

STATE - DEPENDENT ROUTING ~~RULES~~ STRATEGIES

IN TELEPHONE NETWORKS:

A PRELIMINARY STUDY.

SAMUEL ERIDADI KIGUNDU

JUNE 1971

A thesis submitted for the degree of Doctor of Philosophy in the
Faculty of Engineering, University of London.

Department of Electrical Engineering,
Imperial College of Science and Technology,
University of London.

Summary

In this thesis we study the performance of network state-dependent routing rules. Our motivation is to maximize the traffic carrying capacity of any given telephone network for any offered traffic pattern, i.e. to give "the best" overall grade of service.

There are two other considerations. Firstly, I am keenly aware that the optimization of the overall (network) grade of service tends to be accompanied by a debasement in the service given to the smaller parcels of traffic. I have endeavoured to prevent this debasement. Secondly, alternate routing schemes usually give a very poor service when the network is subjected to an overload. I have therefore concentrated on those routing schemes that always give a performance that is at least as good as that given when no alternate routing is used.

The results show that we can carry up to about six per cent more traffic using the state-dependent rules than we can carry without alternate routing. We have succeeded in providing adequate protection to all parcels of traffic, and in realising a consistently good overload performance. Another important and exciting outcome of this study is the concept of "time to congestion" which has played a central role in our routing schemes.

I believe that an improvement in the traffic carrying capacity of six per cent is, by itself, too small to justify the investment necessary to implement the state-dependent routing rules. I say "by itself" because the use of such rules may also have the advantage of enhancing the network security. Nevertheless, the chief recommendation is that this work should be continued and extended; this thesis is only a preliminary study. We would, for example, like to examine the performance of networks having more than half a dozen exchanges. And we look forward to better state-dependent routing rules being devised.

TO

MY WIFE GLADYS

A C K N O W L E D G E M E N T S

The author is deeply indebted to his supervisor, Professor E. C. Cherry, for his sustained enthusiasm, help and advice during the course of his work.

Thanks are due to Mr. R. E. Bogner and Dr. L. Turner of the Imperial College, and Mr. A. C. Cole of the British Post Office for extending their help at various critical stages of the work. I would also like to acknowledge the criticism and help that has come through the innumerable discussions I have had with my student colleagues at the Imperial College.

The unfailing encouragement and moral support from my wife, especially at those times when it seemed that the task I had set myself was practically impossible, is deeply acknowledged. My thanks also go to Miss Barbara Robertson for typing the thesis.

Finally, I would like to thank the Uganda Government without whose financial assistance this work could not have been carried out.

CONTENTS

		Page
	Title page	1
	Summary	2
	Dedication	3
	Acknowledgements	4
	Contents	5
CHAPTER	TITLE	
1.	INTRODUCTION TO THE ROUTING PROBLEM.	10
1.1	Introduction.	11
1.2	The case for a study of advanced routing systems.	11
1.2.1	Routing and advanced signalling systems.	11
1.2.2	Routing and stored program controlled exchanges.	13
1.2.3	Routing and the changing network structure.	13
1.3	The routing problem.	13
1.3.1	The construction of routing rules.	15
1.3.2	The evaluation of routing schemes.	15
1.4	Routing experience in Denmark: a motivation example.	16
1.5	The scheme of this thesis.	18
2.	AN INTRODUCTION TO TELEPHONE SYSTEMS.	20
2.1	Introduction.	21
2.2	Telephone Network Structures.	22
2.3	Some factors affecting the Routing Plan in a practical network.	23
2.3.1	The Transmission Plan.	23
2.3.2	The Switching and Signalling Plans.	24
2.3.3	The Numbering plan.	24
2.3.4	The Network Structure.	25
2.4	Routing in a Hierarchical Network.	25
2.5	Automatic (Sequential) Alternate Routing.	25
2.6	Automatic Dynamic Alternate Routing.	27
3.	ELEMENTS OF TELEPHONE TRAFFIC THEORY.	28
3.1	Introduction.	29

3.2	The nature of telephone traffic.	29
3.3	The units of telephone traffic.	30
3.4	The random arrival of calls.	31
3.4.1	The Poisson process.	31
3.5	The non-random arrival of calls.	33
3.5.1	The smoothness of telephone traffic.	33
3.5.2	Smooth traffic; the Binomial distribution	34
3.5.3	Smooth traffic: the Erlang distribution	34
3.5.4	Rough traffic: the negative Binomial distribution.	35
3.5.5	Rough traffic: the hyper-exponential distribution	37
3.6	Duration of calls: the negative exponential distribution.	38
3.7	Loss formulae.	40
3.7.1	The Erlang loss formula.	41
3.7.2.	The Engset (Erlang's Bernoulli) formula.	44
3.8	Overflow traffic.	47
3.8.1	The Equivalent random theory.	47
4.	ROUTING IN A SWITCH	49
4.1	Introduction.	50
4.2	The 6 to 4 access concentrator problem.	50
4.2.1	State assignment	51
4.2.2	Probabilistic assumptions and network operation	57
4.2.3	Mathematical formulation of the problem.	58
4.2.4	The statistical equilibrium equations for the general state x .	61
4.2.5	Solution of the problem.	62
4.3	The Clos network of switches.	64
4.4	Conclusion	67
5.	SIMULATION OF A TELEPHONE NETWORK	68
5.1	Introduction	69
5.2	The need for models	69
5.3	Advantages of simulation.	71
5.4	Roulette and Event-by-event simulation.	71
5.5	Computer simulation programming languages.	73
5.6	The generation of uniformly distributed random numbers.	73

5.7	Non-uniform continuous random numbers	75
5.7.1	The exponential distribution	76
5.8	The computer simulation of a telephone network	77
5.9	The verification of simulation results.	78
6.	ROUTING IN A NETWORK OF EXCHANGES AND RELATED PROBLEMS	79
6.1	Introduction.	80
6.2	Routing in a network of exchanges	80
6.3	Some problems that are similar to the routing problem.	83
6.3.1	Network flow problems.	84
6.3.2	The inventory problem.	86
6.3.3	Markov processes with rewards.	86
7.	ROUTING IN A FOUR-EXCHANGE NETWORK	88
7.1	Introduction	89
7.2	The direct routing strategy.	89
7.3.	Alternative routing strategies.	90
7.3.1	The three-exchange network.	91
7.4	The four-exchange network.	94
7.4.1	Dynamic routing strategies	96
7.4.2	The hyper-diagonal routing strategy.	96
7.4.3	The min-max routing strategy.	97
7.4.4	The random routing strategy.	102
7.5	Conclusion.	103
8.	STOCHASTIC BEHAVIOUR OF A GROUP OF CIRCUITS : THE TIME TO CONGESTION CONCEPT	104
8.1	Introduction	105
8.2	Transition probabilities in the occupancy of a group of trunks.	106
8.3	The computation of the stationary probabilities.	107
8.4	The fundamental matrix	109
8.5	The first passage time	110
8.5.1	The mean first passage time	111
8.5.2	The variance of the first passage time	112
8.5.3	The probability distribution of the first passage time.	112
8.6	The stochastic behaviour of a group of circuits offered non-random traffic.	113
8.6.1	The transition matrix for smooth traffic	113
8.6.2	The transition matrix for rough traffic	114
8.7	The first passage time distribution in continuous time	116

8.8	The probability density function of $Z = X - Y$	118
8.8.1	The probability of a routing error	119
8.9	Conclusion	119
9	ROUTING STRATEGIES USING THE MEAN TIME TO THE CONGESTION STATE	121
9.1	Introduction	122
9.2	The four-exchange network	122
9.3	The non-symmetrical four-exchange network	128
9.3.1	The protection of the point-to-point grades of service	130
9.3.2	The effect of the variance to mean ratio on the mean time to congestion	133
9.4	A non-symmetrical network with large trunk groups	135
9.5	Conclusion	136
10	ROUTING RULES DEPENDING ON THE DISTRIBUTION OF THE TIME TO CONGESTION	137
10.1	Introduction	138
10.2	The probability that the time to congestion exceeds unity	138
10.3	A comparison of routing strategies used in the four-exchange network	141
10.3.1	Routing strategies to be compared	141
10.3.2	Comparison of overall (network) grade of service	143
10.3.3	Comparison of the point-to-point grades of service	146
10.4	Conclusions	146
10.5	Appendix : Tables of the point-to-point and overall grades of service for the network of Figs. 10.3a and 10.3b.	149
11.	THE SIX-EXCHANGE NETWORK ; FURTHER ROUTING TECHNIQUES	153
11.1	Introduction	154
11.2	Routing in the fully connected six-exchange network.	154
11.2.1	Performance comparisons	157
11.2.2	Overall grades of service	157
11.2.3	Point-to-point grades of service.	157
11.3	Networks in which some exchanges have no direct links to other exchanges	159

11.3.1	Exchange and route partitioning.	160
11.3.2	Exchange lumping.	161
11.3.3	Exchange lumping : an example.	162
12.	CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH.	165
12.1	Conclusions.	166
12.2	Some aspects of the routing problem needing further research.	166
12.2.1	Data processing requirements.	166
12.2.2	Optimal routing rules.	167
12.2.3	Analytical evaluation of network performance.	167
12.2.4	State-dependent rules in a hierarchical network.	168
12.2.5	Routing with imperfect information.	168
12.2.6	Economic considerations.	168
12.2.7	Reduction in network status data.	169
12.2.8	Data and message traffic.	169
	REFERENCES	171

CHAPTER 1 INTRODUCTION TO THE ROUTING PROBLEM

- 1.1 Introduction
- 1.2 The case for a study of advanced routing systems.
 - 1.2.1 Routing and advanced signalling systems.
 - 1.2.2 Routing and stored program controlled exchanges.
 - 1.2.3 Routing and the changing network structure.
- 1.3 The routing problem
 - 1.3.1 The construction of routing rules.
 - 1.3.2 The evaluation of routing schemes.
- 1.4 Routing experience in Denmark: a motivation example.
- 1.5 The scheme of this thesis.

