs of our society. The new systems should provide more efficient, economical, and reliable data communication services.

Recently, community antenna television (CATV) networks have received a great deal of attention not only in the United States, but across the world. The broad-band CATV systems have a capacity 50 times greater than that of the baseband systems. These networks have the ability to serve population areas ranging from 10,000 to 100,000 subscribers. They can carry several independent communication paths in contrast with the single data path of baseband systems, and

bring many types of services to today's home and business environments. The high signal-to-noise ratio S/N in a properly designed CATV based cable plant permits high-quality analog and low-error digital propagation.

Most of the present broad-band CATV networks (including the ones used for two-way interactive communication) have focused on the centralized approach, in which the retransmission facilities are placed at the cable head-end. With a centralized approach, all messages must first be transmitted to a switch located at the head-end, and then retransmitted over the same bus to the proper destination.

However, if the size of the network is large enough to support a large number of stations accessing the network, it may be necessary to search for another approach since the central switch, acting as a bottleneck, will reduce the efficiency of the network. Reference [1] addresses the problem of data transmission over two-way CATV using packet switching techniques. It introduces an alternative model, called the "distributed switching model", which prevents problems of congestion and has several advantages over the centralized case. Then, it looks at some performance criteria associated with packet communication networks designed for transmitting bursty or low duty cycle traffic. These performance parameters include the traffic flow along the network, traffic throughput for each node, and the average message delay. Reference [9] presents the protocol

architecture for such a network. Concomitantly, however, the transfer of different types of information may require different switching techniques. In addition to packet switching, these include circuit switching or hybrid switching. Circuit-switched CATV systems can offer many services to the public which may not be affordable or theoretically obtainable by most of the present telecommunication networks. CATV systems, supplemented by the telephone system, can provide access to almost every household and business. They can support many applications such as voice, data, video, facsimile reproduction, etc.

Most circuit-switched communication networks are designed on a blocking basis. This implies that a certain number of calls will temporarily be rejected by the system. A circuit-switched network will be considered useful if it blocks calls at a rate below the acceptable blocking probability which may vary for different applications.

In this chapter, an n -dimensional birth-death steady-state traffic model is used to evaluate the blocking probability experienced by n different classes of customers. In particular, with n equal to two, i.e., two classes of customers, we are interested in determining the circuit-switched capability of CATV systems with a tree-like topology. To do this, we analyze the performance in terms of end-to-end blocking probability. A large number of papers

have addressed end-to-end blocking probability in circuit-switched networks; examples appear in [2, 10, 11, 12]. Each paper is essentially devoted to a distinct model of a telecommunication system. In this chapter, we carry out a calculation of the blocking probability appropriate to a network with more than one class of customer, such as a CATV system.

Blocking may occur for two reasons: 1) switch failure and 2) the lack of channel availability. We assume here that the switches are nonblocking and, hence, focus on the lack of channel availability only.

In the analysis of end-to-end blocking probability presented here, we have made the following assumptions:

1) The link blocking probabilities are statistically independent from each other.

2) The call interarrival times form a Poisson arrival process regardless of the type of traffic.

3) The number of channels (servers) is the same for all the links.

4) All the stations within the system are assumed equally likely to communicate with one another, i.e., uniform traffic distribution.

5) No priority is given to the users with large traffic requirements.

The independence assumption is the most common one made in the analysis of blocking probability, while the second

assumption, that of a Poisson process, is the standard assumption in queueing theory. The next three assumptions are basically made to simplify the problem as much as possible. In a carefully designed network, one might want to ignore the last assumption by favoring one class of users over another. This could be done by allocating more channels to the users with large traffic requirements (see, for example, [13]).

To calculate the end-to-end probability of blocking, we need to compute the link blocking probability for those links that constitute the path between the call's originating node and the destination node. The first assumption allows us to use queueing theory to obtain the blocking probability related to each link separately. The availability of each link ($1 -$ the blocking probability) depends upon two quantities: 1) link offered load and 2) number of transmission circuits or channels supported by the link. Link offered load depends on the traffic distribution model and the traffic generated by the users, while the second quantity, numbers of channels, is chosen to meet the design requirements.

Following this introduction, the analysis and computational results for a single group of users are discussed in Section 2. Section 3 presents the case for two groups of users with numerical examples included. Section 4 extends

the analysis to the case of more than two groups of users. An appendix describes the recursive formula used in Section 3 to calculate the blocking probability.

2. UNIFORM TRAFFIC DISTRIBUTION WITH ONE GROUP OF USERS

The network topology to be used in this section is shown in Fig. 1. This network model includes the main trunk with intelligent circuit switches installed at different locations on the trunk. The stations (users) are attached to the branches. Let m be the number of switches on the main trunk and n be the number of stations attached to the branches. In the analysis of point-to-point congestion in a two-way distributed CATV system given below, we have assumed Poisson arrivals with arrival rate λ and arbitrary service-time distribution with mean holding time $1/\mu$. (For a loss system the blocking probability does not depend on the service time distribution [14].)

Let

$a = \lambda/\mu$ - the customer offered load in erlangs.

$n(i+1,i)$ - the number of customers (or circuits) using the link between the i^{th} and $i+1^{\text{st}}$ switches in the upstream direction.

$n(i,i+1)$ - the number of customers (or circuits) using the link between the $i+1^{\text{st}}$ and i^{th} switches in the downstream direction.

With the assumption of a uniform traffic distribution, the total number of stations (customers) sharing the link between two successive switches is given by

$$n(i,i+1) = n(i+1,i) \triangleq n_i = \left[\left(1 - \frac{i}{m}\right)ni \right]^* \quad (1)$$

Here $[X]^*$ denotes the smallest integer equal to or greater than X . Equation (1) is derived as follows (see also [1]). Nodes 1, 2, ..., i have a total of (ni) users (or circuits) attached to their branches. Because of the uniform distribution assumption, i/m of these circuits will be destined to the branches attached to nodes 1, 2, ..., i , while $(1 - i/m)$ of these circuits will be destined upstream to nodes $i+1, i+2, \dots, m$. As a result of this symmetrical property, to calculate the link availability we need only to compute the blocking for one direction and use the results to find the congestion between any two terminals in the system. We have assumed, furthermore, that the blocking of each link is sufficiently small that the effect on any other link could be negligible, i.e., the Poisson model assumption is not disturbed for those links on the main trunk, and the link availabilities are independent from each other. Under these assumptions, the Engset formula with blocked calls cleared can be used to estimate congestion for the link between two successive nodes. The link probability of

blocking with c channels per link is then given by

$$P_i(c) = \frac{\binom{n_i}{c} a^c}{\sum_{j=0}^c \binom{n_i}{j} a^j} \quad (2)$$

Note that if the number of users is very large, the simpler Erlang B formula can be used instead of the Engset formula. For the local links (the links between any branch and its corresponding switch), the probability of blocking is given by the same equation except that the number of customers is constant and equal to n . This probability is simply referred to by $P(c)$ here. Having found all the link blocking probabilities, one can use them to obtain the one-way point-to-point congestion between any two customers. This is obtained as follows: suppose station A_i in branch i , attached to the (i) th switch, wishes to communicate with station B_j in branch j . The one-way probability that station A_i will fail to establish a path all the way to station B_j due to the lack of channel availability is given by

$$BP1(A_i, B_j) = (1 - P(c))^2 \prod_{l=i}^{j-1} [1 - P_l(c)] \quad (3)$$

The second term represents the availability of the branch links (source and destination links) times the availability of the $(j - i)$ trunk links located between stations A_i and B_j on the main trunk. As mentioned earlier, the blocking

probability for both directions is the same. The two-way end-to-end probability of blocking for station A_i and B_j is then given by

$$BP2(A_i, B_j) = 1 - [1 - BP1(A_i, B_j)]^2 \quad (4)$$

Although the Engset formula, given by (2), is well known, it is somewhat inconvenient when used to evaluate numerically, especially when c and n are large. The following recursive version can be used, instead of (2), to evaluate the probability of blocking, $P_i(c)$.

$$P_i(k) = \frac{1}{1 + \frac{k}{a(n_i - k + 1)} P_i(k-1)} \quad 1 \leq k \leq c \quad (5)$$

with $P_i(0) = 1$.

2.1 Numerical Analysis

In order to obtain the blocking probability, $BP2(A_i, B_j)$, in a two-way distributed circuit-switched network and to compare the result with the case of a centralized switching network, we have considered the special case of a network with five switches. Figures 2 and 3 show the resultant blocking probability, $BP2(A_i, B_j)$, as function of offered load per user, a , for $n = 1000$ and 2000 users/branch,

respectively. To examine the effect of the number of channels on the blocking probability, we have obtained the results for various numbers of channels in our example ($c=5, 7, \text{ and } 10$), as indicated in Figs.2 and 3.

In each figure, the dotted line shows the two-way blocking probabilities of any two users in the centralized network (obtained by replacing $m = 1$ and n by the total number of customers in the system), whereas the solid lines illustrate the blocking probability for different users with regard to their positions in the distributed network.

As expected, for the same number of channels and the same offered traffic, a distributed circuit switching two-way system provides lower blocking probability than the centralized system. Hence, for the same number of channels and the same blocking rates, the distributed case can handle more offered traffic than the centralized case. This is understandable since in the uniform traffic model assumed for the distributed case, the average number of simultaneous circuits is higher than for the centralized switching case.

3. UNIFORM TRAFFIC DISTRIBUTION WITH TWO GROUPS OF USERS

In many applications, it is desirable to have more than one group of users. Such a group is defined as one with its own holding time and arrival rate. Any communication

system will be efficient and reliable if it can support various classes of customers. In this section and the following subsection, we extend the analysis of the previous section to that of multiple groups of users. First we start with two groups of users and then in Section 4 we extend our analysis to a more general case in which there exist more than two classes of users. The network topology is the same as before. Let us now introduce the following notation:

n_k = number of class k customers attached to each switch
 ($k = 1, 2$)

λ_k = the mean arrival rate of class k customers

$1/\mu_k$ = the mean holding time of class k customers

a_k = load offered by any type k customer in Erlang.

($a_k = \lambda_k/\mu_k$).

To find the blocking probability over a link, assume first that the number of customers of class k using a link is just n_k . This will be modified later to invoke a variable number of customers over the different links. We form the following state-transition-rate diagram for Poisson arrivals, c -server (c -channels), and a finite population queueing system as shown in Fig. 4. In this diagram (i, j) represents the state in which i type 1 and j type 2 calls are present on the link.

Under equilibrium, the rate of flow into state (i, j) must be equal to the rate of flow out of state (i, j) . This

steady-state requirement yields the following equation:

$$(n_1-i+1)\lambda_1 P(i-1, j) + (i+1)\mu_1 P(i+1, j) + (n_2-j+1)\lambda_2 P(i, j-1) + (j+1)\mu_2 P(i, j+1) = \{(n_1-i)\lambda_1 + i\mu_1 + (n_2-j)\lambda_2 + j\mu_2\} P(i, j), \quad i + j \leq c \quad (6)$$

Here $P(i, j)$ is the probability that there are i type 1 and j type 2 stations on the link.

We must also have

$$\sum_{i=0}^c \sum_{j=0}^{c-i} P(i, j) = 1 \quad (7)$$

In order to solve (6) for $P(i, j)$, we use the product method, i.e., $P(i, j) = R(i) * S(j)$. Note that both $R(i)$ and $S(j)$ are just two auxiliary functions needed to obtain $P(i, j)$. Dividing both sides of (6) by $R(i) * S(j)$ and rearranging the i and j terms into two parts, we have

$$\left\{ [(n_1-i)\lambda_1 + i\mu_1] R(i) - [(n_1-i+1)\lambda_1 R(i-1) + (i+1)\mu_1 R(i+1)] \right\} \frac{1}{R(i)} =$$

$$\left\{ -[(n_2-j)\lambda_2 + j\mu_2] S(j) + [(n_2-j+1)\lambda_2 S(j-1) + (j+1)\mu_2 S(j+1)] \right\} \frac{1}{S(j)}$$

As we can see, the left side of this equation is independent of the right. This can be true if and only if both sides of this equation are equal to a constant, say b . It can be shown that the only constant that satisfies this condition is zero. One can, furthermore, simplify this equation into the following set of two equations:

$$\begin{aligned}
(i+1) R(i+1) - a_1(n_1 - i) R(i) &= 0 \\
(j+1) S(j+1) - a_2(n_2 - j) S(j) &= 0
\end{aligned}
\quad 0 \leq i, j \leq c \quad (8)$$

Here $a_1 = \lambda_1/\mu_1$ and $a_2 = \lambda_2/\mu_2$.

The solutions for $R(i)$ and $S(j)$ are given by

$$\begin{aligned}
R(i) &= a_1^i \binom{n_1}{i} R(0) \quad 0 \leq i \leq c \\
S(j) &= a_2^j \binom{n_2}{j} S(0) \quad 0 \leq j \leq c
\end{aligned} \quad (9)$$

The state probabilities $P(i, j)$ are then given by

$$P(i, j) = a_1^i a_2^j \binom{n_1}{i} \binom{n_2}{j} P(0, 0) \quad (10)$$

Using the probability measure, given by (7), we can compute $P(0, 0)$:

$$P(0, 0) = \left[\sum_{i=0}^c a_1^i \binom{n_1}{i} \sum_{j=0}^{c-i} a_2^j \binom{n_2}{j} \right]^{-1} \quad (11)$$

The link is said to be unavailable if all channels in the link are busy. The probability that a link is blocked is simply the probability that the sum of the customers using a link is equal to the number of channels in the link. Denoting this probability by BP and the number of channels by c , we have

$$BP = \sum_{i+j=c} P(i, j) \quad (12)$$

To determine the blocking probability of a path between two users, we must now account for a different number of customers on each link. Recall that n_1 and n_2 represent the number of type 1 and type 2 customers attached to each switch, respectively. Let $n_1(\ell)$ and $n_2(\ell)$ be the number of corresponding type 1 and type 2 customers supported by the link between the (ℓ) th switch and the $(\ell + 1)$ st switch. $n_1(\ell)$ and $n_2(\ell)$ are given by the following equations, similar to (1) in Section 2.

$$\begin{aligned} n_1(\ell) &= \left[\left(1 - \frac{\lambda}{m}\right) n_1 \lambda \right]^* & 1 < \ell < m \\ n_2(\ell) &= \left[\left(1 - \frac{\lambda}{m}\right) n_2 \lambda \right]^* & 1 < \ell < m \end{aligned} \quad (13)$$

Here $[X]^*$ denotes the smallest integer equal to or greater than X . Using (10)-(12) and replacing all values of n_1 and n_2 by $n_1(\ell)$ and $n_2(\ell)$ as given by (13), the blocking probability of link l is obtained as

$$BP_\ell(c) = P(0,0) \sum_{i+j=c} a_1^i a_2^j \binom{n_1(\ell)}{i} \binom{n_2(\ell)}{j} \quad (14)$$

Using (14) and proceeding as in (3), the one-way end-to-end blocking probability of each path connecting two stations A_i and B_j is given by

$$BP_1(A_i, B_j) = 1 - (1 - BP(c))^2 \prod_{\ell=i}^{j-1} \{1 - BP_\ell(c)\} \quad (15)$$

3.1 A Recursive Formula

In practice, equation (14) given above is difficult to apply from a numerical point of view. In Section 2 we showed that instead of using (2) directly, we can apply the recursive formula given by (5) to compute the blocking probabilities. For the present case of two groups of users, the following recursive formula may be used to obtain the blocking probability for each link,

$$BP_{\lambda}(k) = \frac{1}{1 + \frac{1}{BP_{\lambda}(k-1)} \frac{G_{\lambda}(k-1)}{G_{\lambda}(k)}} \quad 1 \leq k \leq c \quad (16)$$

with

$$BP_{\lambda}(0) = 1$$

and

$$G_{\lambda}(k) = a_1 \sum_{j=0}^k \left(\frac{a_2}{a_1}\right)^j \binom{n_1(\lambda)}{k-j} \binom{n_2(\lambda)}{j} \quad (17)$$

For the derivation of (16), see the appendix.

3.2 Average End-To-End Blocking Probability

From (15) for the two groups of users, or (3) for the case of one group of users, one obtains $BP_l(A_i, B_j)$, the

end-to-end blocking probability between user A_i attached to branch i and user B_j attached to branch j . This end-to-end blocking rate is the same for both types of traffic, type 1 and type 2, since the system does not favor one over the other. However, a network designer will also be interested in knowing the overall end-to-end blocking probability of the system. In order to reduce to practical use the equations so far presented for analysis of the end-to-end blocking probability, we calculate the average blocking probability, BP. To find this, we first obtain the average blocking probability experienced by the users in a given branch i . This is given by

$$P(A_i) = \sum_{j=1}^m \text{BP1}(A_i, B_j) \times \text{Prob.}$$

{user A_i on branch i will communicate with user B_j on branch j } (18)

In our analysis we have assumed a uniform traffic distribution. For this case the second term in the RHS of (18) is constant and equal to $1/m$. Hence,

$$P(A_i) = \frac{1}{m} \sum_{j=1}^m \text{BP1}(A_i, B_j) \quad (19)$$

with m the number of switches. Note that the average blocking probability seen by each branch is different. It is apparent that those users located near the center of the network experience less congestion than the users placed

further away from the center. The overall average blocking probability of the network shown in Fig. 1 is now given by

$$BP = \sum_{i=1}^m P(A_i) \times \text{Prob.}$$

{The traffic is generated on the i-th branch} (20)

Assuming a uniform traffic distribution, as above, one then gets

$$BP = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m BP1(A_i, B_j) \quad (21)$$

3.3 Numerical Example

In CATV systems, where two-way interactive data communication services are available, dedicated TV channels are allocated to such services. The requirement to be considered is the number of such channels that must be allocated to these services. A typical application might be the transmission of a single analog video frame on a circuit-switched basis. The user is assumed to use a separate data channel (not discussed here) to request transmission of the frame. Now consider a CATV network with five switches ($m = 5$) and 100,000 users who subscribe to regular TV entertainment programs. Out of these 100,000 users, 40,000 are assumed to

subscribe to two-way interactive data communication services; we study the effect of using circuit switching. In this example we further consider two typical cases. In the first case 20 percent of the two-way subscribers (i.e., 8000) are active, whereas in the second case 40 percent of the subscribers are active (say in peak periods). In each case, the active users are grouped into two classes: home users (class 1), and business users (class 2). Furthermore, let's assume that the users belonging to class 1 (home users) are much less active than the class 2 users (business users) with the ratio of the individual offered load of the first group to the second group, a_1/a_2 , equal to 1/100 (i.e., $a_2 = \lambda_2/\mu_2 = 100 \lambda_1/\mu_1 = 100 a_1$).

In such a network it would be reasonable to assume that the population of the small users (class 1) is 50-200 times larger than that of the big users (class 2). In the example given here we have assumed that the number of class 2 users per branch is $n_2 = 20$ in the first case and $n_2 = 40$ in the second case. Since the total number of subscribers is 8000 (1600 per branch) and 16,000 (3200 per branch) in the two cases, respectively, we have $n_1/n_2 = 79$.

Figs. 5 and 6 plot the results obtained for the blocking probability BP for various values of c . We illustrate these results with a simple example. Consider a typical analog transmission system where each channel is capable of transmitting 30 video frames every second.

Assume that the holding times are the time required to transmit a single frame (i.e., $1/\mu_1 = 1/30$ s). Given the number of channels and the value of BP, we would like to calculate a theoretical upper bound on the mean number of video frames per second that can be offered by any type 1 or type 2 station. Using Figs. 5 and 6, Tables I and II are obtained.

Tables I and II list the mean number of video frames per second that can be generated by an individual type 1 user. The percentage of blocked calls in a typical communication system usually varies between 0.1 and 5 percent. However, in order to examine the effect of the overall blocking probability BP on λ_1 , we have considered three different values of BP as shown in both tables.

Comparing the results of λ_1 for different values of BP in each row, we observe that the higher BP has little effect on λ_1 since n is large and c is relatively small; however, the effect would be noticeable if n were small and/or c were large.

To make use of the results presented in Tables I and II, let us now look at the following example. Consider single-frame video transfer as a representative example of wide-band traffic. We assume that each home user generates one video frame every 5 min on average and each business user generates one video frame every 3 secs. That is,

$\lambda_1 = 0.01\lambda_2 = 3.34 \times 10^{-3}$ (frame/s). Referring to Table I, we see, for $c = 3$ and a total number of 8000 active users, including 100 business users, that BP is about 0.03. The result will change significantly if we add, for example, two channels to the system. For $c = 5$ and $BP = 0.03$, we find that $\lambda_1 = 10.9 \times 10^{-3}$ (frame/s).

Let us now consider the second case. The total number of active users in this case is equal to 16,000, including 200 class 2 users. In order to compare the results to those provided for the first case, we assume that $BP = 0.03$. Table II indicates that when $c = 3$, $\lambda_1 = 1.75 \times 10^{-3}$ (frame/s) or about one frame every 10 min, compared to one frame every 5 min obtained for the first case. If $c = 5$, $\lambda_1 = 5.34 \times 10^{-3}$ (frame/s). For these values of small c and large n , with BP fixed, λ_1 decreases linearly with respect to n .

4. UNIFORM DISTRIBUTION WITH MORE THAN TWO GROUPS OF USERS

Now we extend our analysis to the case of more than two types of traffic. Let us define the following notation:

g = number of classes of traffic

c = number of channels/link/direction

n_k = number of class k customers

$P(i_1, j_2, \dots, l_g) = \text{prob} [i \text{ type } 1, j \text{ type } 2, \dots, \text{ and } l \text{ type } g \text{ customers are in the system and receiving service}]$

a_k = offered traffic (in erlangs) for type k customers.

The evaluation of steady-state probability is analogous to the one described above (for $g = 2$), and is given by

$$P(i_1, j_2, \dots, \lambda_g) = P(0, 0, \dots, 0) a_1^i a_2^j \dots a_g^\lambda \binom{n_1}{i} \binom{n_2}{j} \dots \binom{n_g}{\lambda} \quad (22)$$

where

$$P(0, 0, \dots, 0) = \left[\sum_{i=0}^c a_1^i \binom{n_1}{i} \sum_{j=0}^{c-i} a_2^j \binom{n_2}{j} \dots \sum_{\lambda=0}^{c-(i+j+\dots)} a_g^\lambda \binom{n_g}{\lambda} \right]^{-1} \quad (23)$$

One can use the state probabilities given by (22) and similar equations to those given by (13)-(15) to calculate point-to-point congestion in the circuit-switched communication network shown in Fig. 1.

5. CONCLUSIONS

The purpose of the study discussed in this chapter is to obtain the state probability function for a circuit-switched network in which more than one group of users can concurrently exist. The service quality of a circuit-switched network is measured in terms of the probability of rejecting a call and the delay encountered by the customers

in accessing the common resources. We use an n-dimensional birth-death steady-state model to evaluate the blocking probability experienced by n different classes of users. Using the state probability function for two types of customers, we consider a two-way distributed circuit-switched CATV network with a tree-like structure, and we evaluate the individual end-to-end and the average system blocking probabilities. While the analysis described here has focused on two-way interactive distributed circuit-switched CATV systems, the results can easily be modified to accommodate many existing networks. As mentioned in the beginning, one of the possible techniques that could be used to improve the efficiency of the network is the channel reservation scheme, in which priority is given to users with large traffic requirements.

In order to make the system more attractive, further study is required to evaluate the performance measures of the network under a hybrid (circuit-packet) switching mode operation.

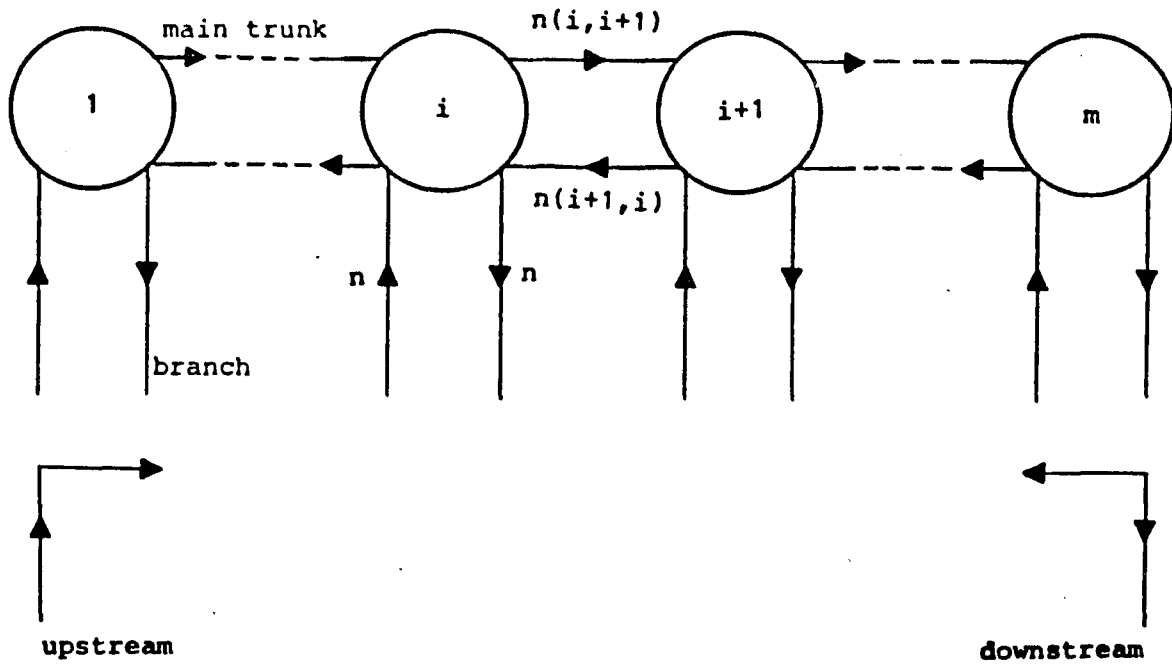


Figure 1. Two-Way Distributed Circuit-Switched CATV System.

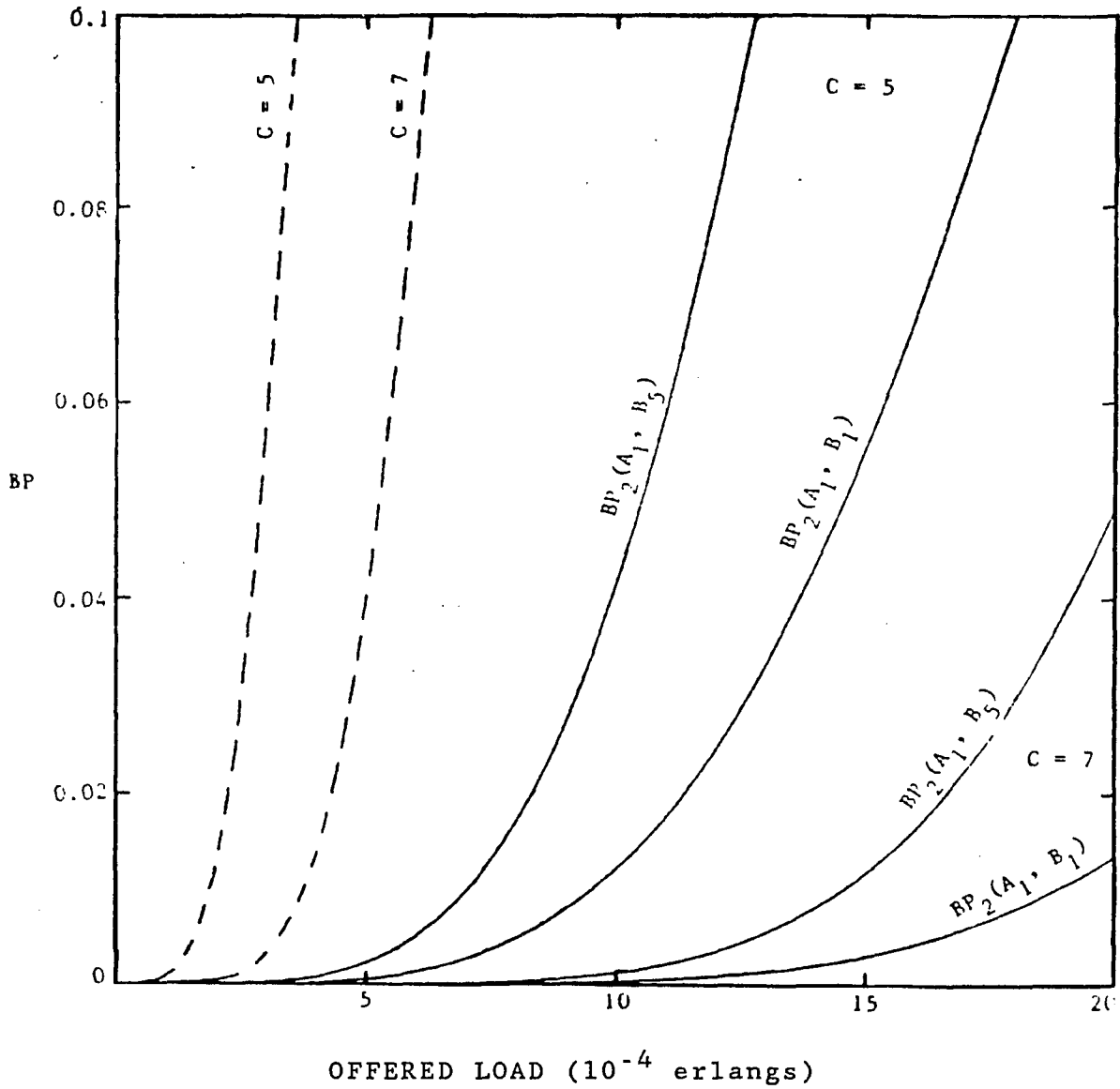


Fig. 2. Blocking probability vs. offered load for $n=1000$ and $m=5$, (total number of active customers = 5000).

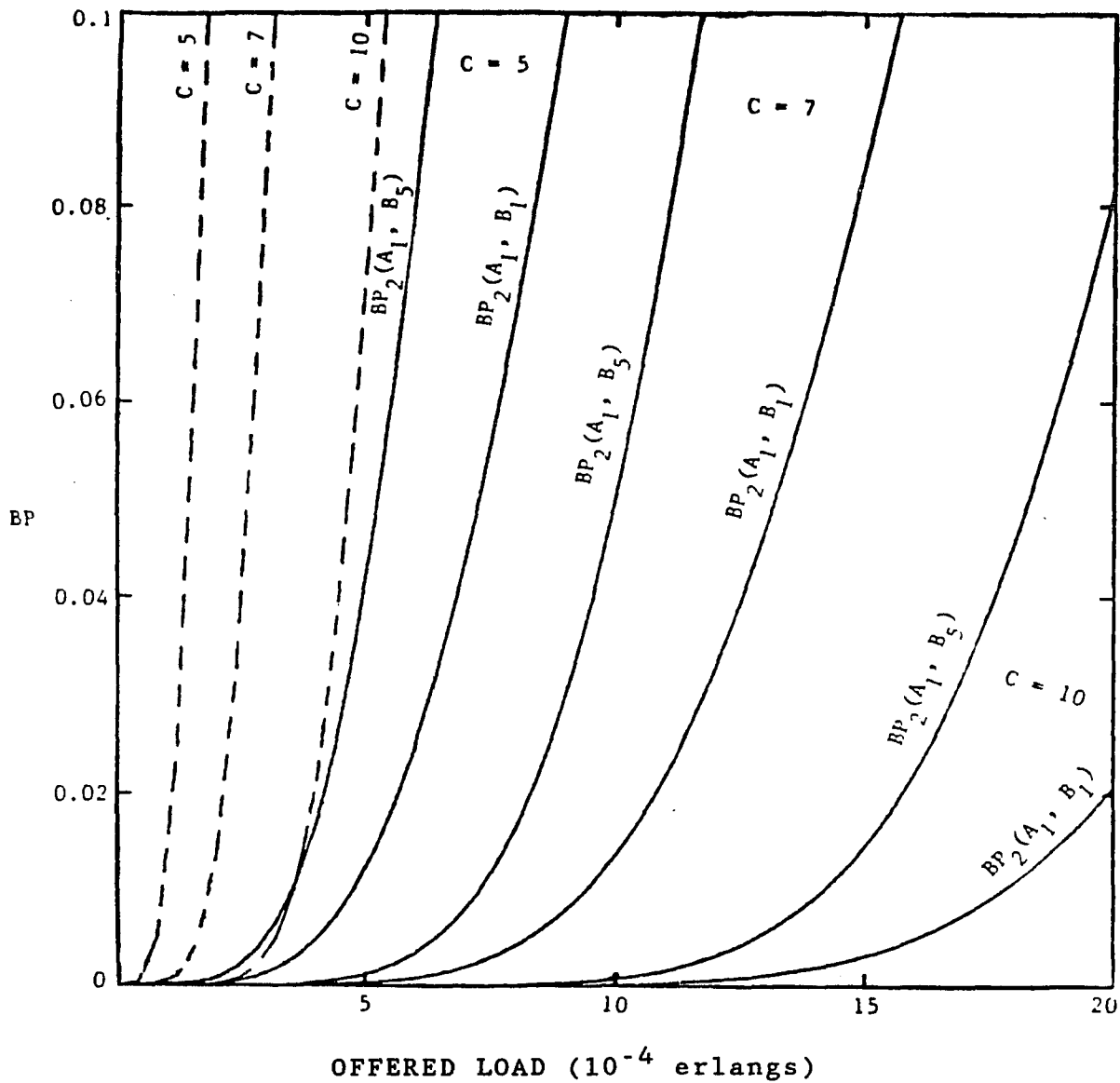


Fig. 3. Blocking probability vs. offered load for $n=2000$ and $m=5$, (total number of active customers = 10,000).

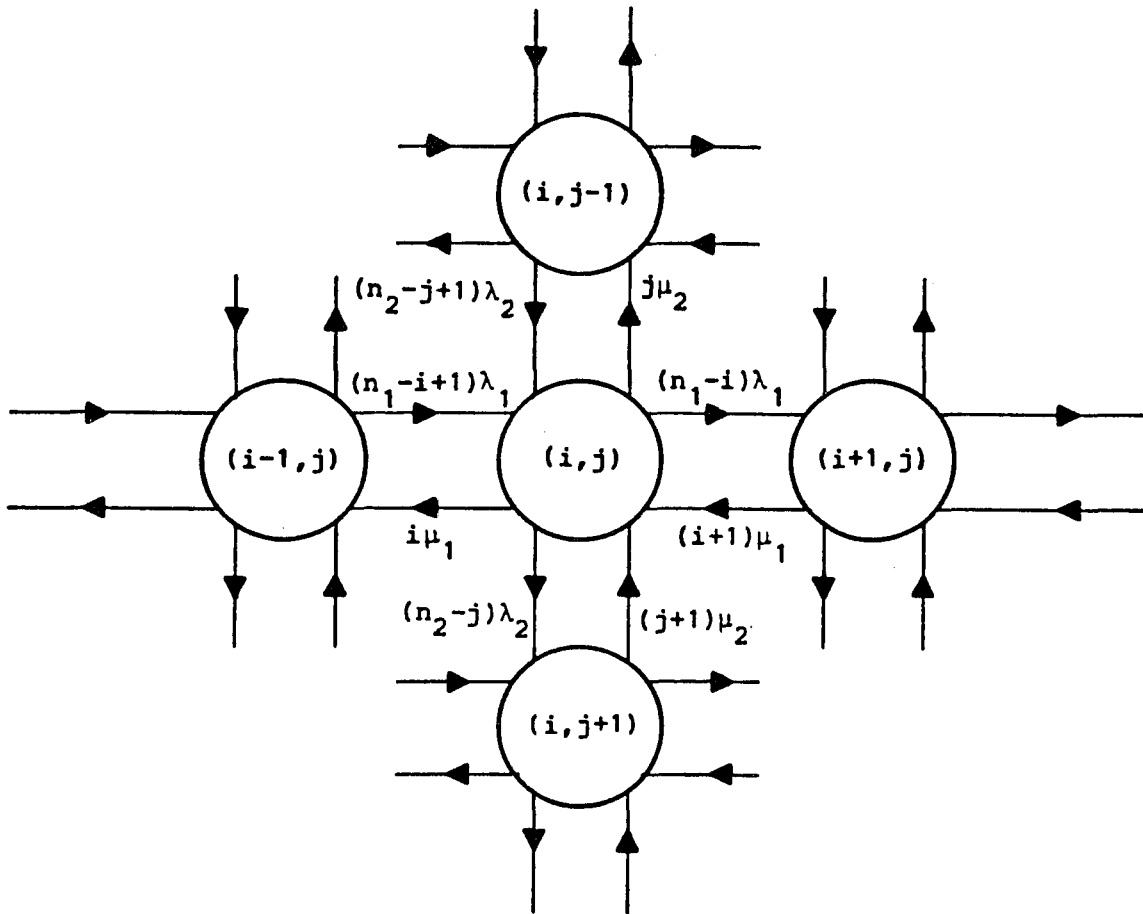


Figure 4. State diagram for two groups of users (case 2).

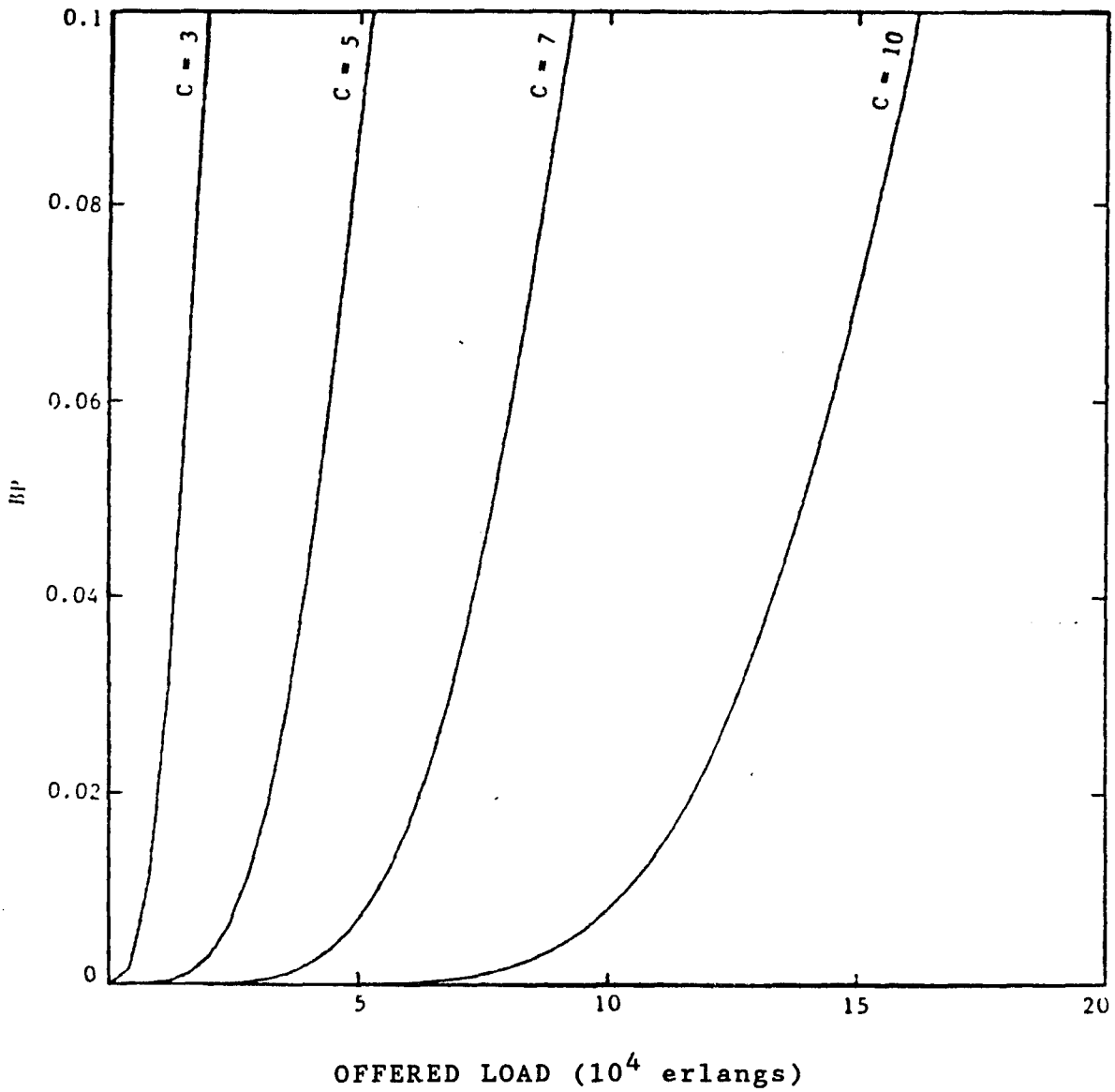


Figure 5. Overall blocking probability vs. a_1 for $m = 5$, $n_1 = 1580$, and $n_2 = 20$. ($a_2 = 100a_1$). Total number of active customers = 8000.