CHAPTER 1 INTRODUCTION TO THE ROUTING PROBLEM.

1.1 Introduction

The goodness of a telephone service is judged, amongst other things, on the proportion of calls that encounter a "busy" tone even if the called subscriber is not using his telephone. Economic considerations make it impossible to dimension the traffic routes so as to guarantee the handling of traffic without any blocking. It is, therefore, vital that we should be able to use the available routes as efficiently as possible. This effectively means that we must continually re-appraise our routing methods in search of better performance.

Routing studies carried out by telephone administrations are, rightly, geared to their immediate routing problems. But it is important that we should also keep an eye on the near and the distant future. In this thesis, we address ourselves to the distant future. We are trying to understand and evaluate the performance of networks, that are not necessarily hierarchical, in which the routing rules are continually updated and matched to the state of the network.

1.2 The case for a study of advanced routing systems.

1.2.1 Routing and advanced signalling systems.

One of the main reasons that have prevented the application of complex routing plans in the commercial network is the lack of signalling facilities that can support such routing schemes. Another main reason, of course, is cost. So what are these qualities that are required of such signalling systems? I consider the following four qualities to be necessary:

1. a fast signalling speed,
2. a large signal set,
3. a means of conveying network-management signals at a reasonable cost,
4. high reliability and very low error rates.

These qualities are, at the moment, only approached by the C.C.I.T.T. No 6 Separate-channel data signalling system, Crew (141). Although originally intended for use in international signalling, its great potential has been recognized, and modifications have been made to it so

that it can be economically applied to national networks. Let us pause, therefore, to look at this signalling system.

The diagram of Fig. 1.1 shows how a separate channel signalling system could be implemented. At each exchange, a processor is required to interpret the data stream, control the switch block, and if a transit exchange, forward the necessary signals to the next exchange, or deal with the subscribers' lines if it is a terminal exchange. The processor also has the duty of requesting re-transmission

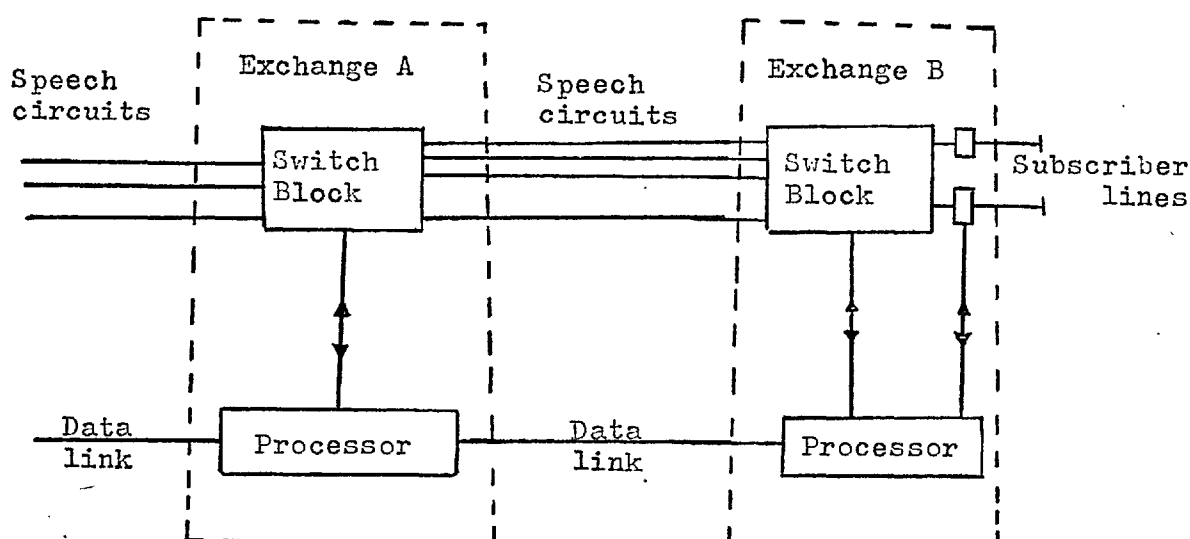


Fig. 1.1 A separate-channel signalling scheme

in the case of parity errors. It is mainly in the error control mechanism that the modified system for national networks realises most of its savings. But because it is unlikely that this method of signalling can be applied economically to the small auxiliary routes, it has been proposed that data-link signals relating to traffic on such routes should be transmitted indirectly via one or more other switching centres so chosen that the data links to be used can be loaded economically.

The complexity of the processors in Fig. 1.1 will depend on the level in the hierarchy that the associated exchange enjoys. For local exchanges, the processor could be very rudimentary. The local exchange would, in fact, be effectively remote processor controlled by its group switching centre, the control being exercised over secured data links.

1.2.2 Routing and stored program controlled exchanges

The advent of electronic exchanges, and stored program controlled (SPC) exchanges in particular, just like the new signalling system, makes the idea of using complex routing plans more meaningful. There are many electronic and quasi-electronic exchanges on the market, and many more in various stages of development. Large SPC exchanges are already in service. Perhaps the most well known is the Bell System's ESS No. 1. Others include L.M. Ericsson's AKE 12 and AKE 13. For the smaller exchanges, of say 1000 lines, the ESS No. 2 has been designed and field tested in the Bell System. Northern Electric's SP-1, and L.M. Ericsson's AKE 11 are also available for the small exchanges. All these SPC exchanges can cope with complex routing schemes quite easily.

1.2.3 Routing and the changing network structure.

There is a tendency for telephone network structures to depart further and further away from the strictly hierarchical arrangement. This trend is likely to continue as high usage routes grow bigger and bigger. We shall then see a highly connected telephone network. Already vigorous discussion is taking place on what constitutes a good structure, Dejean et al (73). Routing plans that have been designed with the hierarchical structure in mind may soon have to be drastically changed.

1.3 The routing problem

This then is a very suitable time to re-examine our routing philosophy. In this thesis, I study and evaluate the performance of small networks using state-dependent routing rules. My model of a telephone network is highly simplified; I view the network as a topological graph with nodes (exchanges) and capacitated links (trunk groups). I am satisfied, though, that this skeleton model is adequate for this preliminary study. As this work is extended, some flesh will have to be added to the skeleton; the limitations of the exchange switching equipment, signalling systems and transmission plans will have to be reflected in the model.

Studies have already been carried out on the optimal routing of telephone calls in small networks of switches, Benes (24, 25) and of exchanges, Grandjean (1), Wilkinson (9) and Weber (30, 31). Such studies have highlighted problems that arise in the use of alternate routing.

The very poor performance of these alternate routing systems under heavy loads has been of great concern. The most generally recommended practice has been to cancel completely any alternate routing if the loading of the network, or part of it, exceeds a certain predetermined level. This measure certainly gives an overall service improvement, but it also leads to certain parcels of traffic being almost completely isolated. Traffic parcels designed to use high efficiency (high usage) routes with facilities for overflowing on to more generously dimensioned routes are very seriously affected. A better method of protecting the network against serious service deterioration is to make a selective cancellation of alternate routes in conjunction with trunk reservation.

With state-dependent routing schemes, the traffic imbalances in the network are smoothly and quickly dealt with, while at the same time maintaining a uniformly good performance even at the highest overloads. Excess capacity in some parts of the network can be used to take up some of the demand in the overloaded sections. In fact, the methods that have been used to take advantage of the different time zones in the international routing of calls, or in countries like the U.S.A. which span several time zones, are now effectively proposed although the networks envisaged in this thesis do not necessarily span more than one time zone. The results obtained when compared with those obtained using different routing rules show that the network's traffic carrying capacity is certainly enhanced.

The small trunk groups that usually suffer from the traffic overflowing on to them from the bigger groups are now protected by a system of automatic selective cancellation of alternate routing. This in turn means that the network need not suffer a catastrophic degradation of service when it is heavily loaded.

The routing schemes that I have constructed and evaluated have not been proved to be optimal. Nevertheless, the methods of study have produced some decision variables that could be of immense value in the construction of even better routing schemes. Perhaps, even optimal routing rules could be constructed using these variables. Of particular importance is the random variable that represents the "time to congestion", and its probability distribution. These quantities are reasonably easy to compute as we shall show in Chapter 8.