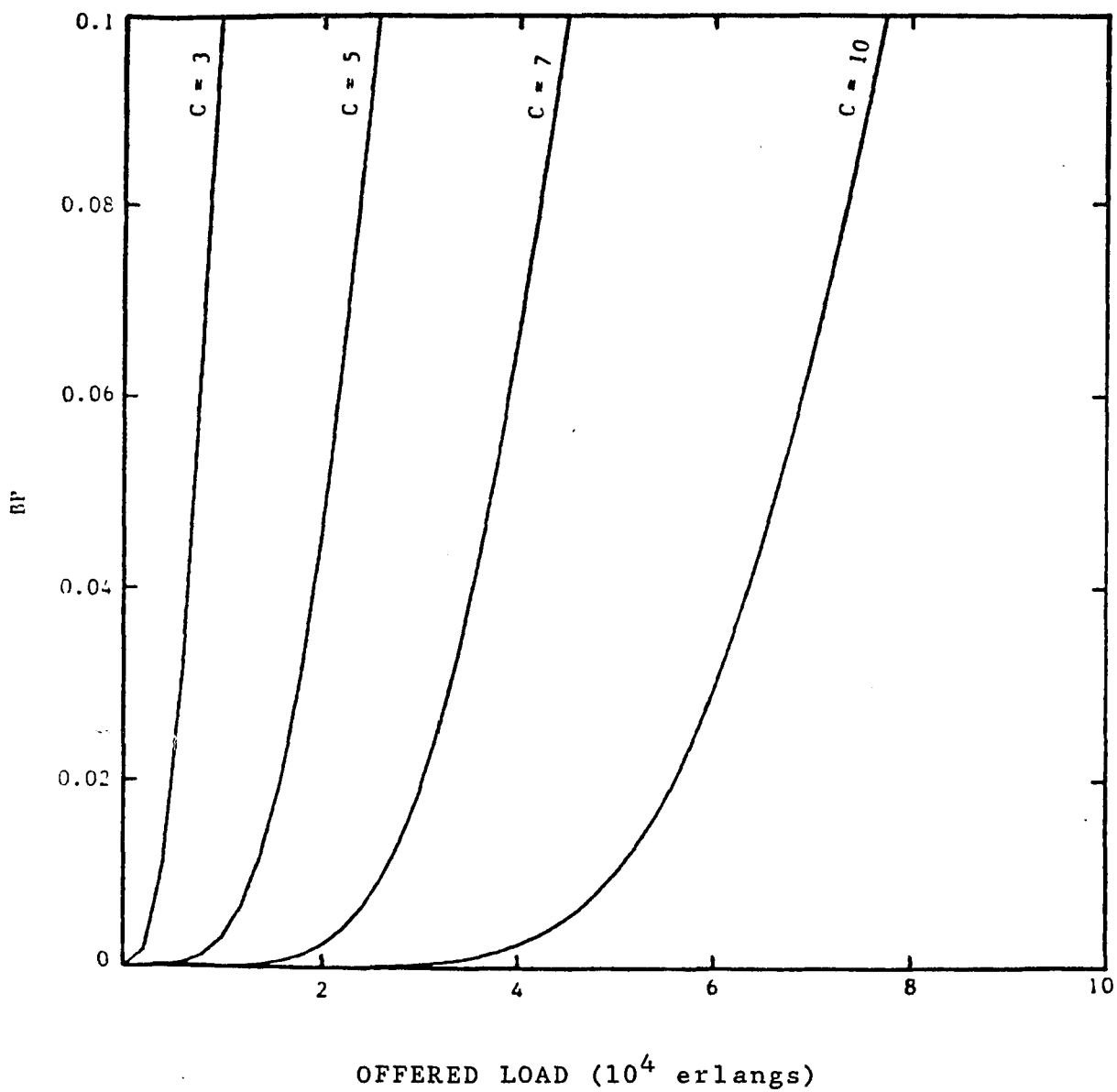


Figure 6. Overall blocking probability vs. \bar{a}_1 for $m = 5$, $n_1 = 3160$, and $n_2 = 40$. ($a_2 = 100a_1$). Total number of active customers = 16000.

Table I: Mean Number of Video Frames per Second, λ_1 , for a Given Blocking Probability, BP, and Number of Channels per Trunk, c.

| BP \ c | 0.01 | 0.03 | 0.05 |
|--------|-----------------------|-----------------------|-----------------------|
| 3 | 2.30×10^{-3} | 3.53×10^{-3} | 4.3×10^{-3} |
| 5 | 8.13×10^{-3} | 10.9×10^{-3} | 12.6×10^{-3} |
| 7 | 1.62×10^{-2} | 1.87×10^{-2} | 2.31×10^{-2} |
| 10 | 3.12×10^{-2} | 3.51×10^{-2} | 4.18×10^{-2} |

$m = 5$, $n_1 = 1580$, $n_2 = 20$

Total number of active users = $m(n_1 + n_2) = 8000$

Table II: Mean Number of Video Frames per Second, λ_1 , for a Given Blocking Probability, BP, and Number of Channels per Trunk, c.

| BP \ c | 0.01 | 0.03 | 0.05 |
|--------|------------------------|------------------------|------------------------|
| 3 | 1.1×10^{-3} | 1.75×10^{-3} | 2.1×10^{-3} |
| 5 | 3.97×10^{-3} | 5.34×10^{-3} | 6.18×10^{-3} |
| 7 | 7.88×10^{-3} | 9.69×10^{-3} | 11.2×10^{-3} |
| 10 | 14.93×10^{-3} | 18.75×10^{-3} | 19.95×10^{-3} |

$m = 5$, $n_1 = 3160$, $n_2 = 40$

Total number of active users = 16000

CHAPTER III

CIRCUIT-SWITCHED ROUTING IN ISDN ENVIRONMENTS

1. INTRODUCTION

In the previous chapter, the modeling of circuit switching networks with more than one group of users was explored. Based on a queueing system with c -servers (c -channels) and n -different types of customers, we developed the model required to predict the link blocking probability. In particular, we classified the users into two groups of users, home-users and business-users, respectively. Using this classification, we applied the model to a CATV network with a tree-like structure and determined the circuit switching capacity in terms of the end-to-end blocking probability and the overall average system blocking rate.

As mentioned in the introduction, one of the most important issues associated with circuit switching networks is the routing problem. The routing algorithm for the

network specified with a tree-like topology described in Chapter II is a trivial problem since for that type of network there is only one possible path between each pair of users. The routing problem becomes an essential issue in a distributed network in which a user may have more than one choice in selecting a path for transmitting information to another user.

The routing problem in a multi-purpose circuit-switched network environment, e.g., ISDN, has not been discussed before. With the emergence of new communication networks, it is necessary to study and gain insights into the problems associated with the performance evaluation of multi-media, circuit switching networks. Focusing on the routing issues in distributed networks with two groups of users and employing the circuit switching technique, we would now like to examine the impact of various routing algorithms on the network performances.

The following general classification for the routing algorithms has been considered by [15]:

- 1) From the network status information standpoint, i.e., traffic load and network topology, routing algorithms can be time invariant (deterministic) or can vary with time (adaptive).

- 2) Based on the availability of the information at different nodes, routing algorithms are performed globally or locally.

3) Routing algorithms can be centralized in a single network component or distributed to each node in the network.

4) On the basis of the optimality criteria, a routing algorithm may be classified according to its tendency towards the optimization of one of the fundamental performance functions. More routing policy classification can be established on the basis of reliability, simplicity, and the size of the network.

The performance of a routing algorithm in a circuit-switched network, as mentioned in the introduction, is usually measured in terms of the end-to-end blocking probability experienced by each user, the total cost, and the overall average network blocking rate. For example, one routing algorithm can be formulated to minimize the individual source-destination blocking rate while another may be responsible for a maximum attainable throughput. A practical routing strategy is defined as one which produces a maximum throughput under the constraint that the individual source-destination blocking rates are maintained below an acceptable level predetermined by the network designer. The reason for imposing this constraint is to assure that maximization of the throughput or, equivalently, the minimization of the network blocking rate, is not done at the expense of an intolerable blocking probability

between two nodes.

On the basis of network status, the routing algorithms are classified as deterministic (fixed) if they are time invariant and do not vary with load fluctuations and topology changes. They are categorized as adaptive (dynamic) if the routing rules are assigned according to the network's current state. Adaptive routing algorithms have received more attention since they can adjust to load fluctuations and network failures.

Three types of dynamic changes in a network are identified [16]: the slow changes which include modification in network topology (link and/or node additions or outages); the medium rate changes due to shifts in the average offered traffic rates; and the rapid changes due to the stochastic fluctuations of call interarrival time and holding time. Using the network's present state, an adaptive (dynamic) routing policy accommodates some or all of the above changes.

Both deterministic and adaptive routing policies are being used in circuit switching networks. A deterministic routing policy, which is usually easier to analyze than an adaptive routing policy, is generally used in the network design process, whereas an adaptive routing policy is implemented during the network operating phase.

Various routing techniques for circuit-switched networks in an environment consisting of only one type of

user have been studied for many years. Examples appear in [3, 8, 17, 18]. Generally speaking, all the current routing techniques in circuit switching networks have at least a distinguishing feature in common, i.e., all the circuit-switched routing techniques can be considered the subclasses of a broad alternate routing class. Depending on the alternate routing plan employed, a call rejected by the first path is allowed to attempt different paths one at a time until either it seizes a free path or is turned down by all permissible paths. A call rejected by all allowable paths is deemed lost and does not retry to establish a path in the near future; as a result it will have no effect on the performance thereafter (loss system).

The purpose of using alternate routing techniques in circuit switched networks is clearly to optimize network performances. To do this, calls are permitted to attempt alternate paths in a sequential manner. The sequential order is given by the routing policy used. The major drawback of using alternate routing techniques is an inevitable computation task. This task is required to update and transfer the routing table from one node to another.

Due to the limitations of network component capabilities such as speed, amount of storage memory, etc., and the extensive computation requirements for determining the best possible alternate routing algorithm, and also

because of other problems involved in the control of traffic flow such as monitoring the traffic and avoiding the looping situation in which a call may visit an intermediate node along its path more than once, the alternate routing strategies used in most of the present circuit-switched networks follow some form of hierarchical routing structure.

The use of hierarchical routing in circuit-switched networks is mainly intended to prevent the network from falling down into situations such as bouncing or looping. In addition, this type of routing will significantly reduce the time required for the calculation of end-to-end blocking probabilities. Nevertheless, the reason for the widespread existence of hierarchical routings has basically been due to switching limitations and computational difficulties. Because of advances in communication technology, such as computer-controlled electronic switching and high capacity interoffice signaling, these problems no longer exist. Therefore, we can extend the current routing rules far beyond the conventional hierarchy. In comparison with nonhierarchical routing networks, routes in hierarchical networks are relatively fixed. This implies that the hierarchical routing scheme disregards the possibility of making use of free alternate paths. This may result in a higher blocking rate compared with that of one without a hierarchical structure. Nonhierarchical routing in ISDN environment can take advantage of traffic fluctuations which

arise from the two classes of users.

In this research, we focus on the routing problem for circuit-switched traffic for nonhierarchical ISDN networks. Following this introduction, a general alternate routing network has been defined in Section 2. Applying the route formulation technique, a mathematical model is developed in Section 3. Then, alternate routing and nonalternate routing are discussed for fully connected networks in Sections 4 and 5, respectively. Section 6 focuses on a specific alternate routing technique called "two-link alternate routing". Numerical analysis is presented in Section 7.

2. GENERAL ALTERNATE ROUTING NETWORK MODEL

The network to be considered here consists of N nodes and L links. Attached to each link there are two classes of customers, class 1 and class 2, respectively. An example of such a network is shown in Fig.1.

In circuit-switched networks, a call is considered a basic unit of traffic. A call is originated at a node (source) and is destined for another node (destination). A path is defined as a set of distinct links used to connect a source node to a destination node. In an alternate routing network, there's more than one path for transmitting a call from a source node to its destination node. Associated with each source-destination node pair, a route is usually

defined as a collection of all the paths that exist between the two nodes.

A call originating at a source node will try to seize a path to establish a circuit to its proper destination node. The choice of selecting a path is entirely controlled by the routing strategy employed. A routing table is assigned to each node. This routing table contains all the information concerning the paths to be used and the order of preference.

Referring to Fig.1 as an example of an alternate routing network, consider the routing of calls from B to D. The complete process of routing the calls between a source-destination node pair can clearly be seen if it is illustrated by a route tree [11]. Route trees provide the same information as given by the routing table. For example, the route tree in Fig.2 shows that there are three paths between B and D. In order of preference, they are: (BD), (BCD), and (BAED). The first path, BD, is called the direct trunk group (in telephone networks) or the primary path. The two remaining paths, (BCD) and (BAED), are called alternate trunk groups or alternate paths.

Suppose that a customer attached to node B wishes to communicate with another customer located at node D. The communication between these two customers will take place if the route (B-D) has at least one free path. In order to transmit the call, an attempt is first made to use the

primary path labeled (BD). If the primary path is not available, the call is then offered to the alternate paths in the order of preference indicated above. The call between B to D will be blocked if each path in the route has at least one busy link.

The basic idea of using alternate routing, as mentioned in the introduction, is to improve the network performance by lowering the end-to-end blocking probabilities. To calculate the end-to-end blocking probabilities, we need to know the number of customers on each link. This means that we must develop an accurate mathematical model to compute the number of customers that are offered to each link.

Applying the route formulation technique, similar to the one used by both [6] and [19], a mathematical model is developed. The models given by both [6] and [19] were used to calculate the carried load for each link on the basis of the Erlang B formula. Considering the case of two classes of customers presented in Chapter II, we are interested in a model which will obtain the number of users offered to each link.

Following the presentation of the mathematical model, the two performance measures are formulated in their most general forms. In order to gain some insight into the network performance measures, the model is then applied to an N-node, fully connected and symmetric nonhierarchical network. Using this type of network topology, alternate and

non-alternate routing procedures are discussed in Sections 4 and 5.

3. MATHEMATICAL ANALYSIS

Using the route formulation technique, a mathematical model for the general alternate routing network is presented in this section. The link blocking probability formula for the case of two classes of users was carried out in Chapter II and is given by

$$P_1 = P(0,0) \sum_{i+j=c} a_1^i a_2^j \binom{n_1}{i} \binom{n_2}{j} \quad (1)$$

where n_k^1 is the number of type k ($k=1,2$) customers supported by link 1.

Let us first introduce the following notation:

- N - number of nodes in the network
- L - number of links in the network
- c_1 - number of channels for link 1
- SD - number of source-destination node pairs, $SD = N*(N-1)$
- N_k^j - number of class k customers for the source-destination pair j , ($k=1, 2$)
- a_k - load offered by any type k customer in Erlang
- r_m^j - the (m) th path for source- destination pair j
- R^j - the total number of paths for source-destination pair j

R_g^j = the first g paths of R^j

$n_k^j(m)$ = number of class k customers served by the (m) th path
for pair j

P_l = blocking on link l

n_k^l = the total number of class k customers offered to link l

nn_k^l = the total number of class k customers served by link l

$P(r_m^j)$ = blocking of path m for source-destination pair j

$P(R_g^j)$ = probability that the first g paths of source-
destination pair j are blocked.

In addition to the above notation, the following assumptions
have been made.

- 1) The link blocking probabilities are statistically independent from each other.
- 2) The call interarrival times form a Poisson arrival process.
- 3) No priority is given to the users with large traffic requirements.
- 4) The time required to establish a path is negligible.
- 5) Paths are disjointed.

As mentioned, the idea is to determine the total number of class k ($k=1, 2$) customers n_k^l , that is offered to link l . n_1^l and n_2^l are two variables needed to compute the blocking on link l . To find n_k^l , first we have to calculate the number of users contributed to each link by a source-destination node pair. Using the notation described above, $n_k^j(m)$ is

first calculated as follows:

Let $M_k^j(m)$ be the number of class k customers that is offered to the m^{th} path for source-destination j . $M_k^j(m)$ is given by

$$M_k^j(m) = N_k^j \times \text{prob. [the first } (m-1) \text{ paths of } R^j \text{ are blocked]}$$

Let the probability that the first $(m-1)$ paths of R^j are blocked be $P(R_{m-1}^j)$. Then, we have

$$M_k^j(m) = N_k^j P(R_{m-1}^j) \quad (2)$$

Using (2), we can now obtain $n_k^j(m)$, the number of class k customers served by the m^{th} path for the source-destination pair j . $n_k^j(m)$ is simply given by the difference in number of class k customers that are offered to the m^{th} and $(m+1)^{\text{st}}$ paths, respectively. Using (2), we get

$$n_k^j(m) = N_k^j [P(R_{m-1}^j) - P(R_m^j)] \quad (3)$$

Now, the total number of class k customers served by the (l) th link is given by

$$nn_k^l = \sum_{j=1}^{N(N-1)} \sum_{\substack{m=1 \\ l \in r_m^j}}^{R^j} n_k^j(m) \quad (4)$$

Finally, the total number of type k customers offered to link 1, n_k^1 , will be

$$n_k^1 = \frac{nn_k^1}{1 - P_1} \quad \begin{cases} l = 1, \dots, L \\ k = 1, 2 \end{cases} \quad (5)$$

Using (1) to (5), the blocking probability of link 1 is determined.

Having found the link blocking probabilities, we are now ready to calculate the path and route blocking probabilities. Under the assumption of independent link blocking probabilities, a path is said to be blocked if at least one of the links in the path is not available. The blocking rate on the m^{th} path of source-destination j is then given by

$$\begin{aligned} P(r_m^j) &= 1 - \text{prob.}[\text{all the links in the path are available}] \\ &= 1 - \prod_{l \in r_m^j} (1 - P_l) \end{aligned} \quad (6)$$

Using (6), the route blocking probability for source-destination j is now determined as follows: The route between source-destination j is blocked if all the paths in the route are busy. This blocking is given by

$$P(R^j) = \prod_{m=1}^{R^j} P(r_m^j) \quad (7)$$

Note that the route blocking probability defined here is the same as the end-to-end blocking rate.

The set of non-linear equations formed by (1) to (7) will completely characterize the steady-state behaviour of the network under general routing strategy. For a given network and traffic requirements, these equations must be solved simultaneously; however, since it is not possible to obtain closed form solutions, an iterative procedure must be used. Starting, for example, with an estimate of the link blockings and using equations (3) through (5), the number of type 1 and type 2 customers for each link is computed. Then, the link blockings are calculated by (1) and the results are compared to their initial values. The iteration is carried out until the differences between all corresponding link blockings in two successive steps are zero or smaller than some predetermined value.