Routing problems divide up naturally into three parts. The first part consists of setting down the requirements that the routing rules should satisfy. Then comes the actual process of devising the routing rules. Finally, the performance of the network under those routing rules must be estimated.

1.3.1 The construction of routing rules

In this thesis, we are looking for routing rules that give a high circuit utility (erlangs per circuit) for all groups of circuits in the network at all levels of traffic. We are satisfied with the performance of the present alternate routing systems at engineered loads, but we are disappointed in their overload performance. We have therefore turned to state-dependent rules to try and rectify this.

We have found no easy or automatic way of constructing routing rules for use in a network of exchanges. In Chapter 6, we have drawn extensively from the work of Dr. V. E. Benes of the Bell Laboratories Inc. U.S.A. to show that some routing rules for networks of switches can be constructed and shown to be optimal using linear programming. But computational difficulties make it almost impossible to apply these methods in a network of exchanges.

The routing rules that we have applied in the earlier parts of this study have been chosen mainly because of their commonsense and intuitive appeal. After evaluating the performance of some networks using these rules, we found that we had an inadequate understanding of the stochastic behaviour of a group of circuits. In Chapter 8 we make a detailed study of this behaviour, and in so doing we develop the "time to congestion" concept which plays a central role in the routing rules which we subsequently develop.

If we are given a set of routing rules, the question of which one to choose arises naturally. In other words, we must evaluate the performance of the network for each of the routing rules.

1.3.2 The evaluation of routing schemes

An analytical treatment of the statistical performance of networks is usually prohibitively difficult. For simple routing rules, the Lee-Le Gall technique, Syski (22) and Grandjean (1), can be used to obtain some approximate results. This technique requires that we assume that the

occupation of any link is independent of the occupations of all other links in the network. This assumption we have to make in a study that relies heavily on alternate routing over two or more links in tandem. Yet unless we make it, we cannot proceed because of the very complex interactions in the network.

Katz (98,99) has, in fact, worked out an algorithm for the statistical performance analysis of a switched alternate routing network using the (independence) assumption that the trunk groups throughout the network are statistically independent of each other. It has been claimed that this algorithm gives results that compare very favourably with those obtained by simulation, and which are obtained in only a fraction of the (computer run) time. The algorithm is suitable only for networks using stage-by-stage (progressive) route control whereby a call is transmitted from one exchange to the next one without any guarantee that a successful path will be obtained through the distant exchange. While this is the method employed by most, if not all, commercial networks, the routing rules studied in this thesis depend on right-through rather than progressive routing. This effectively leaves simulation as the only way of evaluating the network performance under a given routing rule.

Because of the difficulty of analytical evaluation, call-by-call simulation techniques on a computer have been predominantly used to perform steady-state and transient analyses of telephone networks. At the Bell Laboratories, U.S.A., a general telephone network simulator UNISIM has been written for the IBM 7094, Weber and Gimpelson (84).

All the performance evaluations carried out in this study were made by computer simulation using either the GPSS III or the SIMSCRIPT programming languages. Because of the enormous amount of time needed to write and debug computer programs, only a limited range of networks were considered.

1.4 Routing experience in Denmark : a motivation example

Few people will argue that a network's utility cannot be improved by using a more appropriate routing plan. But one suspects that even fewer realise just how much difference a good routing plan can make to the service offered to the public. An example from the Danish telephone network, Asgersen (151), may deepen our motivation in making a closer study of telephone traffic routing.

In 1953 subscriber trunk dialling was introduced in Denmark, and within the next ten years more than half of all long distance calls were subscriber dialled. About 1960 it became clear that the service offered to the long distance traffic was quite unsatisfactory; operator originated test calls showed the unacceptably high grade of service of about 0.25. Further investigation revealed that this high level of blocking could not be explained away by overloaded trunk groups or exchanges. The highest blocking values were, as expected, found for calls destined to, or originated by small exchanges to which no high usage trunk groups existed so that the calls had to follow the backbone route. A full investigation of the influence of the traffic routing plan on the quality of the traffic handling was called for.

The new Danish traffic routing plan, Asgersen (151,102), was the result of the investigation. The tremendous difference made by this new routing plan is shown in Fig. 1.2 re-drawn from Asgersen (151).

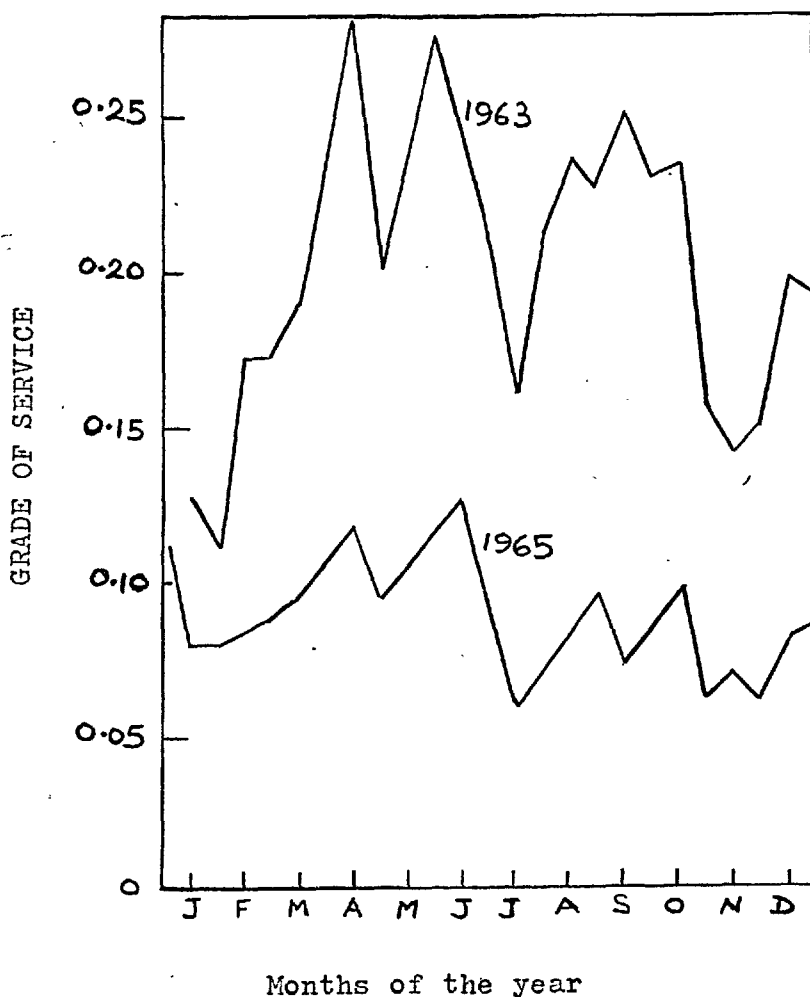


Fig. 1.2 Grade of service in the busy hour

1.5 The scheme of this thesis

This chapter has introduced and justified the routing problem. Chapter 2 is a very brief introduction to telephone systems. It is written especially for the reader whose speciality is not in telephone systems, so that he has a framework and a perspective as he reads the following chapters which analyse the routing problem using a topological graph as a model.

One can hardly discuss the routing problem without frequent and liberal use of telephone traffic theory concepts and terminology. Chapter 3 is designed to introduce the reader to the elements and language of traffic theory. Those who need a deeper understanding of this theory should consult Elldin and Lind (12), and Syski (22).

In Chapter 4 we review and summarize the work which has been done on the routing of calls in a network of switches. We find that these methods cannot be easily adopted for use in the routing problem in a network of exchanges.

Because of the very complex interactions taking place in a telephone network, it is usually prohibitively difficult to obtain analytical performance results. We have, therefore, made extensive use of simulation techniques. Chapter 5 serves to introduce the reader to the simulation concepts and methods that have been used.

Chapter 6 is a brief review of the more important work that has been published on the routing of telephone calls in a network of exchanges. We also examine some problems that are similar to the routing problem. It is on the recommendations for further research explicit or implicit in these studies, and obvious gaps in our understanding of routing processes that we have based our own investigations.

In Chapter 7 we devise and evaluate various routing rules for the three- and four-exchange networks. The performance evaluations are carried out by computer simulation.

After reading Chapter 7, one is convinced that although we appreciate qualitatively the factors involved in the routing problem, we must pursue a quantitative approach. Chapter 8 fulfils this need. In that chapter we re-examine the stochastic behaviour of a single group of circuits offered random or non-random traffic from a routing point of view. It is from this examination that the important "time to congestion" concept emerges.

In Chapters 9 and 10 we use the mean time to congestion and the distribution of the time to congestion, respectively, in devising more routing rules and applying them to the four-exchange network.

We consider the six-exchange network in Chapter 11, and apply to it the routing rules developed in Chapter 10. We also take a brief look at further routing problems that we have not been able to explore more fully, and propose possible solutions or methods of attack.

In Chapter 12 we draw conclusions and make concrete recommendations for further research.