3.1 The End-To-End Blocking Probability

The end-to-end blocking probability for any source-destination node pair can be determined after we have solved the set of seven equations, (1)-(7) and found all the link blockings. As we noted earlier, the end-to-end blocking probability for source-destination pair j , $BP(j)$, is given by the related route blocking. Substituting (6) into (7), $BP_a(j)$, the end-to-end blocking probability for source-

destination node pair j is obtained by

$$BP_a(j) = \prod_{m=1}^{R^j} [1 - \prod_{l \in r_m^j} (1 - P_l)] \quad (8)$$

3.2 The Average Network Blocking Rate

The average network blocking rate is defined as the average of all the end-to-end blocking probabilities experienced by all the customers and is given by

$$BP = \frac{\sum_{j=1}^{N(N-1)} (a_1 N_1^j + a_2 N_2^j) BP_a(j)}{\sum_{j=1}^{N(N-1)} (a_1 N_1^j + a_2 N_2^j)} \quad (9)$$

where $N(N-1)$ denotes the total number of source-destination node pairs in the network.

4. ALTERNATE ROUTING IN FULLY CONNECTED NETWORKS

In the preceding Section, we presented a general mathematical model for calculating the end-to-end blocking probabilities. The model was obtained for a network having an arbitrary topology. There is, however, a special class of

network topology called complete or fully connected topology which makes the end-to-end blocking probability easier to calculate. This type of network topology has shown to be a good model for studying some real networks. For example, see [7] and [8].

In a fully connected network, each node is connected to another node via a direct link. Fig.3 shows such a network with five nodes. Since each node is directly connected to other nodes, an N-node fully connected network has duplex links. Furthermore, the direct link between two nodes is referred to as the direct path (the primary path) and all the other remaining paths between the same two nodes are called alternate paths.

In this section, an N-node, fully connected and symmetrical nonhierarchical network model is assumed. Applying the route formulation algorithm given in the previous section to this network, two different routing procedures will be discussed. They are:

- 1) Non-Alternate Routing
- 2) Two-Link Alternate Routing

In addition to the assumptions described in Section 3, the following assumptions are used in order to simplify our analysis.

- 1) All the source-destination node pairs have an equal number of customers.

- 2) The number of channels is the same for all the links.
- 3) The traffic distribution is uniform.

Under these assumptions, the end-to-end blocking probability for any source-destination pair is the same as the network blocking rate.

5. NON-ALTERNATE ROUTING IN FULLY CONNECTED NETWORKS

Non-alternate routing is the simplest form of routing. In a network employing non-alternate routing, no alternate path has been assigned to any node. A call can only be transmitted to its destination through the direct (primary) path. If the direct path is busy, the call will be cleared and is lost.

Simplicity and stability are the most important properties of this routing technique. It is considered stable, compared to alternate routing, in the sense that the blocking probability changes smoothly with the traffic rate. This subject will be discussed later in the numerical example.

5.1 The End-to-End Blocking Probability

The end-to-end blocking probability for this case is the same as the link blocking probability since no alternate

path has been allowed. This end-to-end blocking probability is shown here by BP_n . Let N_1 and N_2 be the number of type 1 and type 2 customers for each source-destination node pair. The end-to-end blocking probability with c channel per link is then given by (see (1))

$$BP_n = P(0,0) \sum_{i+j=c} a_1^i a_2^j \binom{N_1}{i} \binom{N_2}{j} \quad (10)$$

where a_k is the offered load by any type k ($k = 1, 2$) customer in erlangs.

6. TWO-LINK ALTERNATE ROUTING

Under this scheme, a call blocked by the direct path is allowed to attempt up to s two-link alternate paths. For an N -node fully connected network, s may vary between 1 and $(N-2)$. Depending on the value of N , other alternate paths consisting of more than two links may exist in the network. For example, in a four-node fully connected network, there are two two-link paths and also two three-link paths between any source-destination node pair; however, only the two two-link paths are disjointed.

According to the number of links in each path, alternate paths can be divided into different groups, i.e., all the paths in a group have the same number of links.

Denoting the number of paths in a group by G_p , it can be shown that G_p is given by

$$G_p = \prod_{i=2}^p (N - i) \quad 2 \leq p \leq N \quad (11)$$

where p is the number of links in a path.

The reasons for choosing the two-link alternate routing are:

- 1) The actual blocking performance is mathematically tractable and, thus, its calculation is much easier.
- 2) To be assured that the paths are disjointed.
- 3) To achieve better network utilization by using less network resources.

As mentioned earlier, the end-to-end blocking probability for this case is the same as the route blocking. Furthermore, the average network blocking rate is also equal to the end-to-end blocking probability since both classes of customers are similarly being treated.

As we will see later in examples, the major drawback of this type of routing is the stability problem which arises under heavy traffic conditions. One approach to overcoming this stability problem while still providing relatively better services to customers with large traffic requirements, is to ban the first class of users from attempting alternate paths. Preliminary results have shown

that by superimposing this type of control policy on the top of the alternate routing rule, the stability problem can be improved. This subject will be discussed in the next chapter. Using the alternate routing model discussed in Section 3, the end-to-end blocking probability in the case of two-link alternate routing is now derived.

6.1 End-To-End Blocking Probability

Recall that n_1 and n_2 represent the number of class 1 and class 2 customers attached to each node, respectively. We have assumed that all the source-destination node pairs have the same number of customers and the traffic distribution is uniform. Therefore, the number of class k customers for any source-destination pair is given by

$$N_k^j = \frac{n_k}{N - 1} \quad \left\{ \begin{array}{l} k = 1, 2 \\ 1 \leq j \leq N(N-1) \end{array} \right. \quad (12)$$

where N is the number of nodes in the network.

Assuming a symmetrical network, as above, one then concludes that the link blockings are the same. This probability is simply referred to by P here. The first m paths of each source-destination pair are blocked if the direct path and $m-1$ alternate paths are blocked. Using (6) and noting that an alternate path consists of two distinct links, we get

$$P(R_m^j) = P[1 - (1 - P)^2]^{m-1} \quad 1 \leq m \leq s+1 \quad (13)$$

Substituting (12) and (13) into (3), the total number of class k customers served by the (m)th path for a source-destination j is then obtained as

$$n_k^{j(m)} = \frac{n_k}{N-1} P (1 - P)^2 (2P - P^2)^{m-2} \quad 2 \leq m \leq s+1 \quad (14)$$

The number of class k customers served by the first path (direct path) has to be calculated separately. Let this number be nn_f . nn_f is given by

$$nn_f = \frac{n_k}{N-1} (1 - P) \quad (15)$$

where $(1-P)$ is the availability of the direct path. Substituting (14) and (15) into (4), nn_k , the number of class k customers served by a link is given by

$$nn_k = \frac{n_k}{N-1} (1 - P) \left(1 + P(1 - P) \sum_{j=1}^{N(N-1)} \sum_{\substack{m=2 \\ l \in r_m^j}}^{s+1} (2P - P^2)^{m-2} \right) \quad (16)$$

Considering all the properties of the network employed here, the condition that $l \in r_m^j$ can easily be removed. This is done

as follows. Every call is allowed to attempt up to s alternate paths and since each alternate path consists of two links, there are $2 \times s$ possible alternate links. Thus, the first summation can be replaced by $2 \times s$. Furthermore, due to the fact that most likely every link is equally being used as a part of an alternate path, we can now remove the condition that $l \in r_m^j$ by normalizing the second summation over s . The simplified version of equation (16) is now given by

$$n_{nk} = \frac{n_k}{N - 1} (1 - P) \left(1 + 2P \frac{1 - (2P - P^2)^s}{1 - P} \right) \quad (17)$$

Finally, the total number of class k customers offered to every link is obtained by substituting (17) into (5).

$$n_k^l = \frac{n_k}{N - 1} \left(1 + 2P \frac{1 - (2P - P^2)^s}{1 - P} \right) \quad (18)$$

The end-to-end blocking probability corresponds to the case that all the paths are busy. Using the equation (8), the end-to-end blocking for the two-link alternate routing BP_a is then given by

$$BP_a = P (2P - P^2)^s \quad (19)$$

with

$$P = \frac{\sum_{i=0}^c a_1^i a_2^{c-i} \binom{n_1^1}{i} \binom{n_2^1}{c-i}}{\sum_{i=0}^c a_1^i \binom{n_1^1}{i} \sum_{j=0}^{c-i} a_2^j \binom{n_2^1}{j}} \quad (20)$$

where n_k^1 is given by (18).

7. Numerical Analysis

In this example, we have considered the five-node fully connected and symmetric network model shown in Fig.3. In this model the users have been grouped into two classes, home users (class 1) and business users (class 2). Using the results obtained in preceding sections, we would like to compare the end-to-end blocking performance of alternate routing and nonalternate routing. The following numerical values are assumed.

$n_1 = 600$ (class 1 customers attached to each node)

$n_2 = 20$ (class 2 customers attached to each node)

$c = 5$ (number of channels per link)

Furthermore, we assume that the users belonging to class 2 are much more active than the class 1 users with the ratio of the individual offered load of the second group to the first group $a_2/a_1 = 50$. Using the equations derived in this chapter and the preceding chapter, the results are obtained

and plotted in Fig.4. The three curves labeled $s = 0$, $s = 1$, and $s = 3$ correspond to the end-to-end blocking probability of non-alternate routing, two-link alternate routing with one alternate path, and two-link alternate routing with three alternate paths (full alternate routing).

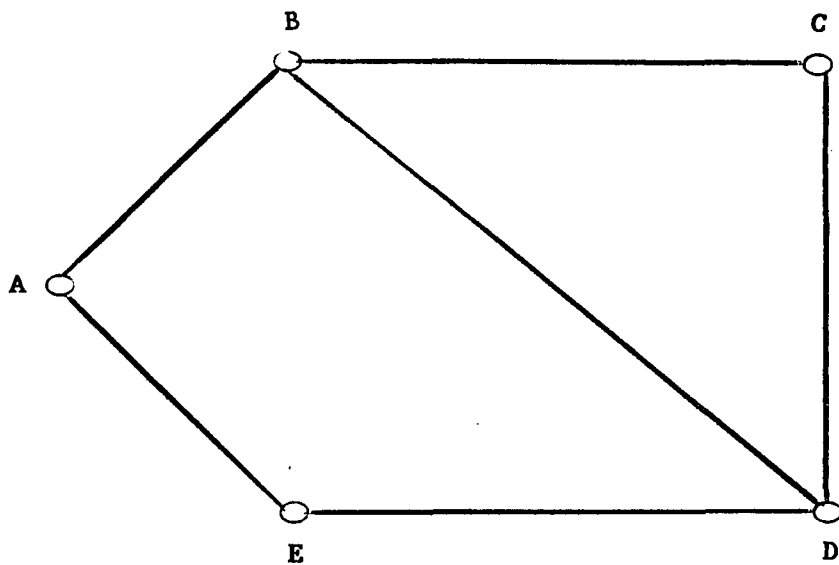
It is shown that with a_1 less than 0.0095 erlangs and $s = 3$, alternate routing improves the blocking performance. Let us assume that the blocking at normal traffic loads and without alternate routing is 0.04. Fig. 4 shows that at this blocking rate, $a_1 = 0.0065$ erlangs. For $a_1 = 0.0065$, we observe that alternate routing with $s = 3$ gives a blocking rate less than 0.01. Now suppose a_1 increases by 46%. This can happen under heavy load conditions. At this load, $a_1 = 0.0095$ erlangs, both non-alternate routing and alternate routing give almost the same blocking. However, if a_1 increases to 0.01, full alternate routing ($s = 3$) shows a performance degradation. This is understandable since under heavy traffic conditions direct paths become less available and more calls attempt alternate paths. We observe that the end-to-end blocking probability of alternate routing is more sensitive to the increase of a_1 than that of non-alternate routing. This sensitivity causes the stability problem which has been mentioned earlier. Comparing the curves labeled $s=3$ and $s=1$, we see that the end-to-end blocking probability of alternate routing with one alternate path is less sensitive

than that of full alternate routing. From these observation we can conclude that alternate routing can improve the blocking performance when the traffic load is not very heavy.

8. CONCLUSION

In this chapter, we have studied the circuit-switched routing in the ISDN environment. We have considered an environment consisting of two user classes: a home user class (type 1) and a business user class (type 2). Using this user classification model, alternate routing and non-alternate routing are discussed. In a comparison of different routing procedures, the major emphasis is placed upon the end-to-end blocking probability and average network blocking rate. The model is then applied to fully connected networks and the two routing techniques are compared. The results show that alternate routing performs better than non-alternate routing under light to moderate traffic conditions. The results also indicate that the alternate routing strategy exhibits a performance degradation, as compared to non-alternate routing, when the traffic load exceeds a specified operating point. This performance degradation becomes more severe as the traffic load increases. To avoid this performance degradation caused by alternate routing during overload conditions, some form of

control mechanism must be used together with alternate routing policy. This subject will be discussed in the next chapter.



○ Node
— Link

Fig.1 A distributed network with 5 nodes and 6 duplex links.

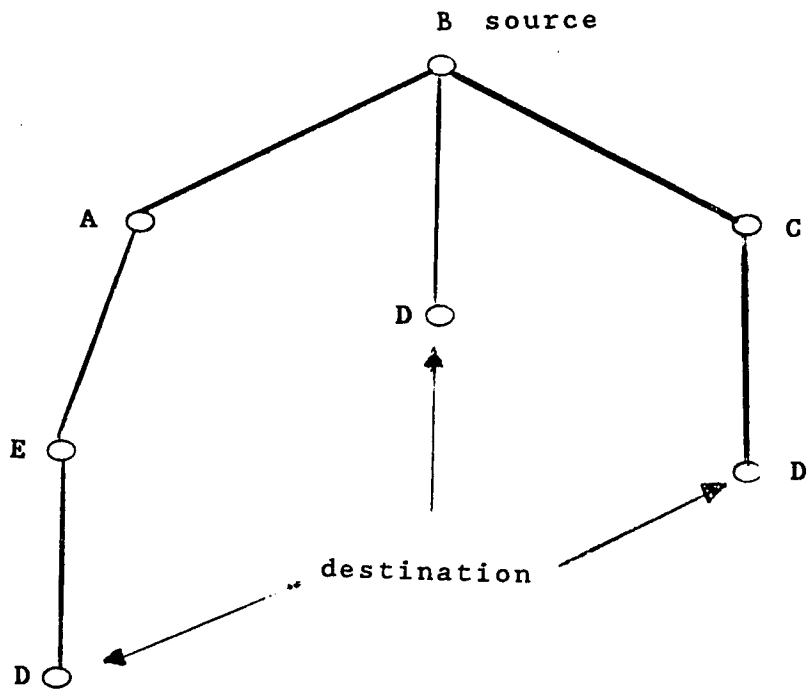


Fig. 2. Route tree for source-destination node pair BD.

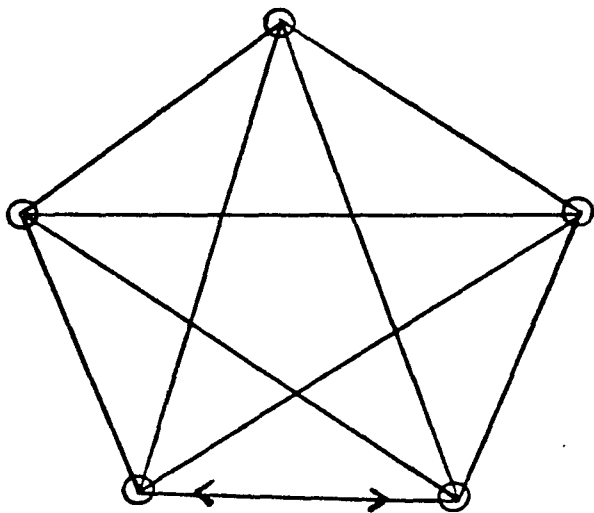


Fig. 3. A 5-node fully connected network

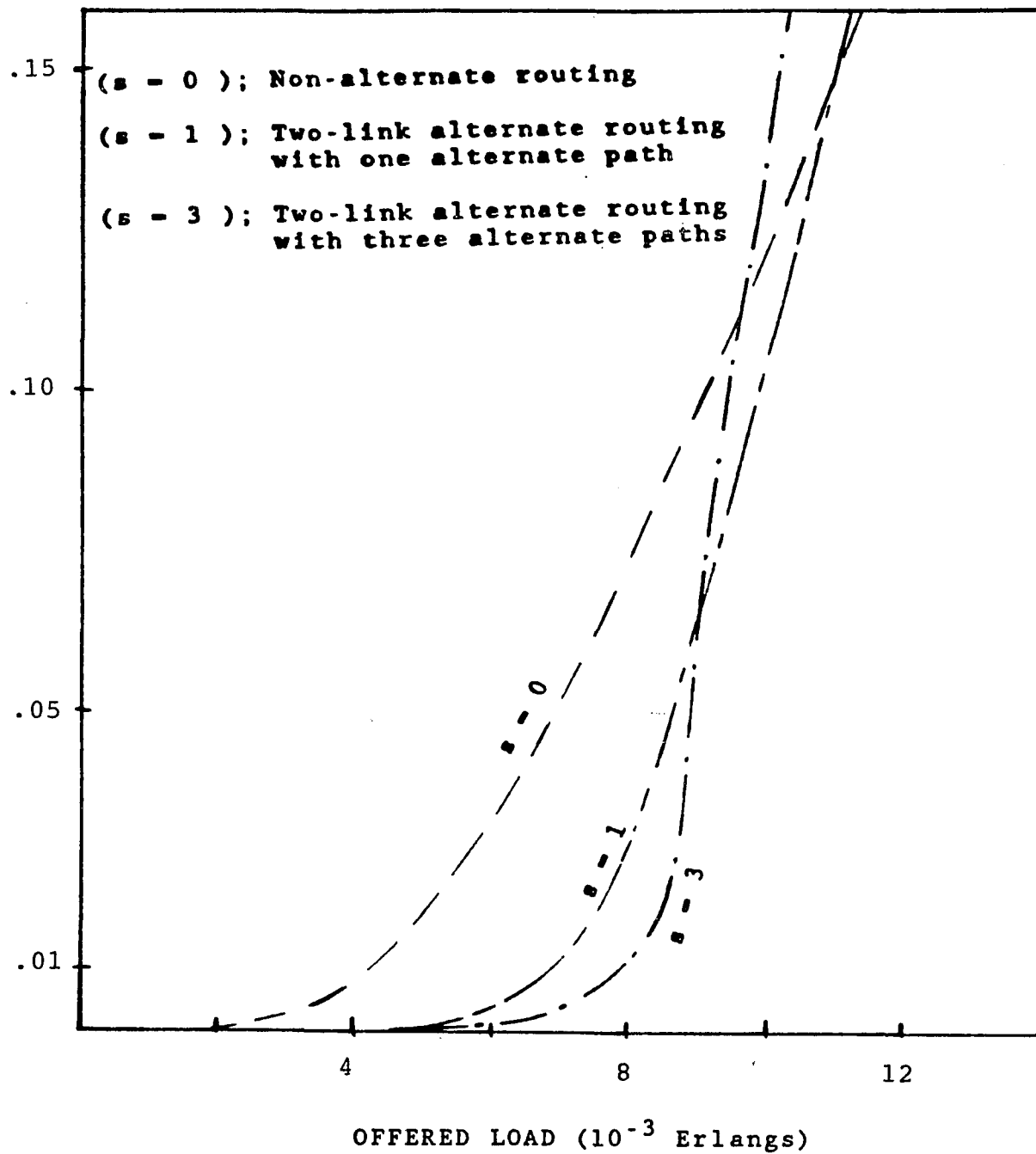


Fig. 4. End-to-end blocking probability vs. a_1
 $(a_2 = 50a_1)$

CHAPTER IV

ALTERNATE ROUTING WITH TRUNK RESERVATION IN ISDN

INTRODUCTION

Alternate routing techniques in circuit-switched networks are used in order to improve network performance as well as to protect the network against components failures. To achieve these goals, a call blocked by its first path is given a degree of freedom as to use in attempting different paths hoping that it will find a free one. However, alternate routing, like any other routing technique, is greatly vulnerable to high traffic demands if it is left uncontrolled. That is, if unexpected overload conditions are allowed to persist, the network performance deteriorates and may even cause congestion. This was indeed the case in the ISDN model studied in chapter III. It was shown that once the offered load exceeds a certain level, the network becomes rapidly congested and its performance falls below that of non-alternate routing. As the name "uncontrolled

alternate routing" suggests, in an uncontrolled alternate routing network, the overflow traffic is unconditionally allowed to try alternate paths even if the offered traffic exceeds the design limit.

At first glance, it may seem that alternate routing policy is helpful in improving network performance since it provides more than one possible path between any source-destination node pair. But, as is well known, this is not always true; it depends also on the traffic intensity and the availability of resources. The results in the preceding chapter indicated that the unconstrained alternate routing policy is superior to non-alternate routing only at normal to moderate traffic loads. It was also shown that under heavy traffic conditions, alternate routing without flow control is subject to performance degradation and the network performance is greatly sensitive to the offered traffic deviation.

The key to this behaviour lies in the fact that uncontrolled alternate routing causes inefficient utilization of the network resources. The performance degradation in pure alternate routing networks occurs because alternate paths are usually longer than the primary paths. This means that a call carried via an alternate path requires more network resource than when it is sent to its destination via the primary path. When the input traffic exceeds a certain

level, more calls are blocked by the direct paths and as a result they retry alternate paths. As the traffic load increases, the links become less and less available. Even if the input traffic continues to increase linearly, the net traffic load that must be supported by each link increases more rapidly and links will be utilized more. The actual load offered to each link is the sum of direct calls and alternate calls. At normal loads, links are more available and consequently fewer calls are refused by direct paths. Under this condition, the overflow traffic would have a better chance to use a secondary path to establish an end-to-end connection. On the other hand, at high traffic loads, links are less idle. This being the case, more calls are alternately routed and therefore links are used to support more calls. So, the total load on each link is considerably higher than if alternate routing were not allowed and, consequently, link blocking probabilities will be much higher. Under this condition, even though calls have more choices in attempting different paths, this advantage is offset by the fact that the links are highly utilized.

A simple analytical model may be helpful in seeing how an uncontrolled alternate routing behaves as the traffic load changes. To do this, let us consider a 3-node complete interconnected network. For the sake of simplicity, furthermore, assume that the traffic distribution throughout the network is uniform and all links are duplex with an

equal number of channels. For this model, since each source-destination pair has one direct path and one alternate path, the end-to-end blocking probability is given by $P_a = P(2P - p^2)$ where P is the link blocking rate. Let P_n be the end-to-end blocking probability for the same network but without alternate routing. Note that in fully connected networks, P_n is the same as the link blocking probability. Comparing P_n with P_a , alternate routing performs better as long as P_a is less than P_n . If P_n , on the other hand, is less than P_a , the reverse is true, that is, the alternate routing becomes worse as compared to non-alternate routing.

Let Δ be the difference between P_n and P_a . Figure 1 shows Δ as a function of P_n for two different cases. The two cases differ from each other in the number of channels (circuits) that are assigned to the links ($c = 5$, $c = 10$). Both curves show two different regions, regions I and II. In region I, Δ is negative and alternate routing is more advantageous than non-alternate routing. The second region, considerably larger than the first, is the region where alternate routing causes performance degradation.

An engineered network is usually designed with link blocking probabilities ranging between 0.001 to 0.05. Referring to the curve labeled $c = 5$, we see that alternate routing performs much better than non-alternate routing within this range of operation. At some points, it could

provide a grade of service much higher than non-alternate routing. For example, for P_n equal to 0.075, we see that alternate routing gives an improvement of about 60%. In spite of the fact that the first region covers a relatively small portion of P_n , an engineered network is usually operating in this region.

However, it may occasionally happen that the input traffic is much higher than the normal traffic load. Figure 1 shows that with P_n greater than A, alternate routing performs poorly. From Fig.1, we see that in the neighborhood of point A, Δ is very sensitive to P_n . Since P_n is itself a positive function of the input traffic, we conclude that a small change in the input traffic around this point would have a great effect on the performance of the routing. The curve corresponding to $c = 10$ is given to show the effect of c on the alternate routing method. Both curves follow a similar pattern except that for $c = 10$, alternate routing begins to deteriorate faster than the case in which $c = 5$. Therefore, it is true to say that the region in which alternate routing shows a better performance than non-alternate routing reduces as the number of circuits on the links increases. But it is also true that a network with a large number of circuits is usually designed for less blocking rate. Thus, we observe again that alternate routing gives a relatively better performance in the normal traffic load region.

Nevertheless, as mentioned earlier, it is very likely that any network can be overloaded and if that happens, we've seen that uncontrolled alternate routing becomes even worse than non-alternate routing. From these observations, we realize that some form of routing control is clearly needed to protect the network from overload conditions.

Generally speaking, the problem of overload conditions in alternate routing may be eliminated using one of the two following methods. In one approach, P_n is periodically measured and if P_n is found to be in the second region, immediate action is necessary to ban all the overflow traffic from using secondary paths. Another approach is to apply some sort of state dependent control mechanism to alleviate the overflow traffic as it begins to build up. The first approach is less attractive since it requires continuous observation and an extensive computation task.

In this paper, we focus on the second approach, i.e., a control policy which is more adaptive and varies with input traffic. One way to implement this policy is to introduce a condition for using alternate paths. For example, one can reserve a certain number of channels (circuits) specifically for direct calls. Under this condition, a call blocked by its direct path can try an alternate path if the number of free channels on each link of the alternate path is greater than the number of channels reserved for the directly routed

traffic. This method is well known as the "Trunk Reservation Scheme" and has been studied by many people. For example, see [6, 8, 13].

In the case of an infinite source of users, the analysis of the trunk reservation technique is easily carried out. Using a single birth-death steady-state traffic model, one can first formulate the state probability function and then evaluate the link blocking probability and the possibility that the overflow call is able to use the alternate path. The analysis is similar to the one used to obtain the Erlang B formula. The results of trunk reservation schemes have shown that an appropriate trunk reservation algorithm can prevent congestion and operate within an acceptable range of network performance.

In this chapter, we intend to apply this method to the case of two types of users. In a community of heterogeneous users, trunk reservation technique has not been discussed before. In the case of two classes of users, the trunk reservation scheme could be applied using one of the following cases. In one case, alternately routed calls belonging to class 1 or class 2 are allowed to attempt alternate paths if the number of circuits that are used by the same type of users is less than a specified value. In particular, two threshold values are assigned to each link, one for each group of users. A call blocked by its first path is allowed to try a secondary path if and only if the

number of circuits used by the same type of users on each link of the secondary path is less than the corresponding threshold value. We refer to this case as "non-uniform trunk reservation". In the second case, one threshold is just assigned to each link. Here, a call belonging to either class is rejected by a secondary path if the number of idle circuits on each link of the path is less than the link threshold value. In this case both groups of users are subject to the same condition. This case is here referred to as "uniform trunk reservation."

Following this introduction, non-uniform and uniform trunk reservation models are presented in Sections 2 and 3, respectively. Section 4 introduces the trunk reservation into alternate routing networks and the model is then applied to fully connected networks. Section 5 includes numerical examples and remarks.

2. NON-UNIFORM TRUNK RESERVATION (case I)

In this section, we introduce the non-uniform trunk reservation model and evaluate the link performance. To determine the link performance, we form the following state-transition-rate diagram for Poisson arrivals, c -servers (channels), and a finite population queueing model (see figure 2). Let n_k and $n_k = r_k n_k'$ ($k = 1, 2$) be respectively

the number of class k customers that are offered directly and alternately to the link. Thus, the total number of class k customers will be $N_k = n_k + n'_k = n_k(1+r_k)$. From a total of N_k , n'_k customers are subject to trunk reservation. They would be able to use the link if the number of circuits occupied by the same class of traffic is less than t_k ($k = 1, 2$), where t_k is the threshold value assigned to the link for type k traffic.

In Fig. 2, b'_i and b''_j are the number of class 1 and class 2 customers that may be present on the link during the transition period from state (i, j) to the state $(i+1, j)$ and to the state $(i, j+1)$, respectively. Furthermore, (i, j) represents the state in which i type 1 and j type 2 calls are in progress.

In contrast to the case studied before (Ch. II), b'_i and b''_j depend, in this case, on the number of channels used by the corresponding class of users. For each class, there exist two cases. Considering b'_i , for example, either i is less than t_1 or it is equal to or greater than t_1 . Under this scheme, a type k call can try alternate routing if the number of type k calls that are in progress is less than t_k . b'_i and b''_j for both cases are now given by

$$b'_i = \begin{cases} N_1 - i & i < t_1 \\ n_1 - i + t_1 \frac{n_1}{N_1} - N_1 - i & i \geq t_1 \end{cases} \quad (1)$$

$$b''_j = \begin{cases} N_2 - j & j < t_2 \\ n_2 - j + t_2 \frac{n_2}{N_2} = N_2 - j & j \geq t_2 \end{cases} \quad (2)$$

Equation (1), for example, is obtained as follows. When $i < t_1$, both direct and alternate calls from type 1 are allowed to enter the system. Under this condition, b'_i is just equal to N_1 , the total number of type 1 customers, minus i , the number of type 1 users that are already in progress. When the system reaches the threshold value, i.e., $i \geq t_1$, only direct calls can have access to the system. Without alternate routing, b'_i would be equal to $(n_1 - i)$. But due to the alternate routing, not all i channels are used by directly routed calls. t_1 channels out of i are shared by both directly and alternately routed type 1 traffic. Therefore, the portion of t_1 channels that is used by alternate calls before the system reaches the threshold must be added to $(n_1 - i)$. Similarly, b'_j is obtained.

Under equilibrium, the steady-state requirement yields the following equation:

$$\begin{aligned} [i \mu_1 + j \mu_2 + \lambda_1 b'_i + \lambda_2 b''_j] P(i, j) = \\ \lambda_1 b'_{i-1} P(i-1, j) + (i+1) \mu_1 P(i+1, j) + \\ \lambda_2 b''_{j-1} P(i, j-1) + (j+1) \mu_2 P(i, j+1) \end{aligned} \quad (3)$$

where $p(i, j)$ stands for the probability that i type 1 and j type 2 customers are present in the system. Referring to figure 2, depending on i and j , four different regions can be realized. They are:

- 1) $i \leq t_1$ and $j \leq t_2$
- 2) $i \geq t_1$ and $j \leq t_2$
- 3) $i \leq t_1$ and $j \geq t_2$
- 4) $i \geq t_1$ and $j \geq t_2$

The general state-flow-rate equation given by (3) is satisfied by all the above four regions. Using the product method introduced in Chapter II, it can be shown that each region has a product form solution similar to the one given by equation (10) in Chapter II. In order to prove that the complete solution is also in product form, we must show that the boundary conditions are satisfied in all four regions. The solution to each region is now given

region I

$$P(i, j) \stackrel{\Delta}{=} P_1(i, j) = P'_1 a_1^i a_2^j \binom{N_1}{i} \binom{N_2}{j} \quad i \leq t_1 \text{ and } j \leq t_2 \quad (4.1)$$

region II

$$P(i, j) \stackrel{\Delta}{=} P_2(i, j) = P'_2 a_1^i a_2^j \binom{N_1}{t_1} \binom{N_2}{j} \prod_{k=0}^{i-t_1-1} \frac{N_1-k}{t_1+k+1} \quad i \geq t_1 \text{ and } j \leq t_2 \quad (4.2)$$

region III

$$P(i, j) \triangleq P_3(i, j) = P'_3 a_1^i a_2^j \binom{N_1}{i} \binom{N_2}{t_2} \prod_{k=0}^{j-t_2-1} \frac{N_2-k}{t_2+k+1}$$

$$i \leq t_1 \text{ and } j \geq t_2 \quad (4.3)$$

region IV

$$P(i, j) \triangleq P_4(i, j) = P'_4 a_1^i a_2^j \binom{N_1}{t_1} \binom{N_2}{t_2} \prod_{k=0}^{i-t_1-1} \frac{N_1-k}{t_1+k+1} \prod_{k=0}^{j-t_2-1} \frac{N_2-k}{t_2+k+1}$$

$$i \geq t_1 \text{ and } j \geq t_2 \quad (4.4)$$

where P'_1 to P'_4 are some constants to be determined. The boundary conditions are

$$P_1(t_1, j) = P_2(t_1, j)$$

$$P_3(i, t_2) = P_1(i, t_2)$$

$$P_2(i, t_2) = P_4(i, t_2) \quad (5)$$

$$P_3(t_1, j) = P_4(t_1, j)$$

$$\sum_{l=1}^4 \sum_{\substack{\text{all possible} \\ i \text{ and } j}} P_l(i, j) = 1$$

Using these boundary conditions, we get

$$P'_1 = P'_2 = P'_3 = P'_4 \triangleq P(0, 0) \quad (6)$$

Having found the state probability function, we can now

evaluate the link performance. The link is blocked to direct calls if the sum of the customers (type 1 and type 2) using the link is equal to the number of circuits in the link. Denoting this blocking probability by P, it is given by

$$P = \sum_{i+j=c} P(i, j) \quad (7)$$

In addition, the link would be blocked to alternately routed type k customers if the number of channels occupied by any type k customers is at least equal to t_k . Let P_k ($k = 1, 2$) be the probability that alternate type k calls are blocked. P_k is given by

$$P_1 = \sum_{j=0}^{c-t_1} \sum_{i=t_1}^c P(i, j) \quad (8)$$

$$P_2 = \sum_{i=0}^{c-t_2} \sum_{j=t_2}^c P(i, j)$$

3. UNIFORM TRUNK RESERVATION (case 2)

In this case, the decision to allow alternately routed calls to attempt an alternate link is based on the number of busy channels in the link. In particular, for a link with c number of channels, (c-t) channels are reserved for direct

calls and the remaining t channels are used by both alternately and directly routed calls. With i and j respectively the number of type 1 and type 2 customers using the link, an alternately routed call is allowed to use the link if $i+j < t$. Therefore, the probability that a link is blocked to directly routed calls is different from the probability that the same link will be blocked to alternately routed calls. However, both types of alternately routed calls are subject to the same blocking rate (uniform trunk reservation). Let P be the probability that a direct call is blocked and P_{12} be the probability that an alternate call is blocked by the link. In this section, we will try to estimate P and P_{12} .

Depending on the total number of customers on the link, state (i, j) can be in one of the two possible regions. Referring to the state diagram in Fig.3, we see that the state (i, j) may be in region I if $i+j < t$, or in region II if $i+j \geq t$. Here, b'_i and b''_j are, as in the previous section, the possible number of type 1 and 2 customers given that the system is in the state (i, j) . For this case, b'_i and b''_j are interchangeably related to each other and depend on the region where they are. When $(i+j) < t$, both alternately and directly routed calls are allowed to use the link. Thus, b'_i and b''_j are simply given by

$$b'_i = n_1 + n'_1 - i = N_1 - i \quad i+j < t \quad (9)$$

$$b''_j = n_2 + n'_2 - j = N_2 - j$$

In region II where $i + j \geq t$, alternately routed calls are rejected. Under this condition, b_i and b_j are calculated as follows. Given the state (i, j) and $i + j \geq t$, t channels out of $(i + j)$ channels are occupied by both directly and alternately routed calls. Under the steady-state condition, t channels are shared by both traffics according to

$$t_1 = t \frac{i}{i + j} \quad (10)$$

$$t_2 = t \frac{j}{i + j}$$

and, of course, $t = t_1 + t_2$. Now, we must determine the portions of t_1 and t_2 that are allocated to alternately routed type 1 and type 2 traffic, respectively, before the system enters the second region. With n_k and n'_k as described in section 2, t_k channels are divided between type k customers, both direct and alternate calls, with a ratio proportional to the population sizes of users. That is, with a total number of $N_k = n_k + n'_k = n_k(1+r_k)$ users, $t_k n'_k / N_k = t_k r_k / (1+r_k)$ are assigned to alternately routed type k calls and $t_k n_k / N_k = t_k / (1+r_k)$ are given to directly routed calls. Therefore, with the system in the state (i, j) , the total number of direct type k ($k=1, 2$) calls that are in progress

is given by

$$b_i = i - \frac{r_1}{1 + r_1} t_1 = i - \frac{r_1}{1 + r_1} t \frac{i}{i + j} \quad (11)$$

$$b_j = j - \frac{r_2}{1 + r_2} t_2 = j - \frac{r_2}{1 + r_2} t \frac{j}{i + j}$$

In a network in which both types of traffic are subject to the same trunk reservation rule in every link, the blocking probabilities experienced by both types of users are the same. Thus, $n'_1/n_1 = r_1 = n'_2/n_2 = r_2 = r$ and b'_i and b''_j can be found by

$$b'_i = n_1 - b_i = n_1 + \frac{r}{1 + r} t \frac{i}{i + j} - i \quad i+j \geq t \quad (12)$$

$$b''_j = n_2 - b_j = n_2 + \frac{r}{1 + r} t \frac{j}{i + j} - j$$

Using equations (9) and (12), now we would like to solve the balance equation (3) to find the state probability function, $P(i, j)$. Applying the product method, as before, to this system, it can be shown that each region has a product form solution. However, it is not possible to find a general closed form solution, similar to section II, that can satisfy the boundary conditions. The reason for not

having a complete product form solution is because b'_i and b''_j are not independent from each other; both b'_i and b''_j depend on i and j and are not separable.

In the remainder of this section, we use some approximations to solve the above system equation and estimate the link performance. Starting with equation (12), we try now to approximate b'_i and b''_j in such a way that they become independent from each other. The approximation of b'_i is given in the following. A similar approximation is applied to b_j . Rewriting the equation (12), we have

$$\begin{aligned}
 b'_i &= n_1 + \frac{r}{1+r} t \frac{i}{i+j} - i \\
 &= n_1 - \frac{i}{1+r} \left(1 + r \left(1 - \frac{t}{i+j} \right) \right)
 \end{aligned}
 \tag{13}$$

In this equation with $(i+j) \geq t$, $\frac{t}{i+j} \leq 1$ and since r , the fraction of direct calls that are rejected by the first path, is usually much less than 1, $r \left(1 - \frac{t}{i+j} \right) \ll 1$ and can be negligible when compared to 1. Denoting the approximated version of b_i by bb_i , we get

$$b'_i \sim bb_i = n_1 - \frac{i}{1+r}
 \tag{14}$$

Furthermore, $N_1 = n_1(1+r)$. Substituting this in (14), we have

$$b'_i - bb_i = \frac{1}{1+r} (N_1 - i) \quad i+j \geq t \quad (15)$$

Similarly for b_j

$$b''_j - bb_j = \frac{1}{1+r} (N_2 - j) \quad i+j \geq t \quad (16)$$

Now we use equations (9), (15), and (16) to solve the equilibrium equation (3). Using the product method and proceeding as in Chapter II, the solution to region I and region II is given by

$$P(i, j) \triangleq P_1(i, j) = P_1 a_1^i a_2^j \binom{N_1}{i} \binom{N_2}{j} \quad i + j \leq t \quad (17)$$

$$P(i, j) \triangleq P_2(i, j) = P_2 a_1^i a_2^j \left(\frac{1}{1+r}\right)^{i+j} \binom{N_1}{i} \binom{N_2}{j} \quad i + j \geq t \quad (18)$$

here, $a_k = \lambda_k / \mu_k$ is the load offered by any type k customer in erlangs and P_1 and P_2 are the two constants to be determined using the boundary conditions. Boundary conditions are

$$\sum_{\substack{\text{all possible} \\ i \text{ and } j}} P(i, j) = 1 \quad (19)$$

and

$$P_1(i, j) \Big|_{i+j=t} = P_2(i, j) \Big|_{i+j=t} \quad (20)$$

Using (20), we get

$$P_2 = P_1(1 + r)^t \quad (21)$$

Now, using both equations (19) and (21), P_1 is obtained as

$$P_1 = \left[\sum_{i+j=0}^t a_1^i a_2^j \binom{N_1}{i} \binom{N_2}{j} + (1+r)^t \sum_{i+j=t+1}^c a_1^i a_2^j \binom{N_1}{i} \binom{N_2}{j} \right]^{-1} \quad (22)$$

We are now ready to evaluate the probability that a link is blocked to a customer. Under this scheme, the probability that a link is blocked to a direct call, type 1 or type 2, is the same as the probability that all the channels in the link are busy. Denoting this probability by P , P is given by

$$P = \text{Prob.}(i + j = c) = \frac{P_1}{(1+r)^{c-t}} \sum_{i=0}^c a_1^i a_2^{c-i} \binom{N_1}{i} \binom{N_2}{c-i} \quad (23)$$

where P_1 is given by (22). Next, we obtain the probability that an alternate call is blocked by a link. First, note that in uniform trunk reservation, both types of alternately

routed calls are subject to the same link blocking. Now let P_{12} be the probability that an alternate call is rejected by the link. P_{12} is given by

$$P_{12} = \text{Prob.}(i + j \geq t) = 1 - \text{Prob.}(i + j < t) =$$

$$1 - P_1 \sum_{i=0}^{t-1} a_1 \binom{N_1}{i} \sum_{j=0}^{t-1-i} a_2 \binom{N_2}{j} \quad (24)$$

In the following section, we will apply the above two cases to alternate routing networks, and numerical comparisons of fully connected networks will be presented.