CHAPTER 2. AN INTRODUCTION TO TELEPHONE SYSTEMS

- 2.1 Introduction
- 2.2 Telephone Network Structures
- 2.3 Some factors affecting the Routing Plan in a practical network
 - 2.3.1 The Transmission Plan
 - 2.3.2 The Switching and Signalling Plans
 - 2.3.3 The Numbering Plan
 - 2.3.4 The Network Structure
- 2.4 Routing in a Hierarchical Network
- 2.5 Automatic (Sequential) Alternate Routing
- 2.6 Automatic Dynamic Alternate Routing

CHAPTER 2. AN INTRODUCTION TO TELEPHONE SYSTEMS

2.1 Introduction

The main objective of telephone traffic engineering is to formulate a set of procedures for the design and day-to-day management of a telephone system that will result in the most economic provision of telephone exchange switching equipment, trunk and junction circuits, subject to providing a satisfactory service at an acceptable cost to the users of the system.

Because of economic considerations, telephone systems are designed in such a way that there is a limit to the number of simultaneous calls they can handle. For most systems the limit is expressed in terms of "the grade of service", which is the probability that a call made in the busy hour will fail because of equipment congestion. This is the probability that a calling subscriber will get a "busy tone" signal.

An objective that practically all (national) telephone administrations are aiming at is complete subscriber trunk dialing (S.T.D.), and eventually International subscriber dialing (I.S.D.). This objective necessitates the use of common control methods for telephone switching so as to exploit the network to the maximum for routing traffic. This routing philosophy underlies the concept of network management which effectively means that the common control of an exchange should have data on the ability of other parts of the network to carry traffic, and thus, at any one time, the ability to pass a particular call. As the ability of the network to carry traffic will continually change, it is clear that the data must be continually updated.

This flexible routing scheme would have its greatest impact in international long distance systems, but the basic concepts could also be applied to national, and even local networks.

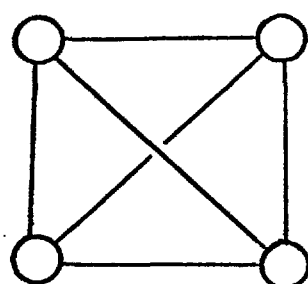
It is evident that rigid routing rules, such as those used in the hierarchical structure, prevent the network being exploited as a whole for routing traffic. The alternative, which allows full exploitation of the network, is to allow a particular call to be routed in any way consistent with transmission considerations, depending upon the free availability of circuits in the network as a whole. This alternative is possible with network management, and this thesis is the result of the introductory work done on trying to increase the traffic handling capacity of a network by using a flexible routing scheme.

But before plunging into the routing problem let us very briefly examine the telephone network structure and the routing schemes as they are currently applied.

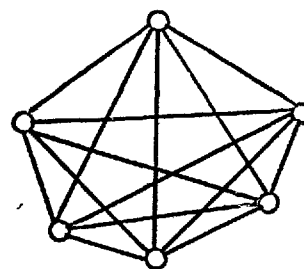
2.2 Telephone network Structures

In trying to understand the telephone network structures of present day networks, it is instructive to review network developments right from the beginning, Wernander (62).

If a network is made up of a small number of subscribers geographically near each other, it is best covered by a complete mesh, Fig. 2.1.



4 subscribers



6 subscribers

Fig. 2.1 The mesh network

Many intercom sets and small internal telephone services have a complete mesh structure. For n subscribers, it is readily seen that $n(n-1)/2$ direct connections are needed. If n is greater than 10 or 20, it becomes imperative to have a centralized switching exchange. This results in a star network as in Fig. 2.2

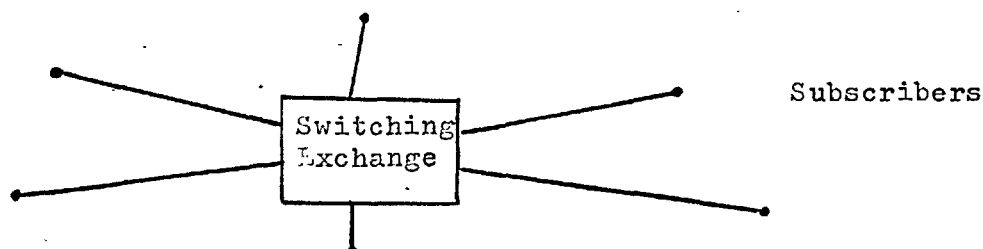


Fig. 2.2 A Star network

By a similar argument, the local switching exchanges are mesh connected by junction circuits until there are too many local exchanges. Then a higher order exchange, a Group Switching Centre (G.S.C.) is made

the centre of a star of local exchanges. This very simplified procedure underlies the way the hierarchical network of Fig. 2.3 has evolved.

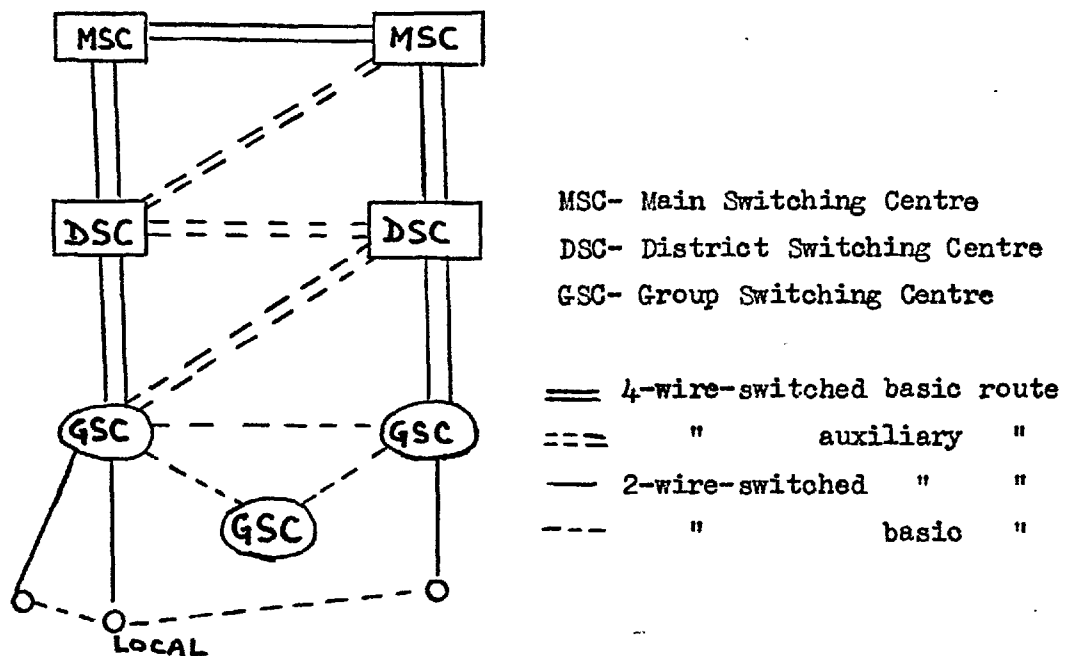


Fig. 2.3 The British Telephone Network

This hierarchical structure has evolved quite naturally in all networks of any complexity, with more or less levels of hierarchy, and varied nomenclature for the exchanges. For example, the U.S. and Italy have hierarchies of five levels, West Germany four levels, while Switzerland has only three.

The basic or backbone structure of the network provides an assurance that any telephone can be connected to any other telephone, and the auxiliary (direct) routes are permitted and provided where sufficient traffic exists to justify them. These auxiliary routes are quite often provided on a high-efficiency (high usage and high loss) basis, with the final or ultimate grade of service being given by the backbone network.

2.3 Some factors affecting the Routing Plan in a practical network.

The routing plan of a network is very closely tied up with the switching plan, the transmission plan, the actual structure of the network, and to some (ever decreasing) extent the numbering plan.

2.3.1 The Transmission Plan

The main aim of the transmission plan is to ensure that all subscribers can get a reasonably good quality connection. Good in the sense that the noise on the line, the cross-talk, the distortion (including echoes for long distance calls), and the actual transmission losses are kept to an acceptable level. Fig. 2.4 shows the transmission (loss) plan for the British (trunk transit) network.

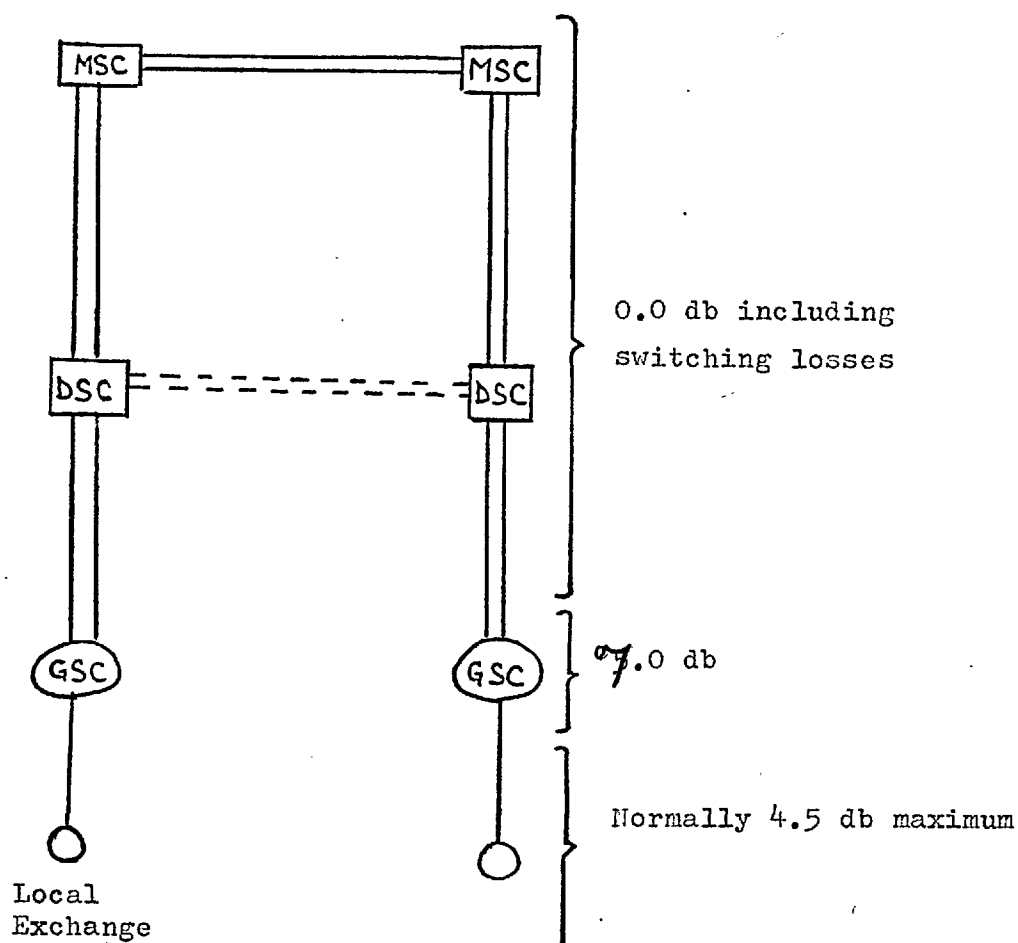


Fig. 2.4 Transmission plan for the trunk transit network.

Seen as a routing constraint, the transmission plan limits the number of links in tandem that a call can use in order that the maximum loss of 15 db may not be exceeded.

2.3.2 The Switching and Signalling Plans

The transmission losses due to switching have been mentioned in section 2.3.1, and will not be considered any further. But there is much to be said about switching delays. We shall content ourselves, however, by saying that the use of many links in tandem is undesirable because it leads to long post-dialling delays that may result in calls being abandoned. There are some purely technical reasons too. If the links are seized link-by-link, then circuits may be engaged for appreciably long times only to find that the last link to the required subscriber is congested.

2.3.3 The Numbering Plan

In step-by-step (usually Strowger) systems where the routing of a

call is under the direct control of the dialled digits, the routing plan cannot be divorced from the numbering plan. The use of Register-Translators provides the only way of separating the routing and numbering plans. This separation is very welcome from both the routing and numbering points of view.

2.3.4 The Network Structure

The provision of circuits between the different exchanges effectively determines the structure of the network. This dimensioning of the routes is made on the assumption, explicit or implicit, that the telephone traffic is going to be routed in a particular way. The routing plan is thus built into the network structure.

2.4 Routing in a Hierarchical Network

Consider a call from one local exchange A to another local exchange B, Fig. 2.5

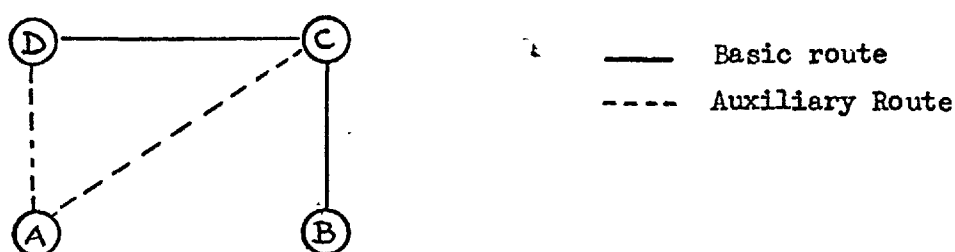


Fig. 2.5 Routing a local call

If, because of sufficiently high traffic between A and B, a direct junction route had been provided between A and B, then the call would be offered to the A-B junction route in the first place. If this route is either congested or non-existent, the call may be connected through the tandem exchange C. If this is not possible, the backbone route A-D-C-B is used.

2.5 Automatic (Sequential) Alternate Routing

In the Strowger step-by-step system, the routing of the telephone calls is directly dependent on the dialed digits. Even with the Director system which can give a fixed translation code, the basic situation is still the same. And this is that every call from a particular exchange to another exchange must pass over the same route whether it is direct, or it involves one or more tandems. Such a system demands that the number of circuits on each route should be carefully planned to provide the specified grade of service. If the traffic

forecasts are not sufficiently accurate, or if the traffic pattern has a tendency to shift, the specified grade of service cannot be realised.

In common control systems using automatic alternate routing, the testing of the routes is performed by the common control equipment. The exchange identification digits are translated by the Translator into routing digits, then the common control equipment tests the first-choice route. If this is congested, a second choice route is tested. This procedure continues until either a free route is found or the final route is also found to be congested in which case the call fails and busy tone is returned.

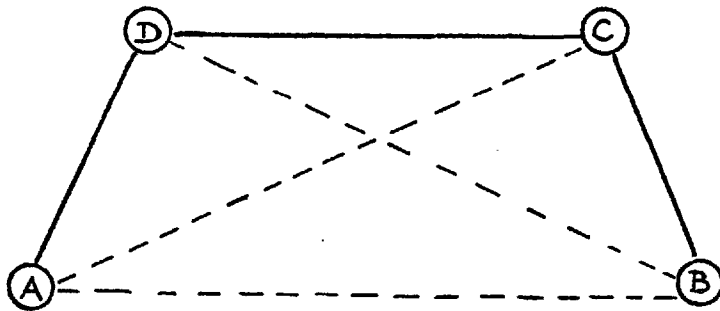


Fig. 2.6 An example to demonstrate alternate routing

Using the diagram of Fig. 2.6 we can draw a route tree which shows the sequential order in which the routes are searched for a call from exchange A to exchange B as in Fig. 2.7

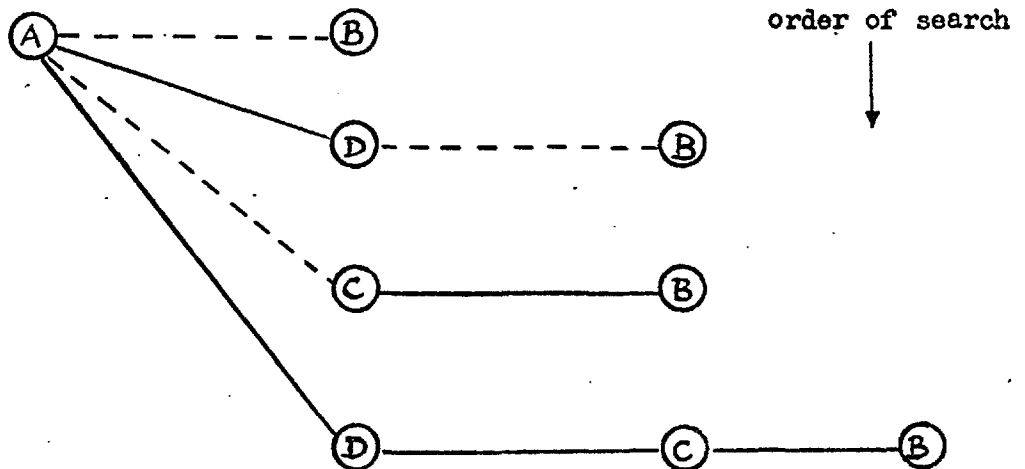


Fig. 2.7 The Route tree for a call from A to B

The British Post Office does not use automatic alternate routing even though it presently has available equipment for testing up to three routes (including the first choice route). Automatic alternate routing is, however, used in the U.S. and in Sweden.

2.6 Automatic Dynamic Alternate Routing

We have discussed automatic sequential alternate routing. This is a static routing plan that does not take into account the momentary distribution of traffic. Thus regardless of whether the traffic pattern has shifted or not, the same routes are searched in the same order. In automatic dynamic alternate routing, on the other hand, all the available alternate routes are considered together, and the route that is expected to lead to the least congestion is chosen, basing the decision on the instantaneous state of the network.

In this thesis, we examine various dynamic alternate routing strategies. But first, we must review the mathematics of telephone traffic in Chapter 3.