4. ALTERNATE ROUTING WITH TRUNK RESERVATION

The network model to be considered here is again a fully connected network with N nodes and $N(N-1)/2$ duplex links. The general alternate routing technique described in sections 2 and 3 of the last chapter can easily be modified to include both trunk reservation schemes in networks with different topologies; however, since the results, even for a small network, are not mathematically tractable and we are interested in numerical comparison rather than simulation, we have decided to consider fully connected networks. Furthermore, the alternate routing and non-alternate routing in fully connected networks have already been discussed and

the results are available for comparison. In addition to the standard assumptions used in the performance evaluation of almost every circuit-switched network, three more assumptions are adopted here to simplify the analysis. They are:

- 1) All the source-destination node pairs have the same number of type k customers.
- 2) c, number of channels, is the same for every link.
- 3) The traffic distribution is uniform.

The following is the list of notations used in this section. The notations are analogous to the notations used in Chapter III.

- N_k^j = number of class k customers for the source-destination j (k=1, 2)
- a_k = load offered by any type k customer
- r_m^j = the m^{th} path for the source-destination j
- R^j = the total number of paths for the source-destination j
- R_g^j = the 1^{st} g paths of R^j .
- N_k = the total number of class k customers offered to a link
- nn_k = the total number of class k customers served by a link
- $n_k^j(m)$ = number of class k customers served by the m^{th} path for the source-destination j
- $P_k(r_m^j)$ = Prob. (the m^{th} path of the source-destination j is blocked)

$P_k(R_g^j)$ - Prob. (the 1st g paths of the source-destination j are blocked)

Using the above notations, $n_k^j(m)$, the total number of class k customers served by the mth path for the source-destination j is given by (see equation (3) in chapter III)

$$n_k^j(m) = N_k^j [P_k(R_{m-1}^j) - P_k(R_m^j)] \quad (25)$$

In fully connected networks, the first path of a source-destination node pair is the direct link between the two nodes. A call, type 1 or type 2, is blocked by its 1st path if all the channels in the link are busy. This blocking is shown here by P and is the same for all the links, that is

$$P_k(R_1^j) = P \quad 1 \leq j \leq N(N-1) \quad (26)$$

For alternately routed calls, type 1 or type 2, an alternate path is available if every link in the path satisfies the corresponding trunk reservation condition. Recalling the above assumptions, we see that the alternate link blocking probability is also the same for all links. Let P_k be the probability that an alternate link is not available to alternately routed type k traffic. For the case of uniform trunk reservation, an alternate link is blocked to both types of users with an equal probability, that is, $P_1 = P_2$. This case will be considered later. Moreover, we consider

here the two-link alternate routing algorithm which was introduced in Chapter III. Therefore, an alternate path consists of two links and the probability that an alternate path is blocked to a type k customer is obtained as

$$P_k(r_m^j) = 1 - (1 - P_k)^2 = 2P_k - P_k^2 \quad \begin{matrix} k=1, 2 \\ 1 \leq j \leq N(N-1) \\ 1 \leq m \leq s+1 \end{matrix} \quad (27)$$

where s is the number of alternate paths that a call may attempt. It may vary, for two-link alternate routing, between zero to N-2 providing that $N \geq 3$. Now, the probability that the 1st m paths of a class k customer are blocked is given by

$$P_k(R_m^j) = P(2P_k - P_k^2)^{m-1} \quad (28)$$

Substituting (26) and (28) into (25), we get

$$\begin{aligned} n_k^j(1) &= N_k^j(1 - P) & m=1 \\ & & (29) \\ n_k^j(m) &= N_k^j P(1 - P_k)^2 (2P_k - P_k^2)^{m-2} & m \geq 2 \end{aligned}$$

We can now calculate nn_k , the total number of class k customers that are served by a link. For the derivation, see equation (16) in Chapter III. nn_k is given by

$$nn_k = N_k^j \{ (1 - P) + 2P[1 - (2P_k - P_k^2)^s] \} \quad (30)$$

This equation can be decomposed into two parts. The first part, $N_k^j(1 - P)$, is the number of type k customers that are directly served by the link. The second part, $2N_k^jP[1 - (2P_k - P_k^2)^s]$, is the number of type k customers that are alternately served by the link. Thus, the number of type k customers that are directly or alternately offered to each link will be

$$n_k = \frac{N_k^j(1 - P)}{1 - P} = N_k^j \quad (31)$$

and

$$n_k = r_k n_k = 2P \frac{N_k^j}{1 - P_k} [1 - (2P_k - P_k^2)^s] \quad (32)$$

Comparing (32) with the notation $n_k = r_k n_k$, we see that

$$r_k = \frac{2P}{1 - P_k} [1 - (2P_k - P_k^2)^s] \quad (33)$$

Now, let BP_k represent the end-to-end blocking probability of type k customers. BP_k is simply obtained by noting that a type k call is rejected if it has tried, in addition to the direct path, up to s alternate paths and failed to establish an end-to-end connection. This blocking is given by

$$BP_k = P(2P_k - P_k^2)^s \quad (34)$$

where P and P_k are given by (7) and (8). n_k and n_k , the two

variables needed to compute P and P_k , are shown by (31) and (32). In this case, the average network blocking probability is determined by

$$BP_{ant} = \frac{\sum_{k=1}^2 a_k n_k BP_k}{\sum_{k=1}^2 a_k n_k} \quad (35)$$

here, the abbreviation "ant" stands for "Alternate routing with Non-uniform Trunk reservation".

For the case of uniform trunk reservation, we indicated earlier that the alternate link blocking probability for both types of customers is the same. So, in the uniform trunk reservation $P_1 = P_2 = P_{12}$, and one can easily modify the above equations to obtain the end-to-end blocking probability. This is done by replacing both P_1 and P_2 by P_{12} . In addition, the end-to-end blocking probability, for this case, is the same as the average blocking rate. BP_{aut} , the end-to-end blocking probability of any type of customer, type 1 or type 2, for alternate routing with uniform trunk reservation is now given by

$$BP_{aut} = P(2P_{12} - P_{12}^2)^s \quad (36)$$

with P and P_{12} given by (23) and (24).

5. NUMERICAL EXAMPLES

Using the equations derived in this chapter, we would now like to compare the circuit-switched performance of various alternate routing and non-alternate routing policies. The results are obtained for the fully connected and symmetrical network model. The performance measures used for the comparison are the end-to-end blocking probability and the average network blocking rate. However, for a fully connected network, the average end-to-end blocking rate is the same as the end-to-end blocking probability except for the case of the alternate routing with non-uniform trunk reservation. In order to gain more insight into the performance of different routing strategies, we have considered several examples. Each example is used to explore and answer one of the several questions arising in the comparison. In particular, we are interested in comparing the performance of routing models under different design variables such as the user population and the channel capacity.

First, we begin with the same network model used in the example that was given in the previous chapter. Figure 4 plots the end-to-end blocking probability as a function of a_2 . Here, $c=5$ and the number of type 1 and type 2 customers that are directly attached to each node are chosen

to be 600 and 20, respectively. It is also assumed that the ratio of the individual offered load of the second group to the first group of users, a_2/a_1 , is 50. In the 5-node fully connected model used in this example, the maximum number of possible two-link alternate paths is three ($s \leq 3$); however, as far as heavy load conditions are concerned, we have seen that $s = 1$, that is, one alternate path only, gives a better performance than the cases in which $s = 2$ and $s = 3$. Figure 4 shows that with $s = 1$, the two-link alternate routing without trunk reservation (full alternate routing) would provide an end-to-end blocking probability lower than that of non-alternate routing if a_2 is smaller than a_2 . When $a_2 = a_2$, both non-alternate and full alternate routing give the same blocking rate. At this point, $BP_a = BP = 14.5\%$. In a non-alternate routing network, a grade of service of about 3% is usually considered a typical blocking rate for normal traffic loads. Figure 4 shows that with BP equal to 3%, $a_2 = 0.3$ erlangs. Comparing a_2 with 0.3, we notice that a_2 is about 70% greater than 0.3. This implies that alternate routing would perform better than non-alternate routing even if the traffic load increases up to 70% of its normal value.

Under heavy load conditions, as we can see from figure 4, using full alternate routing can actually worsen the network performance, i.e., $BP_a > BP$ when a_2 exceeds a_2 . The curve labeled BP_{ant} corresponds to the alternate routing method with non-uniform trunk reservation. For this curve,

in addition to $s = 1$, t_1 and t_2 are selected to be 1 and 3. That is, an alternately routed type k ($k = 1, 2$) call may attempt an alternate path if the number of circuits occupied by the same type of calls is less than t_k . As we mentioned earlier, the reason for using this constraint is to block alternately routed traffic more often and, therefore, to prevent links from becoming highly utilized. It also helps to protect the directly routed traffic against overload conditions. Inspection of figure 4 shows how alternate routing with non-uniform trunk reservation can considerably improve the network performance. For example, at the point where the two curves BP_{ant} and BP intercept, we find that the blocking probability is 29%. With $BP = 29\%$, a_2 is found to be 0.8 which is almost 170% greater than the normal load - a substantial improvement as compared to full alternate routing.

In the next example, we consider the alternate routing with uniform trunk reservation. In figure 5, we plot the end-to-end blocking probability as a function of a_2 for three different cases. They include alternate routing with $s=1$ and $t=5$ (BP_a), alternate routing with $s=1$ and $t=3$ (BP_{aut}), and non-alternate routing (BP). The network specifications are assumed to be the same as before. Referring to figure 5, we see that alternate routing with uniform trunk reservation displays a blocking performance

similar to the case of non-uniform trunk reservation. It shows how, by dedicating a small number of circuits to directly routed calls, this technique can improve the network performance. When compared with the alternate routing technique ($t=5$), the improvement in network performance becomes more noticeable as the traffic load increases.

The main difference between the two trunk reservation techniques lies in the fact that in a non-uniform case, one can also give priority to a class of users. This priority is usually given to users with large traffic requirements, as we've seen in figure 4 ($t_1=1$ and $t_2=3$). To see the effect of non-uniform trunk reservation on an individual user, BP_1 and BP_2 , the end-to-end blocking probabilities of type 1 and type 2 users are plotted in figure 6. In addition, two more curves are shown as two reference curves and correspond, respectively, to non-alternate routing and the average network blocking probability of non-uniform trunk reservation. From the figure, it is apparent that in the region in which the average blocking rate, BP_{ant} , is less than BP , BP_1 is much higher than BP_2 , a situation which is usually undesirable. It is even shown that BP_1 becomes considerably higher than the case of non-alternate routing.

In order to improve BP_1 , in figure 7, t_1 is increased from 1 to 2. In comparing the curves in figure 7 with the ones in figure 6, we can see that in the region where the

input load is high, there is a trade-off between BP_1 and the average end-to-end blocking probability. BP_1 is improved in figure 7 but, on the other hand, the average blocking probability during overload conditions (i.e., the time when the prevention of congestion is most needed) is also increased as compared to the one in figure 6. This is in fact one of the disadvantages of non-uniform trunk reservation. Moreover, in a network using non-uniform trunk reservation, additional information and memory are required at each node of the network. This information is necessary to keep track of each class of users so that a decision as to whether or not to allow a particular type of call to use an alternate link can be made at each node. This will further complicate the routing procedure and thus makes this routing policy less advisable. Uniform trunk reservation, on the other hand, not only yields a performance comparable to the non-uniform case but is also simpler to implement.

More comparisons between the two trunk reservation schemes are made and the results are listed in Tables 1 and 2. In view of these results, we may conclude that alternate routing with non-uniform trunk reservation is a good candidate only if one wishes to offer better service to a specific group of users. On the other hand, when network performance, simplicity, and fairness are concerned, alternate routing with uniform trunk reservation seems to be

a proper choice. It does not favor any group of users over another and yet is able to improve the network performance during overload conditions.

Finally, we would like to examine the effect of link capacity, c , on alternate routing. Corresponding to $c=10$ and 15 , two sets of curves are plotted in figure 8. Each set contains two curves, one for non-alternate routing and one for alternate routing with trunk reservation. n_1 and n_2 are identical in both cases and $a_2=50a_1$. In addition, we have assumed that in each case 20-percent of the circuits (channels) is reserved exclusively for the directly routed traffic. That is, $t=8$ when c is equal to 10 and $t=12$ for $c=15$. Here, T_c ($c=10, 15$) represents the end-to-end blocking probability where the two curves meet each other. For a given traffic requirement, alternate routing performs better than non-alternate routing if the end-to-end blocking probability is less than T_c . The results for $c=10$ and $c=15$ indicate that T_c is affected by the number of channels, that is, $T_{15} < T_{10}$. This implies that as the channel capacity increases, the alternate routing starts to degrade earlier and, as a result, will become less preferred. In order to fully explore this situation, more curves are shown in figure 9. In this figure, T_c is plotted as a function of the channel capacity c . The network parameters used in this example are given in Table 3. Both curves show that T_c decreases as c increases.

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The major objective of this work was to analytically gain insight into the circuit-switched performance of the ISDN. In Chapter I, we first reviewed the recent evolution in the area of communication technology, described the ISDN environment, and gave the motivation for this research. We then presented the circuit-switched routing problem and outlined the framework of our research.

In Chapter II, we developed a mathematical model for evaluating the circuit-switched performance of a network with two groups of users (i.e., a community of heterogeneous users). As a specific illustration of a typical application, the model was then used to determine the circuit-switched capacity of CATV systems in terms of end-to-end blocking probability. Finally, in Chapter II, we extended our analysis to the case of more than two types of traffic.

In Chapter III, we discussed the routing problem and presented a general alternate routing algorithm applicable to the ISDN with an arbitrary topology. Next, we applied that algorithm to fully interconnected and symmetrical networks and compared alternate routing with non-alternate routing. The performance criteria used in the comparison were the average network blocking rate and end-to-end blocking probability. For light to moderate traffic loads,

the results of the analysis showed a performance improvement when using alternate routing. Under heavy traffic load conditions, however, we found that the performance of alternate routing deteriorates as compared to that of non-alternate routing. It is not surprising that alternate routing was not found helpful under these conditions. This is because beneficial alternate routing relies solely on the availability of network resources. In alternate routing networks, the average amount of resources involved with each communication increases since alternate paths are usually longer than primary paths. At normal loads, there exists a sufficient spare capacity in a network that can ultimately be used by alternately routed traffic. As the traffic load increases, links become highly utilized. Furthermore, because of the strong relationship between the link blocking probability and the offered load, a small change in the latter, due to the use of alternate routing, will have a great effect on the link blocking probability.

In Chapter IV, the heavy load performance of alternate routing was further discussed and the trunk reservation technique for the case of two groups of users was introduced. In particular, two different approaches to implement this policy were considered: 1) non-uniform and 2) uniform trunk reservation schemes. The results of the trunk reservation were then applied to fully connected networks and all four alternate and non-alternate routing methods

were numerically compared. The results indicated that the trunk reservation scheme can be used as an effective way to overcome the performance degradation caused by uncontrolled alternate routing. It was shown how, for example, by dedicating a small number of channels in each link to directly routed traffic, this technique can improve the network performance.

Unquestionably, this research has only scratched the surface of one of the design issues related to ISDN's. There appear to be two major areas in which further research is needed.

1) The comparison of various routing strategies based on fully connected networks seems to be incomplete. In order to make a satisfactory comparison, one would like to extend the analysis to include networks with different topologies; network topology plays an important role in routing performance.

2) In this work, we have assumed a fixed channel capacity and, therefore, eliminated the problem of bandwidth allocation among circuit- and packet-switched traffic. The assumption of fixed channel capacity for circuit-switched traffic is valid as long as the ISDN consists of two separated networks, one for circuit-switched traffic and the other for packet-switched traffic. In the next generation of ISDN's, however, it is expected that both types of traffic

will share the same transmission media. In that case, the problem of bandwidth allocation and routing must be investigated simultaneously and more research effort is required to study the interaction between the two types of traffic.

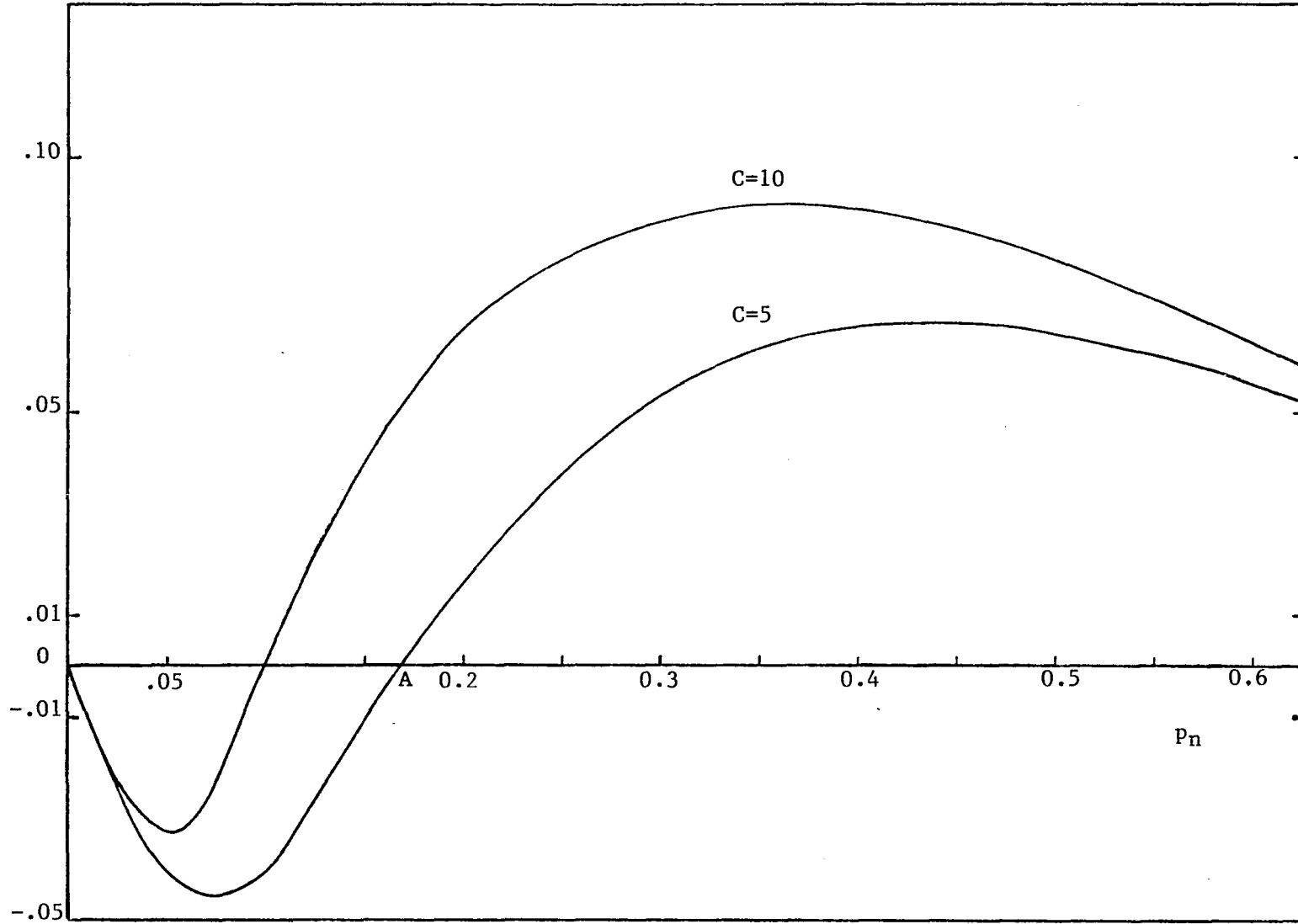


Fig.1 Performance comparison of non-alternate and alternate routing procedures.