CHAPTER 3 - ELEMENTS OF TELEPHONE TRAFFIC THEORY

- 3.1 Introduction
- 3.2 The nature of telephone traffic
- 3.3 The units of telephone traffic
- 3.4 The random arrival of calls
 - 3.4.1 The Poisson process
- 3.5 The non-random arrival of calls
 - 3.5.1 The smoothness of telephone traffic
 - 3.5.2 Smooth traffic: the Binomial distribution
 - 3.5.3 Smooth traffic: the Erlang distribution
 - 3.5.4 Rough traffic: the negative Binomial distribution
 - 3.5.5 Rough traffic: the hyper-exponential distribution
- 3.6 Duration of calls: the negative exponential distribution
- 3.7 Loss formulae
 - 3.7.1 The Erlang loss formula
 - 3.7.2 The Engset (Erlang's Bernoulli) formula
- 3.8 Overflow traffic
 - 3.8.1 The Equivalent random theory

CHAPTER 3. ELEMENTS OF TELEPHONE TRAFFIC THEORY

3.1 Introduction

This chapter is written for the benefit of those readers who may not be familiar with the basic concepts of telephone traffic theory. As such, it contains no original work.

I have set out to introduce the reader to five basic aspects of telephone traffic theory:

1. The nature of telephone traffic.
2. The units in which telephone traffic is measured.
3. A mathematical description of telephone traffic that recognizes the differences between various kinds of traffic.
4. The loss formulae that give the probability of finding all circuits in a group engaged.
5. A description of overflow traffic.

These traffic theory concepts run through this thesis and provide the language in which the routing problem is most naturally formulated. We begin by examining the nature of telephone traffic.

3.2 The nature of telephone traffic

It is usually assumed in studies of telephone traffic that the telephone calls arise (collectively) by pure chance because there are so many subscribers that one can assume that their actions are uncorrelated by anything except the time of day. There are, however, situations which can influence a large number of subscribers to make calls; for instance, a national public holiday like Christmas, a local or national disaster, or any other occurrence of national interest. Since most switching systems are only designed to operate under average conditions, they will be unable to deal with such changes in the traffic pattern. Manual intervention is then necessary, for example to cut-off all but a limited number of telephones.

Although telephone calls usually originate from independent actions of individual subscribers, the telephone traffic so produced has definite statistical properties, like means and variances, that can be measured.

3.3 The units of telephone traffic

There are three major, inter-related factors that define telephone traffic:

- (a) The average number of calls N that are originated in some specified interval of time T .
- (b) The average holding (conversation) time t of these calls.
- (c) The average number of simultaneous calls A during the time interval T .

These factors are evidently related by

$$A = \frac{Nt}{T} \quad \dots\dots\dots 3.1$$

The following definitions have been adopted by the C.C.I.T.T. (Comite Consultatif International de Telegraphique et Telephonique)

- " The amount of traffic carried by a group of circuits, or a group of switches, during any period is the sum of the holding times expressed in hours."
- " The traffic flow on a group of circuits, or a group of switches, equals the amount of traffic carried, divided by the duration of the observation, provided that the period of observation and the holding times are expressed in the same units. The traffic flow is expressed in erlangs."

The erlang is a measure of the average number of simultaneous calls. From the relation 3.1 it is clear that the average traffic intensity in erlangs is numerically equal to the average number of calls which originate during the average holding time of a call. This equality is quite often used as the definition of an erlang.

In the United States, the ccs (hundred call seconds) is also used as a unit of traffic flow. Numerically

$$1 \text{ ccs} = \frac{1}{36} \text{ erlangs}$$

But for calculations it is generally necessary to express the traffic in erlangs.

The process describing telephone traffic is fully defined if one knows:

1. the times of arrival of calls,
2. the holding times of the successful calls,
3. the disposition of unsuccessful calls.

3.4 The random arrival of calls

In Section 3.3, it was assumed that we know the average number of calls N arriving during a given time interval T . This mean of the random process $\dot{x}(t)$ describing the arrival of calls, while sufficient for defining the traffic intensity, it is not enough for the more detailed calculations that are necessary in traffic theory. We must, therefore, postulate the probability density function which gives rise to this mean.

3.4.1 The Poisson Process

It turns out that if we use the Poisson process as our arrival process, we get a very good approximation to the arrival of calls in a real telephone network.

To use the Poisson process, and talk of Poissonian or Poisson traffic, we must make the following assumptions:

- (a) the number of subscribers is so large that assuming it to be infinite would cause a negligible error.
- (b) the traffic has not been modified by any switching arrangement leading to loss or delay. We are thus excluding smoothed and overflow traffic.
- (c) the individual calls are not related in any way.
- (d) calls may be originated at any moment during the period under consideration.

Then the probability of precisely n calls arriving during an observation period t is given by

$$\begin{aligned} \text{Prob (exactly } n \text{ calls in } t) &= p_n(t, \lambda) \\ &= \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad \begin{matrix} n = 0, 1, 2, \dots \\ t \geq 0 \end{matrix} \\ &= 0 \quad n < 0 \end{aligned}$$

The parameter λ is the calling rate, i.e. the average number of calls arriving per unit of time. It is convenient to assume, without any loss of generality, that the mean holding time is also the unit of time.

The average number of calls arriving during the period t is the mathematical expectation of $P_n(t, \lambda)$.

$$\sum_{n=0}^{\infty} n P_n(t, \lambda) = \lambda t \quad \dots \quad 3.2$$

By dividing by t in equation 3.2, we can verify that λ is the average number of calls arriving per unit time, and is, therefore, by definition the offered traffic.

The Poisson density function is illustrated in Fig. 3.1 for two values of λt .

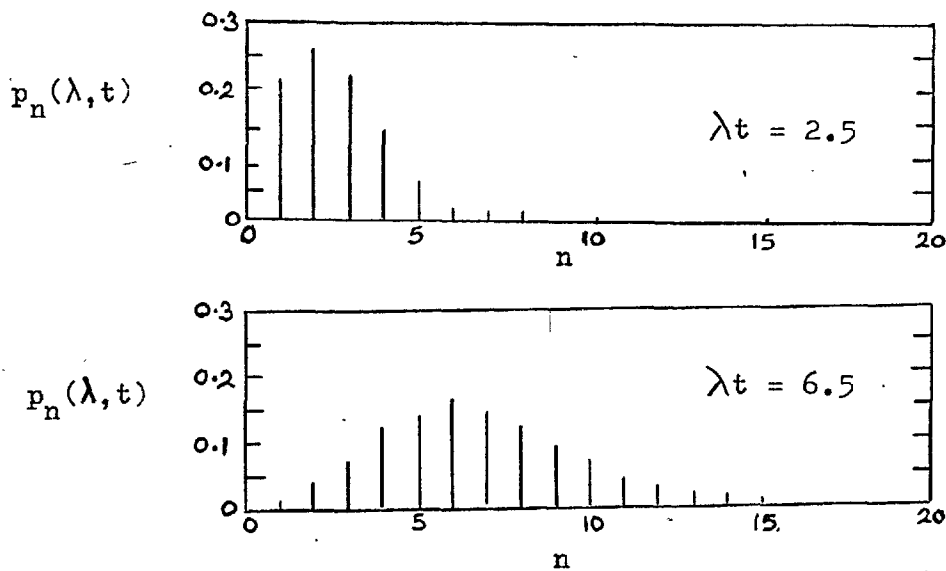


Fig.3.1 The Poisson density function for two values of λt .

It can be shown that for the Poisson process the mean is always equal to the variance; this is the only distribution for which this equality always holds.

The probability that no calls arrive during a time interval t is

$$p_0(\lambda, t) = e^{-\lambda t}$$

Regarding this as a complementary distribution function, the corresponding probability density function

$$p(t) = \lambda e^{-\lambda t}$$

is the negative exponential density function. We shall have more to say about this p.d.f. in Section 3.6 when we discuss the holding time distribution. It suffices to say that $p(t)$ is the density function of the inter-arrival interval, i.e. the time between two call arrivals. The mean time between the arrival of calls is therefore

$$E(t) = \lambda \int_0^{\infty} t e^{-\lambda t} dt = \frac{1}{\lambda}$$

For Poissonian traffic, therefore, the inter-arrival time has an exponential distribution with a mean of $1/\lambda$. This fact will form the basis for the generation of random traffic in the simulation of telephone systems.

The probability of a call arriving in a small interval of time dt is obtained by expanding $p_1(dt, \lambda)$

$$\begin{aligned} p_1(dt, \lambda) &= \lambda dt e^{-\lambda dt} \\ &= \lambda dt + o(\lambda dt)^2 \\ &= \lambda dt \quad \text{as } dt \rightarrow 0 \end{aligned}$$

3.5 The non-random arrival of calls

In Section 3.4.1, we assumed that the traffic described by the Poisson process had not been modified by any switching arrangement leading to loss or delay. This excluded smoothed and overflow traffic. We are now going to attempt a description of this "non-random" traffic.

3.5.1 The Smoothness of telephone traffic

The most usual measure of the smoothness of traffic is the ratio α of the variance to the mean. For random traffic, $\alpha=1$ as shown in Section 3.4. A ratio of zero means perfectly smooth traffic in which each terminated call is immediately replaced by a newly originated one. A ratio α greater than one defines rough traffic.

The other measure of the smoothness of telephone traffic is the ratio of the standard deviation to the mean.

3.5.2 Smooth traffic: the Binomial distribution

Smooth traffic can be described approximately by the Binomial distribution. We make two assumptions:

- (a) the number of subscribers is finite, and equals n
- (b) the probability that an idle subscriber makes a call is constant, and equal to p .