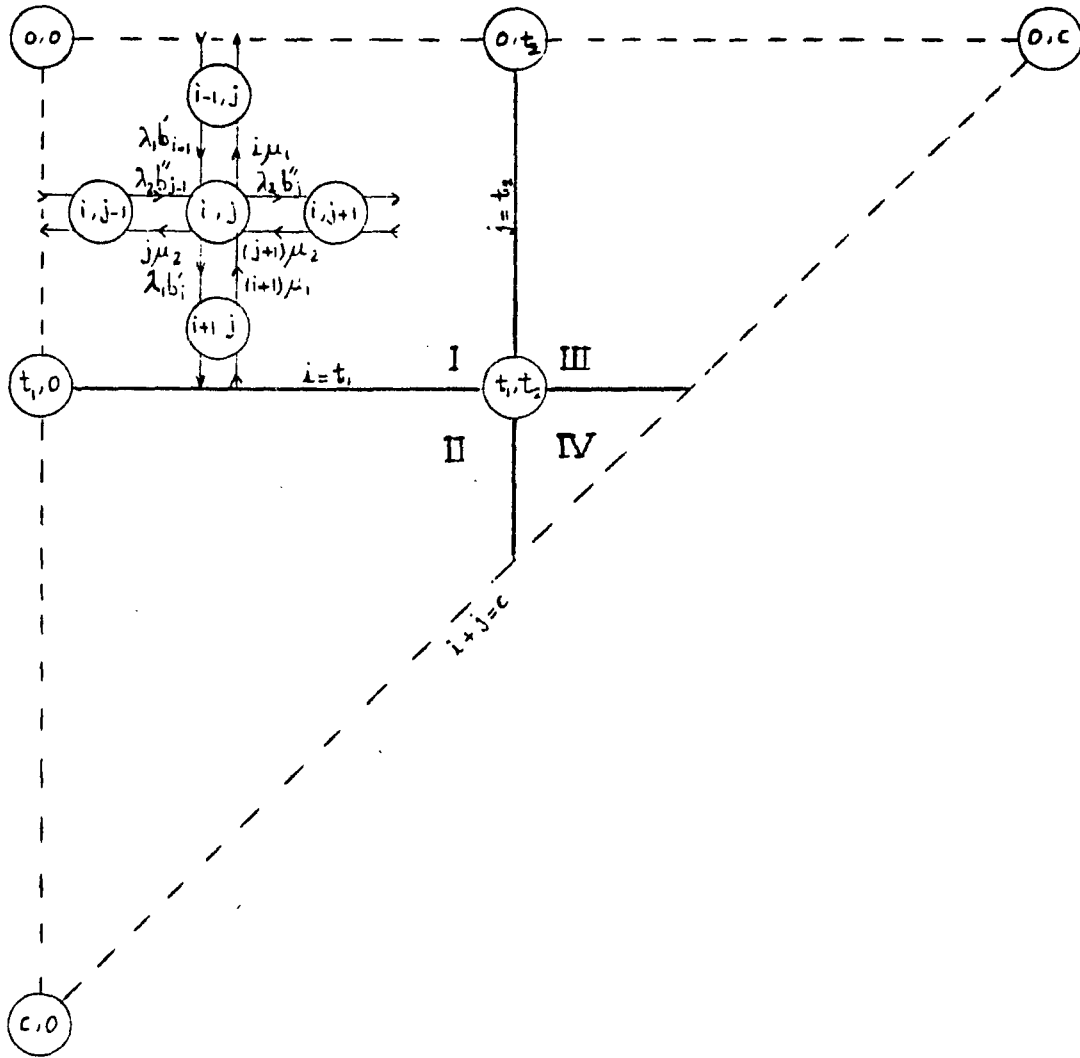


Fig.2 Non-Uniform Trunk Reservation (case I)

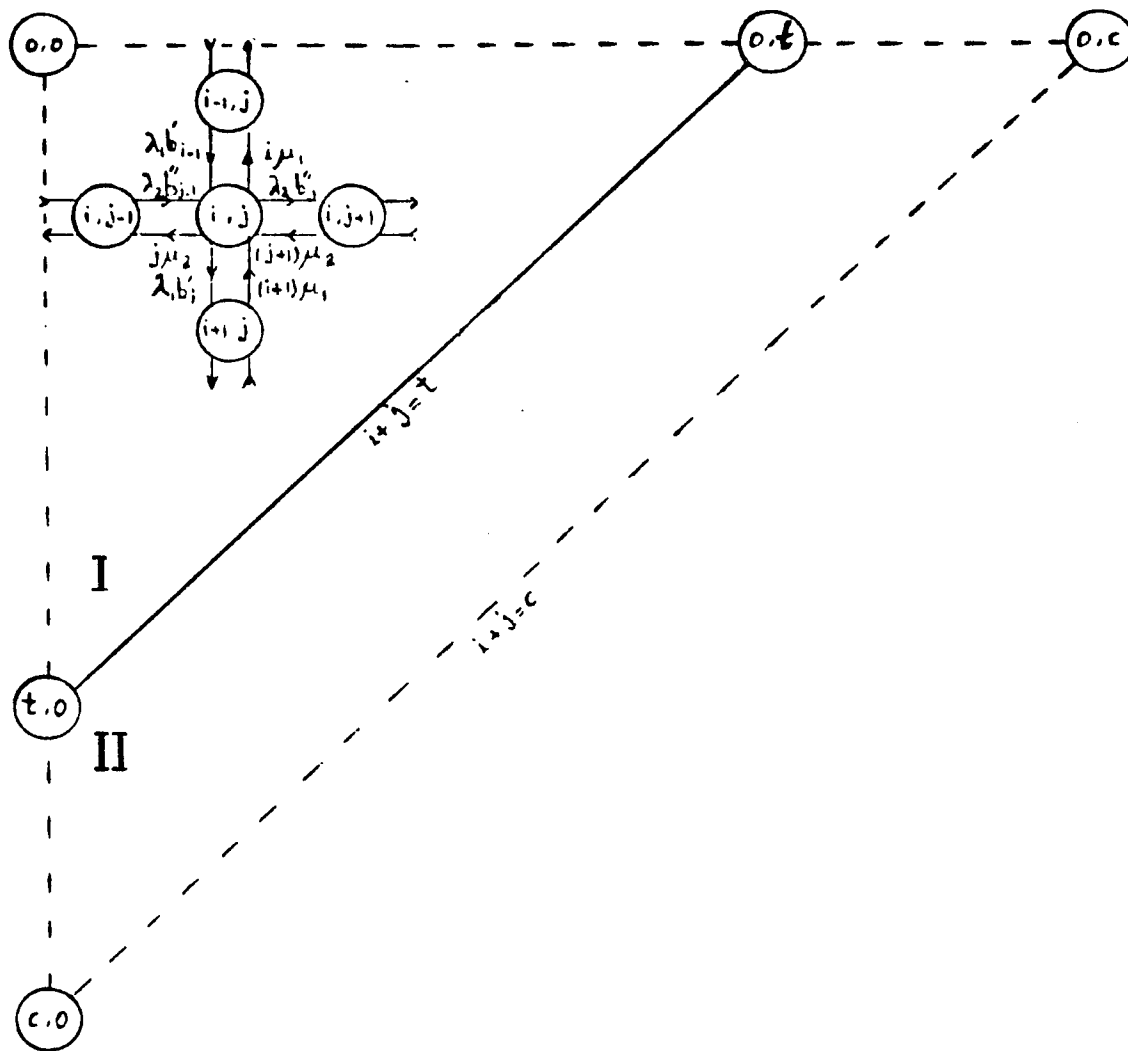


Fig.3 Uniform Trunk Reservation (case II)

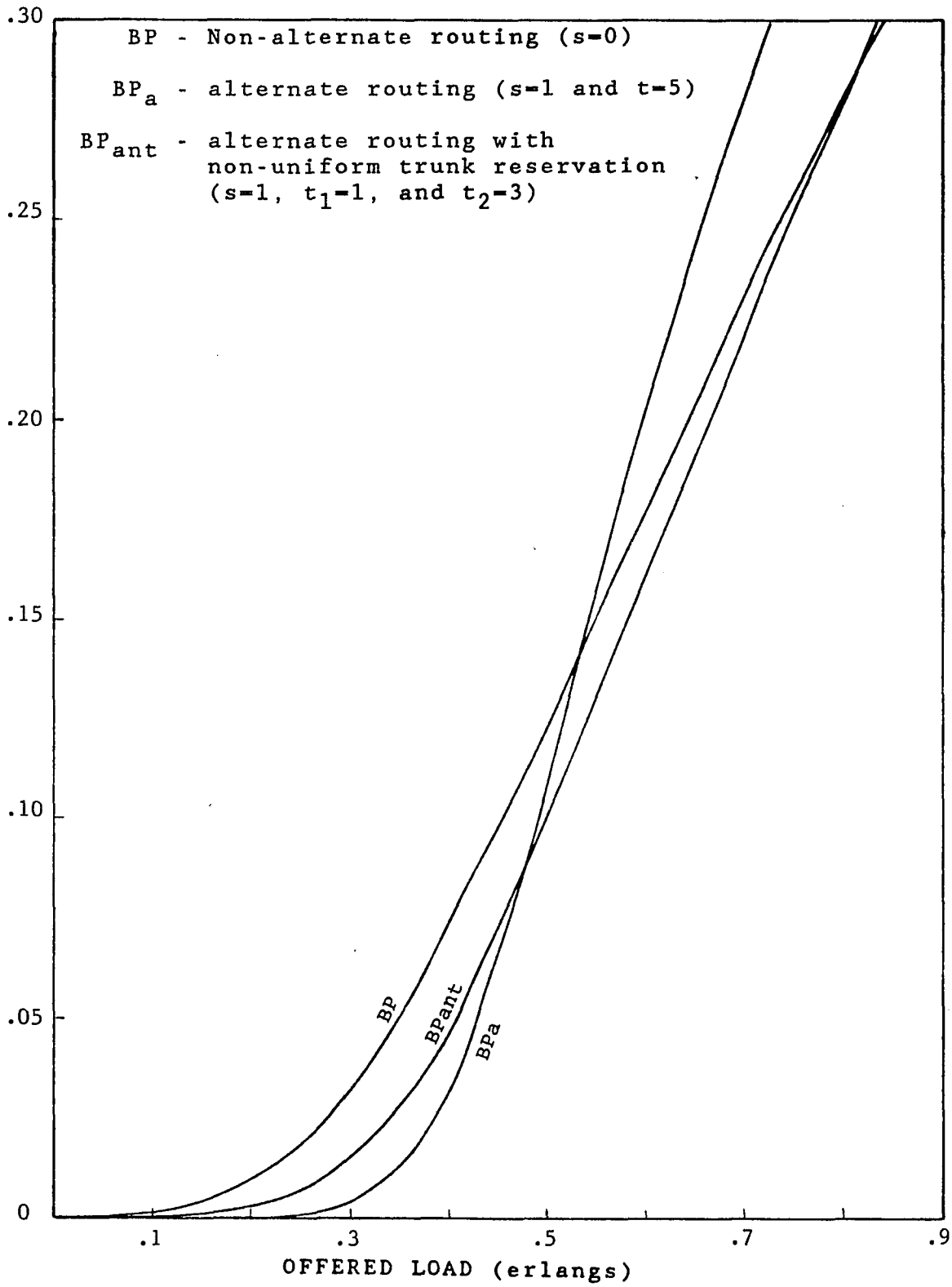


Fig. 4 End-to-End Blocking Probability vs. a_2

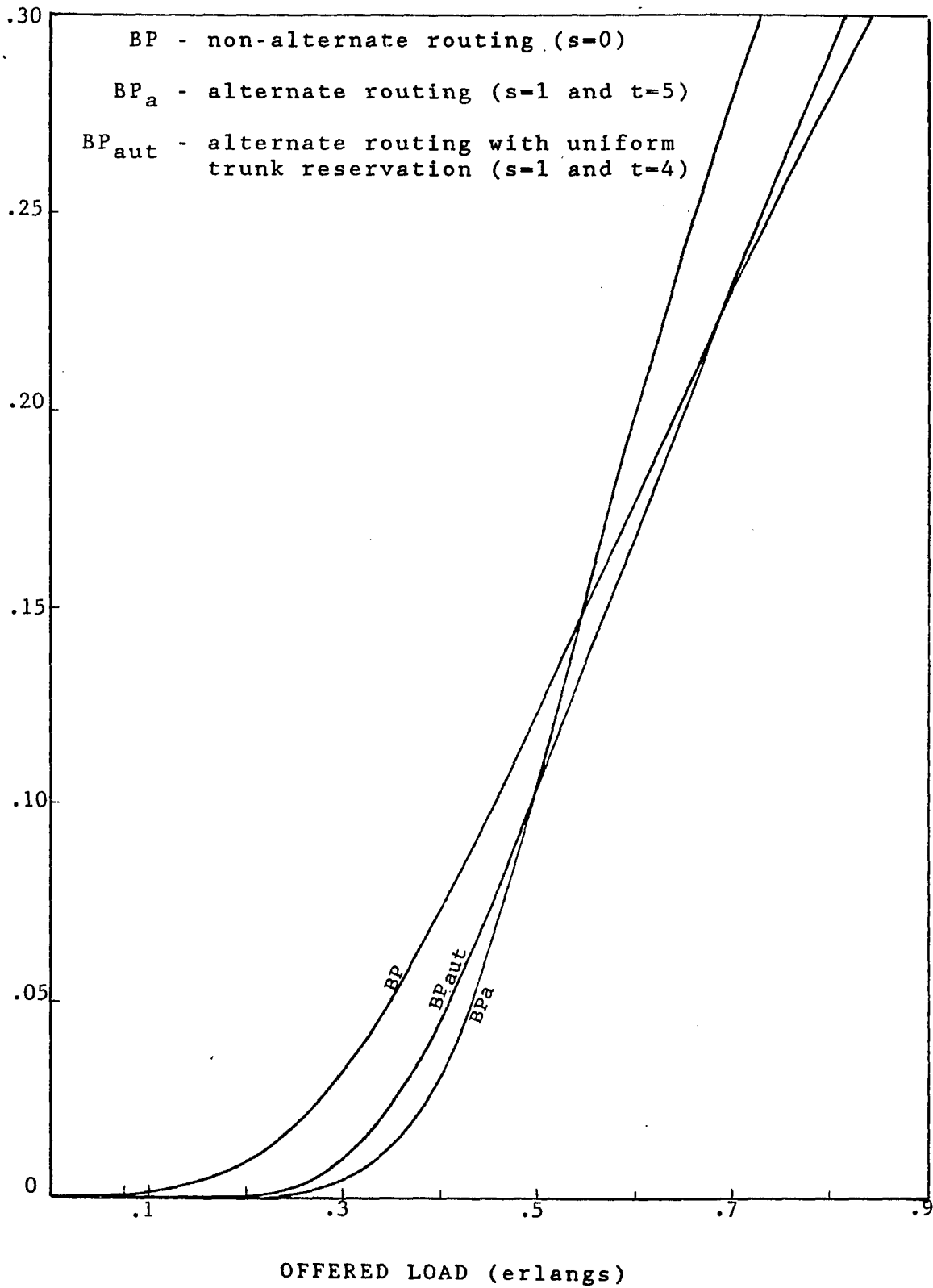


Fig. 5 End-to-End Blocking Probability vs. a_2

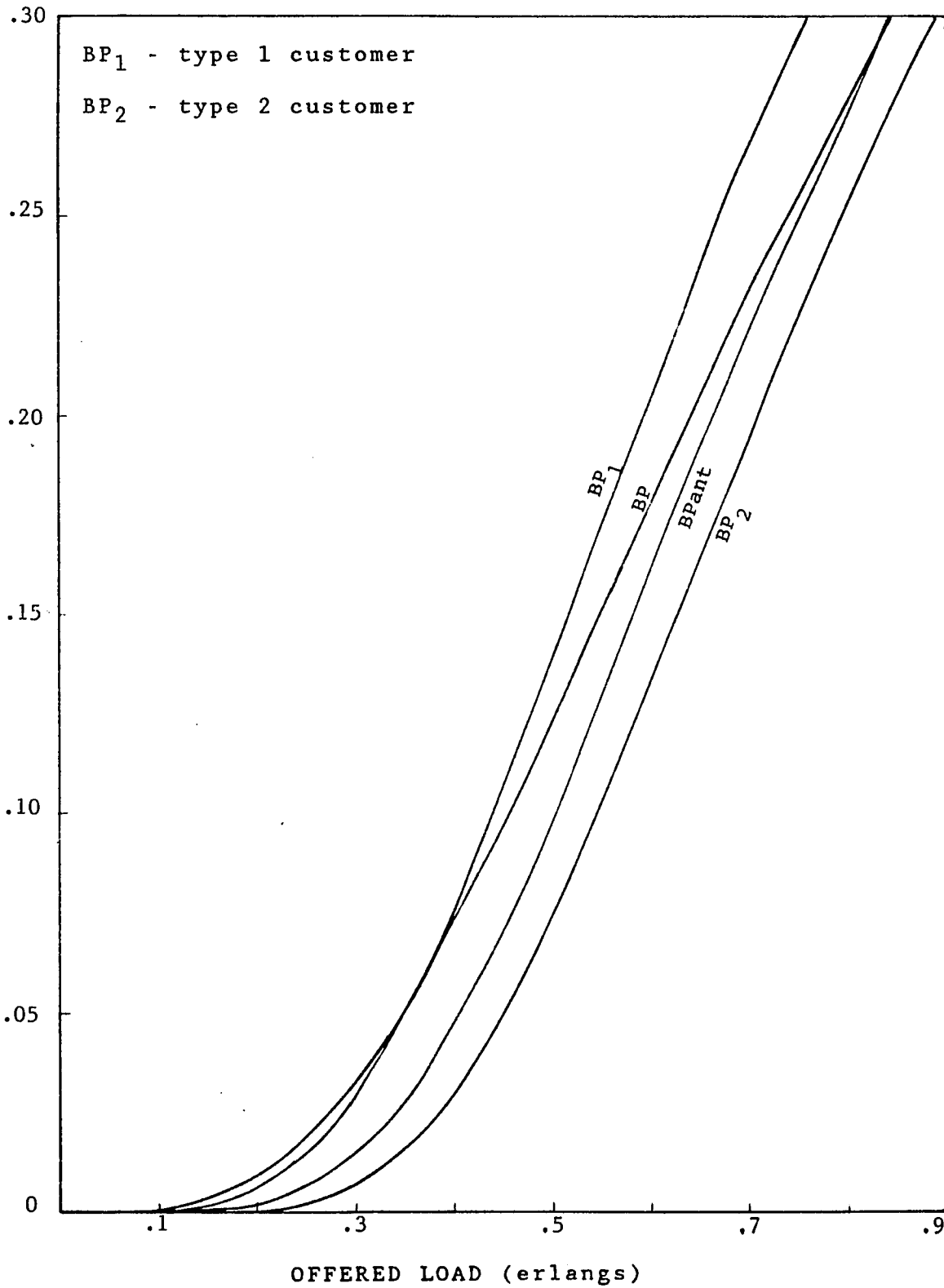


Fig.6 End-to-End Blocking Probability vs. a_2
(non-uniform trunk reservation with $t_1=1$ and $t_2=3$)