Let x be the number of calls in progress simultaneously at any given time. The experiment behind this distribution is to take the n subscribers and find the probability $p(x)$ that x of them are engaged; out of a sequence of n Bernoulli trials we are looking for x successes. $p(x)$ is then given by

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n$$

The mean m , and the variance v are given by

$$\begin{aligned} m &= E(x) = np && (= \text{offered traffic by definition}) \\ v &= E[(x-m)^2] && (= \text{variance of offered traffic}) \end{aligned}$$

$$\text{and } \rho = \frac{v}{m} = 1-p.$$

$$\text{Therefore } n = \frac{m}{1-\rho}$$

Thus if we are given the mean and the variance of the traffic, both n and p that define the Binomial distribution are easily determined.

3.5.3 Smooth traffic : the Erlang distribution

An alternative description of smooth traffic is the Erlang distribution, named after A.K. Erlang who found this family of distributions to be approximately representative of smooth traffic.

The probability density function is

$$p(t) = \begin{cases} (k\lambda)^k \left[\frac{e^{-k\lambda t}}{(k-1)!} \right] t^{k-1}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

where k is a positive integer greater than zero

The cumulative complementary distribution $P_c(t)$ is

$$P_c(t) = 1 - \frac{1}{(k-1)!} \int_0^t \lambda t^k (k) \quad t \geq 0$$

where the incomplete Gamma function $\int_y (z)$ is defined by

$$\int_y (z) = \int_0^y t^{z-1} e^{-t} dt \quad \text{real part of } z > 0$$

The case $k = 1$ corresponds to the poisson process as can be easily verified by substitution in $p(t)$ and $P_c(t)$ using the fact that $\int_y(1) = 1 - \exp(-y)$

This Erlang distribution of type k with parameter λ is, in fact, the k -fold convolution of the negative exponential distribution with parameter $k\lambda$. The mean m , and the variance v are given by,

$$m = 1/\lambda$$

$$v = 1/k\lambda^2$$

3.5.4 Rough traffic: the negative Binomial distribution

Rough traffic can be described approximately by the negative Binomial distribution, Wilkinson (129), Mina (40).

In the experiment underlying the (positive) Binomial distribution we had a fixed number n of trials out of which we were looking for x successes. In the negative Binomial distribution, on the other hand, we go on making more trials until we get x successes. The random variable of interest is the number of trials that we have to make before we obtain x successes. Let p be the probability of a success, and $q (=1-p)$ the probability of a failure. Suppose there are k failures before the first success. The probability of this event, that the first success happens at the $(k+1)$ -th trial, is

$$g(k) = q^k (1-q) \quad k = 0, 1, 2, \dots$$

This is known as the geometric distribution.

The probability that the r -th success occurs at the trial numbered $x+r$ is the same as the joint probability that there are $r-1$ successes in $x+r-1$ trials and a success on the $(x+r)$ -th trial. It is given by

$$p_r(x) = \binom{x+r-1}{r-1} p^{r-1} q^x p \quad x = 0, 1, \dots$$

Using the identity

$$\binom{x+r-1}{r-1} = \binom{x+r-1}{x} = (-1)^x \binom{-r}{x}$$

We can therefore re-write $p_r(x)$ as $p_r(x) = \binom{-r}{x} (-q)^x (1-q)^r, x = 0, 1, \dots$

Note that we have now tacitly resumed our assumption that we have a countably infinite number of subscribers as we did in the Poisson process.

Although in the derivation of $p_r(x)$ we assumed that r was a positive integer, $p_r(x)$ remains meaningful for any non-negative real number r .

Note :

$$\binom{r}{k} = \frac{r(r-1)\dots(r-k+1)}{k(k-1)\dots(1)} \quad \text{for all real } r \text{ and integer } k \geq 0$$

$$= 0 \quad \text{integer } k < 0$$

The mean m , and the variance v of the negative binomial distribution are

$$m = \frac{kq}{1-q}$$

$$v = \frac{kq}{(1-q)^2}$$

And

$$\lambda = \frac{v}{m} = \frac{1}{1-q}$$

$$\tau = \frac{m}{-1}$$

The parameters r and q which define the distribution are therefore easily determined from m and v .

3.5.4.1 The ergodicity of the negative Binomial process

In both the positive and the negative Binomial processes, the calling rate $\lambda(x)$ is dependent on the number x of calls in progress. For the positive Binomial process, $\lambda(x)$ decreases as x increases, and it is therefore intuitively obvious that the system will attain statistical equilibrium. In the negative Binomial case, on the other hand, $\lambda(x)$ increases with x . It can be shown, Wallstrom (148), that this process will also attain statistical equilibrium if for large x , $\lambda(x)$ does not increase faster than a linear function of x in which the coefficient of x is less than unity. The negative Binomial process satisfies this condition.

3.5.5 Rough traffic : the hyper-exponential distribution

The hyper-exponential distribution defines a class of distributions that have a standard deviation that is greater than the mean. This class of distributions can, therefore, be used to approximate rough traffic. But it has, in fact, found more application in queuing theory than it has in telephone traffic theory. The distribution $F(t)$ can be written as,

$$F(t) = \sum_{k=1}^n a_k (1 - \exp(-\lambda_k t)) \quad , \text{ for } t \geq 0, \quad \text{with } a_k, \lambda_k > 0, \\ n \geq 1, \text{ and } \sum_{k=1}^n a_k = 1 \\ = 0 \quad , \quad t < 0$$

The mean $E(t)$, and the second moment $E(t^2)$ are given by Cohen(39) as

$$E(t) = \sum_{k=1}^n a_k / \lambda_k$$

$$\text{and} \quad E(t^2) = 2 \sum_{k=1}^n a_k / \lambda_k^2$$

3.6 Duration of calls: the negative exponential distribution

In a study of telephone traffic, we must postulate a probability distribution function for the holding time of the successful calls. This is evidently necessary since the telephone traffic we are trying to describe is essentially a birth and death process.

It is usually assumed that the duration of the calls follows the negative exponential distribution. Thus the probability that a particular call does not terminate before a duration t , or equivalently, that the holding time exceeds a time interval t , is given by

$$P_c(t) = e^{-t/\tau}, \quad \text{where } \tau \text{ is a parameter}$$

(The subscript c stands for complementary, because $P_c(t)$ is a complementary cumulative distribution function). The probability density function is given by

$$p(t) = \frac{d}{dt} [1 - P_c(t)] = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

and is illustrated for some values of τ in Fig.3.2.

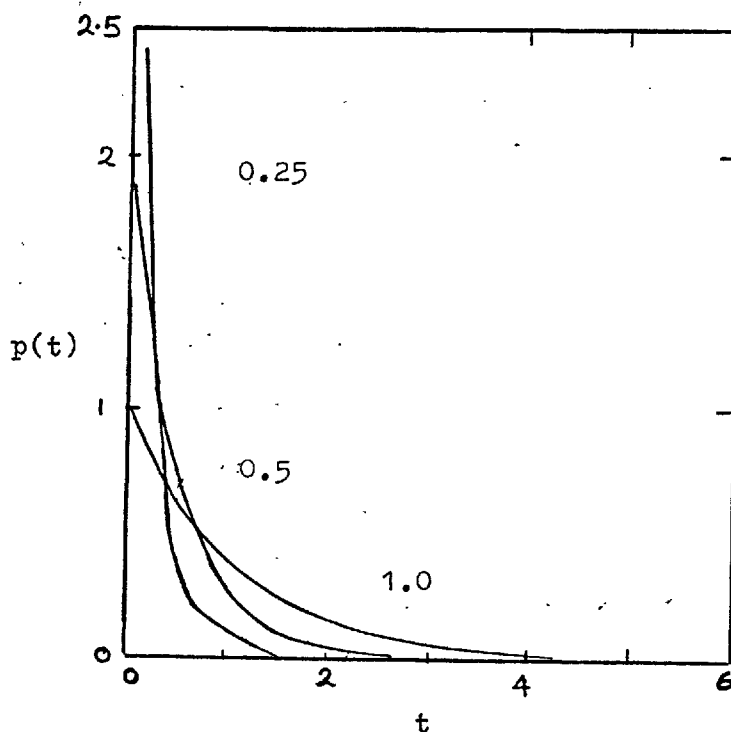


Fig.3.2 The negative exponential density function for several values of τ .

The mean holding time $E(t)$ is

$$E(t) = \int_0^{\infty} \frac{t}{\tau} \exp(-t/\tau) dt = \tau$$

The parameter τ , therefore, stands for the mean holding time. The probability that a call terminates in a small interval of time dt is obtained by expanding the distribution function.

$$\begin{aligned} 1 - P_c(dt) &= 1 - \exp(-dt/\tau) \\ &= dt/\tau + o(dt^2) \end{aligned}$$

So that in the limit as $dt \rightarrow 0$, the probability of a particular call terminating is proportional to dt . If we have n calls, the probability of one out of the n calls terminating in the time interval dt is evidently $n dt/\tau$.

From repeated series of observations of real telephone traffic, Kruithof(21), it has been found that the negative exponential distribution function fits the observations reasonably well, Figs. 3.3 and 3.4.