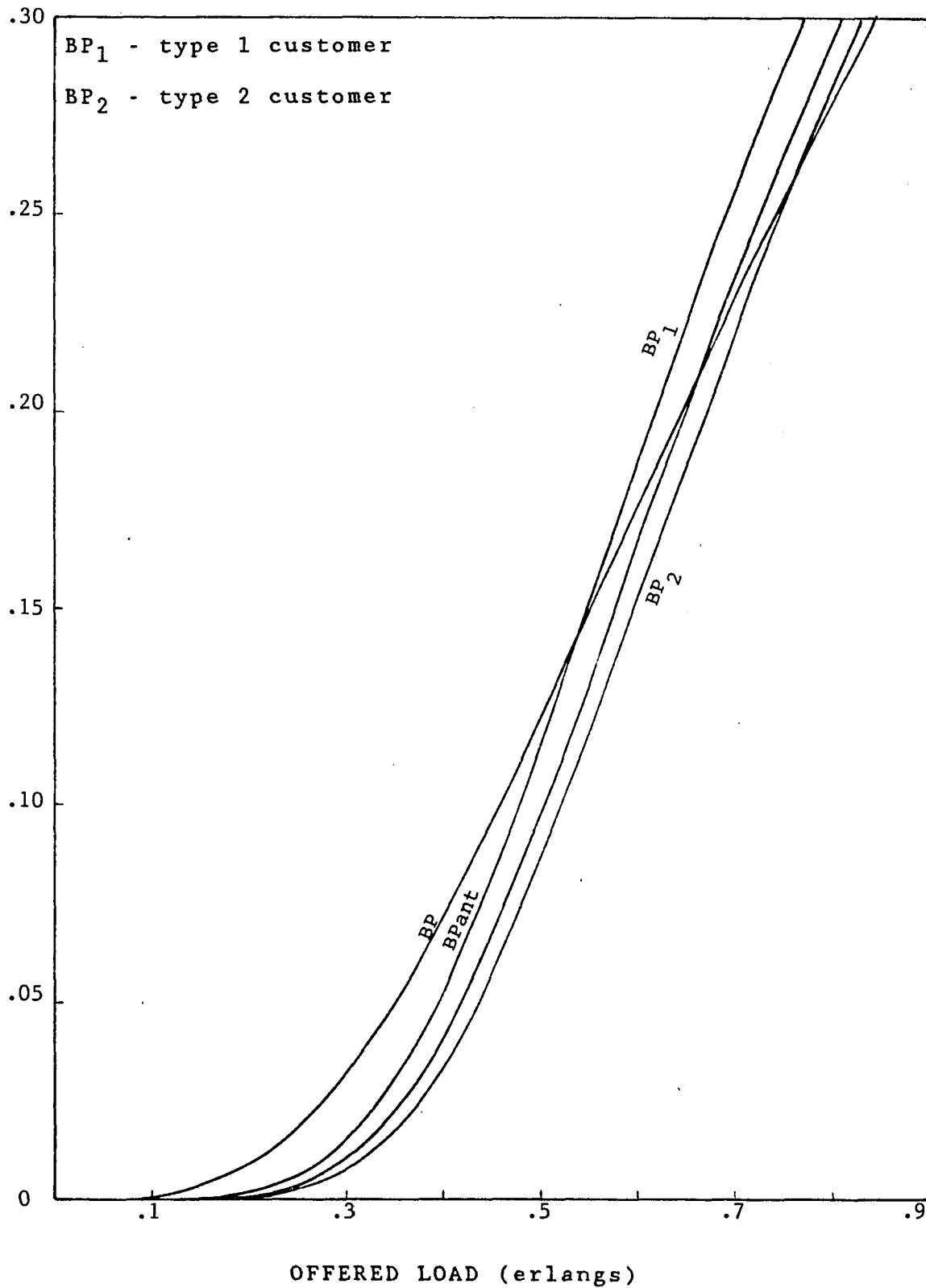


Fig.7 End-to-End Blocking Probability vs. a_2
(non-uniform trunk reservation with $t_1=2$ and $t_2=3$)

TABLE I: End-to-End Blocking Probability

| a ₂ | non-uniform trunk reservation* (10 ⁻²) | | | uniform trunk reservation** (10 ⁻²) | non-alternate routing (10 ⁻²) |
|----------------|---|-----------------|-------------------|---|--|
| | BP ₁ | BP ₂ | BP _{ant} | BP ₁ =BP ₂ =BP _{ant} | |
| .10 | .0012 | .0001 | .0005 | .0004 | .032 |
| .15 | .0620 | .0124 | .0289 | .0467 | .461 |
| .20 | .750 | .274 | .433 | .756 | 2.17 |
| .25 | 4.21 | 2.45 | 3.03 | 3.95 | 5.64 |
| .30 | 11.87 | 8.96 | 9.93 | 9.75 | 10.54 |
| .35 | 20.51 | 17.31 | 18.38 | 16.34 | 16.14 |

Table I. Comparison of non-uniform and uniform trunk reservation schemes (n₁=500, n₂=20, and c=10)

* t₁=4, t₂=7, and s=1

** t=8 and s=1

TABLE II: End-to-End Blocking Probability

| a ₂ | non-uniform trunk reservation* (10 ⁻²) | | | uniform trunk reservation** (10 ⁻²) | non-alternate routing (10 ⁻²) |
|----------------|---|-----------------|-------------------|---|--|
| | BP ₁ | BP ₂ | BP _{ant} | BP ₁ =BP ₂ =BP _{ant} | |
| .10 | .0039 | .00002 | .0011 | .0001 | .0330 |
| .15 | .374 | .0210 | .130 | .0618 | .907 |
| .175 | 1.57 | .207 | .661 | .532 | 2.40 |
| .20 | 5.22 | 1.36 | 2.66 | 2.62 | 4.84 |
| .225 | 12.24 | 5.25 | 7.58 | 7.08 | 8.12 |
| .25 | 20.08 | 11.32 | 14.24 | 12.51 | 11.95 |
| .275 | 26.91 | 17.58 | 20.70 | 17.87 | 16.04 |

Table II. Comparison of non-uniform and uniform trunk reservation schemes (n₁=1000, n₂=40, and c=15)

* t₁=5, t₂=15, and s=1

** t=14 and s=1

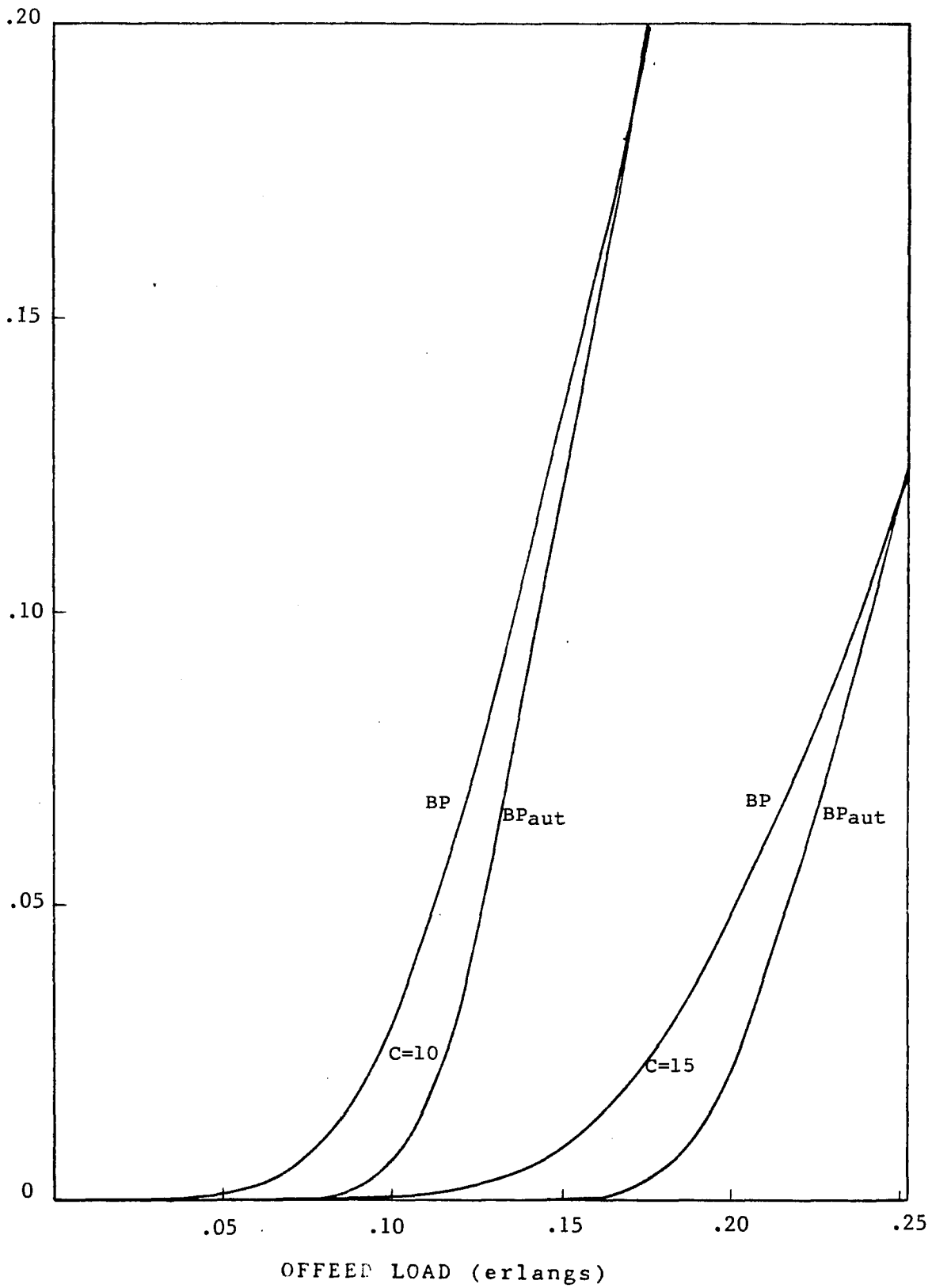


Fig.8 End-to-End Blocking Probability vs. a_2 for c=10 and 15

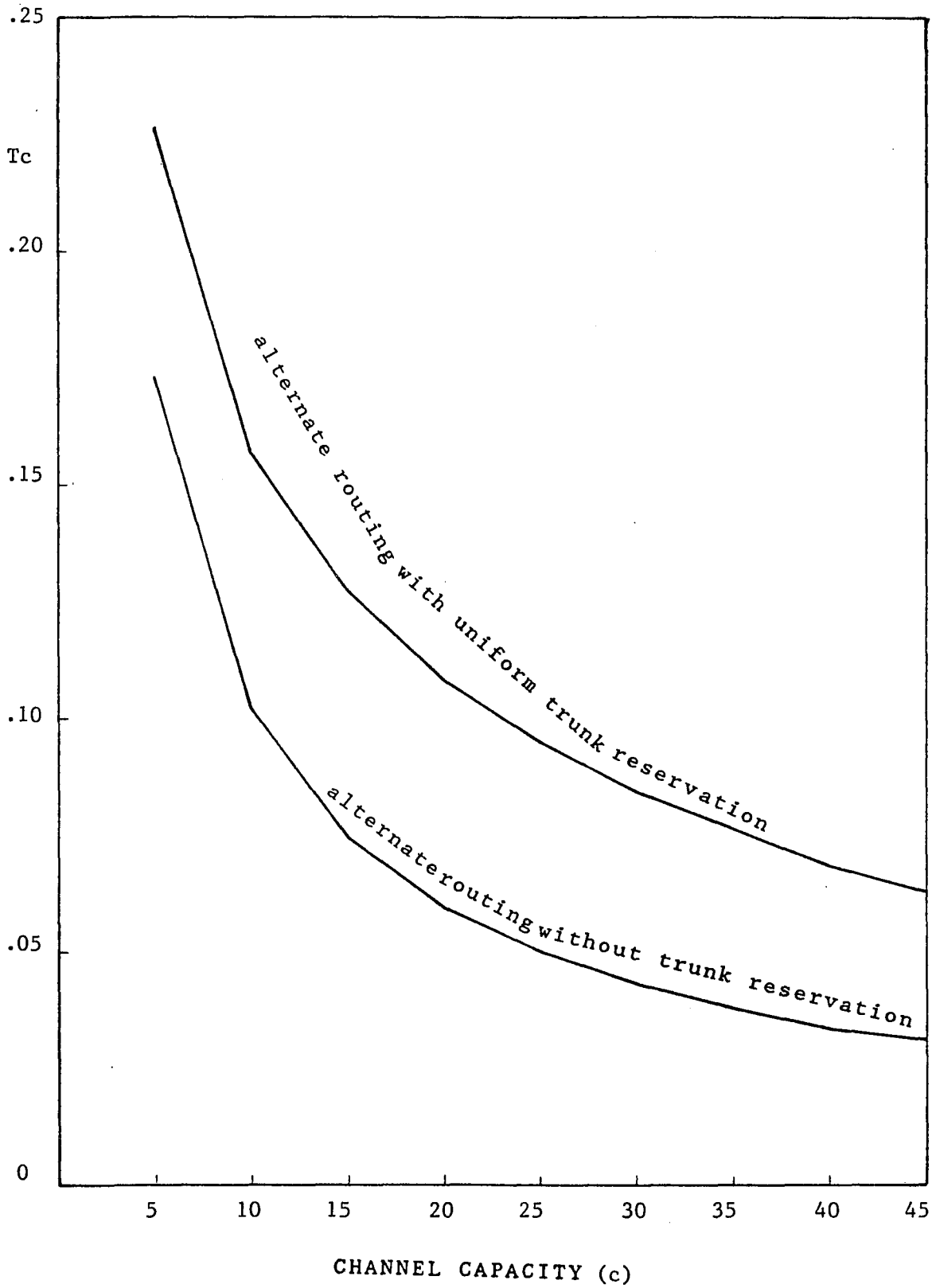


Fig.9 Effect of c on Alternate Routing

Table III. Network parameters used in Fig.9

| c | n ₁ | n ₂ | t |
|----|----------------|----------------|----|
| 5 | 300 | 10 | 4 |
| 10 | 600 | 20 | 8 |
| 15 | 900 | 30 | 12 |
| 20 | 1200 | 40 | 16 |
| 25 | 1500 | 50 | 20 |
| 30 | 1800 | 60 | 24 |
| 35 | 2100 | 70 | 28 |
| 40 | 2400 | 80 | 32 |
| 45 | 2700 | 90 | 36 |

APPENDIX

Let $BP_1(k)$ be the probability that an arriving customer in the (1)th link with k channels ($1 \leq k \leq c$) is blocked. Using (11) and (14), $BP_1(k)$ is deduced:

$$BP_1(k) = \frac{\sum_{i=0}^k a_1^{k-i} a_2^i \binom{n_1(l)}{k-i} \binom{n_2(l)}{i}}{\sum_{i=0}^k a_1^i \binom{n_1(l)}{i} \sum_{j=0}^{k-i} a_2^j \binom{n_2(l)}{j}} = \frac{G_1(k)}{H_1(k)} \quad (A-1)$$

Rewriting $H_1(k)$, we get

$$\begin{aligned} H_1(k) &= \sum_{i=0}^{k-1} a_1^i \binom{n_1(l)}{i} \sum_{j=0}^{k-i-1} a_2^j \binom{n_2(l)}{j} + a_2^{k-i} \binom{n_2(l)}{k-i} + a_1^k \binom{n_1(l)}{k} \\ &= H_1(k-1) + G_1(k) \end{aligned} \quad (A-2)$$

Substituting (A-2) into (A-1) and dividing both the numerator and denominator of the resulting equation by $G_1(k)$, we get

$$BP_1(k) = \frac{1}{1 + \frac{H_1(k-1)}{G_1(k)}} \quad (A-3)$$

The second term in the denominator can be written as

$$\frac{H_1(k-1)}{G_1(k)} = \frac{H_1(k-1)}{G_1(k-1)} \times \frac{G_1(k-1)}{G_1(k)} = \frac{1}{BP_1(k-1)} \times \frac{G_1(k-1)}{G_1(k)} \quad (A-4)$$

Finally, the recursive formula given by (16) is obtained by substituting (A-4) into (A-3).

REFERENCES

1. T. Saadawi and M. Schwartz, "Distributed Switching for Data Transmission over Two-Way CATV," IEEE Journal on Selected Areas on Commun., March 1985.
2. R.I. Wilkinson, "Theories for Toll Traffic Engineering in the U.S.A.," B.S.T.J., 35, No.2, pp. 421-514, March 1956.
3. J.J. Pilliod, "Fundamental Plans for Toll Telephone Plan," B.S.T.J., 31, pp. 832-850, September 1952.
4. J.S. Turner, "Design of an Integrated Services Packet Network," submitted for publication.
5. G.R. Ash, R.H. Cardwell, and R.P. Murray, "Design and Optimization of Networks With Dynamic Routing," B.S.T.J., 60, No.8, pp. 1787-1820, October 1981.
6. J.M. Akinpelu, "The Overload Performance of Engineered Networks With Nonhierarchical and Hierarchical Routing," B.S.T.J., 63, No.7, pp. 1261-1281, September 1984.
7. R.S. Krupp, "Stabilization of Alternate Routing Networks," IEEE Int. Commun. Conf., No. 3I.2, Philadelphia, 1982.
8. T.-K. G. Yum and M. Schwartz, "Comparison of Routing Procedures for Circuit-Switched Traffic in Nonhierarchical Networks," IEEE Trans. Commun., Vol. COM-35, pp. 535-544, May 1987.
9. T. Saadawi, N. Jain, and M. Schwartz, "Protocol for a Distributed Switching Interactive CATV Network," IEEE INFOCOM'86, April 1986, Miami, FL.
10. M. Nesenbergs, "Three Typical Blocking Aspects of Access Area Teletraffic," IEEE Trans. Commun., Vol. COM-28, pp. 1662-1667, September 1980.
11. W.S. Chan, "Recursive Algorithms for Computing End-To-End Blocking in a Network with Arbitrary Routing Plan," IEEE Trans. Commun., Vol. COM-28, February 1980.
12. J.S. Kaufman, "Blocking in a Shared Resource Environment," IEEE Trans. Commun., Vol. COM-29, pp. 1474-1481, October 1956.

- 13 A. Girard, "Blocking Probability of Noninteger Groups with Trunk Reservation," IEEE trans. Commun., Vol. COM-33, February 1984.
- 14 J. Riordan, "Stochastic Service Systems," John Wiley and Sons, New York, 1969.
- 15 M. Gerla, "Deterministic and Adaptive Routing Policies in Packet-Switched Computer Networks," Proc. 3rd IEEE Data Commun. Symp., pp. 23-28, 1973.
- 16 A. Livne, "Dynamic Routing in Computer Communication Networks," Ph.d Dissertation, Polytechnic Univ., 1977.
- 17 P.M. Lin, B.J. Leon, and C.R. Stewart, "Analysis of Circuit-Switched Networks Employing Originating-Office Control with Spill-Forward," IEEE Trans. Commun., Vol. COM-26, pp. 754-765, June 1978.
- 18 B. Gopinath, J.M. Garcia, and P. Varaiya, "Blocking Probability in a Switching Center With Arbitrary Routing Policy," B.S.T.J., Vol. 63, No. 5, pp. 709-720, May-June 1984.
- 19 R.L. Franks and R.W. Rishel, "Overloaded Model of Telephone Operation," B.S.T.J., Vol. 52, No. 9, pp. 1589-1615, November 1973.