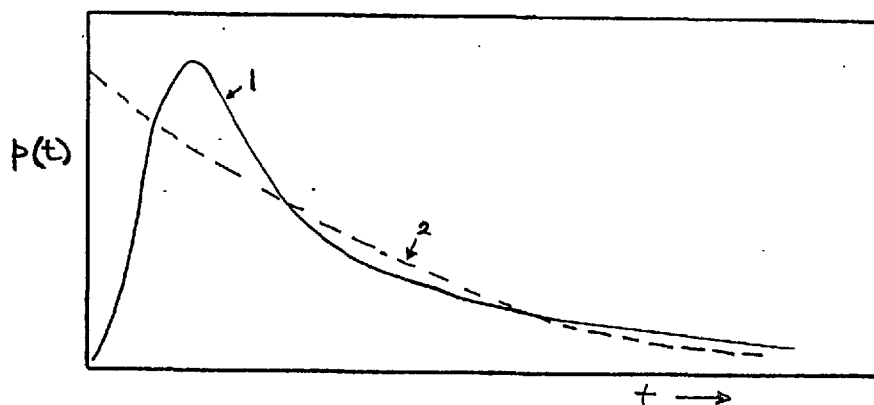


Fig.3.3 Distribution of holding times *for local calls*
 1. Observed distribution
 2. Exponential distribution

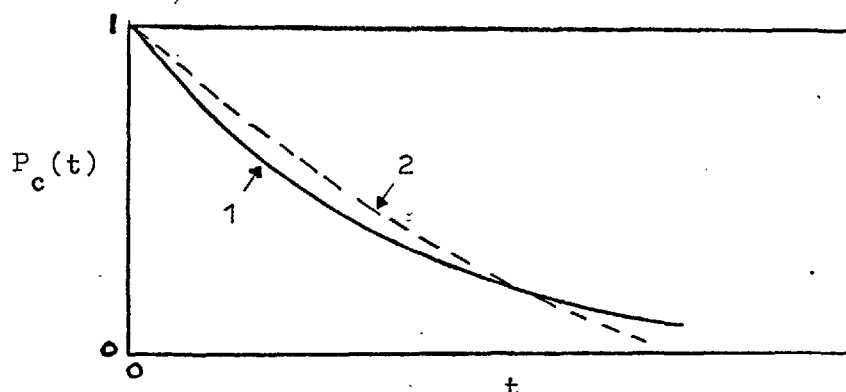


Fig.3.4 Distribution of holding times for local calls
 1. Observed distribution
 2. Exponential distribution

Actually the distribution of the holding times is not vitally important in the full availability case, and many loss formulae, including Erlang's formula, make no assumptions about the distribution apart from requiring that it has a finite mean. The distribution is important, however, for limited availability and overflow traffic situations.

3.7 Loss Formulae

In a telephone system working on a demand basis, it is not always possible to connect an attempted call. The reasons are entirely economic. The complete elimination of congestion would mean the use of leased circuits instead of a switched circuit network, resulting in expenses that are judged to be unsuitable because of their influence on the tariff charges. It is important, therefore, in attempting to evaluate the performance of a network to study first the loss of a simple system offered random or non-random traffic.

The simplest system is a network consisting of only two exchanges, A and B, connected by a group of n two-way circuits, Fig 3.5.

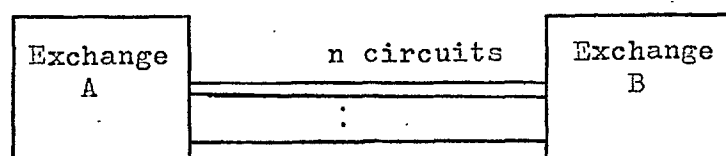


Fig. 3.5 A two-exchange network

Let us assume, without any loss of generality, that all the traffic is originated by the subscribers connected to exchange A. We are now interested in the probability that a call originated at A will find all n circuits busy, i.e. the probability that the call *is lost*.

3.7.1 The Erlang loss formula

Since this is the most important formula, we shall deal with it first.

Let us assume that:

- (1) The inter-arrival times of the calls are independent random variables with the same negative exponential distribution with mean $1/A$.
- (2) The holding times and the inter-arrival times are independent of each other.
- (3) The holding times are independent random variables, all with mean equal to μ . We further assume, without any loss of generality, that $\mu = 1$, i.e. the mean holding time is being used as the unit of time.

We have now defined a Birth and Death random process a realisation of which is shown in Fig. 3.6.

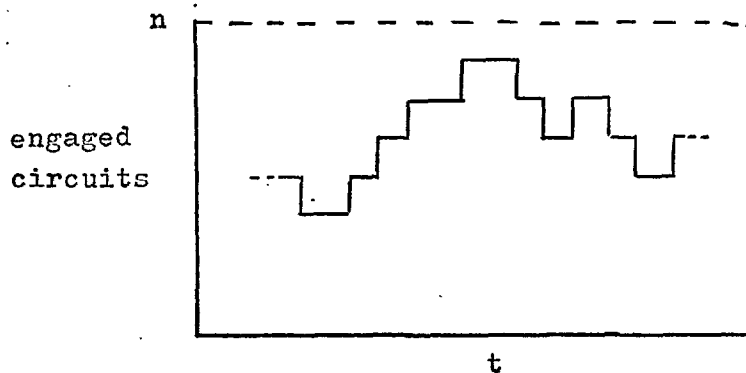


Fig 3.6 A realisation (sample function) of the traffic Birth and Death Process.

Let p_x denote the stationary probability that x out of the n circuits are in use (the system is in state x) at any given time. We can write down the statistical equilibrium equations for state x using Fig.3.7.

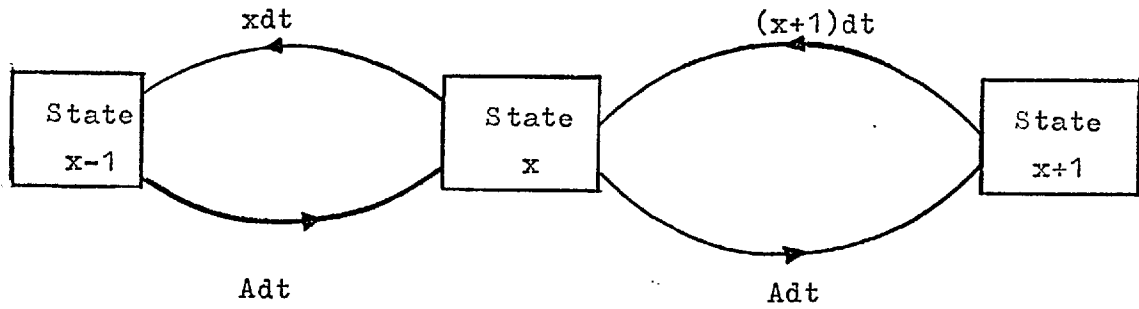


Fig. 3.7 Transitions into and out of state x

The total probability of a transition out of state x in time dt is equal to the total probability of a transition into state x under statistical equilibrium. Therefore,

$$(xdt + Adt + o(dt))p_x = (Adt + o(dt))p_{x-1} + (x+1)dtp_{x+1}$$

In the limit as $dt \rightarrow 0$,

$$(x + A)p_x = Ap_{x-1} + (x + 1)p_{x+1}, \quad x = 1, 2, \dots, n-1$$

Therefore

$$p_{x+1} = \frac{(A+x)p_x - Ap_{x-1}}{x+1}$$

By considering the equilibrium conditions for state 0, we find that

$$Adtp_0 + o(dt) = dtp_1$$

And in the limit as $dt \rightarrow 0$,

$$p_1 = p_0 A$$

Using these starting values, we can determine all the p_x by induction. We find that

$$p_x = \frac{A^x p_0}{x!}$$

Now the probabilities sum to unity.

Therefore

$$p_0 = \frac{1}{\sum_{m=0}^n \frac{A^m}{m!}}$$

So the probability of finding x circuits in use is

$$p_x = \frac{\frac{A^x}{x!}}{\sum_{m=0}^n \frac{A^m}{m!}} = E_n(A, x)$$

$E_n(A, x)$ is known as Erlang's formula. Whenever it is understood that $x = n$, it is usually written as $E_n(A)$, and is the probability that a call arriving at any time t will find all the n circuits busy, i.e. the probability that a call will be blocked.

3.7.1.1 Some properties of Erlang's formula

- (a) Because we assumed a Poisson arrival of calls in the derivation of Erlang's formula, we implicitly assumed that we had an infinite number of subscribers. This assumption was necessary so that the probability of a transition from state x to state $x+1$ is constant at A for all $x=0, 1, \dots, n-1$. Clearly, if the number of subscribers is finite but large compared with n , this transition probability will remain reasonably constant, and we can still use Erlang's formula without much error.
- (b) Erlang's formula was derived assuming the holding times to have a negative exponential distribution but it can be shown that it is valid provided only that the holding time has a finite mean.
- (c) We can express the loss formula $E_n(A)$ in terms of the incomplete Gamma function, Syski (22), as

$$E_R(A) = \frac{A^R e^{-A}}{\Gamma_A(R+1)}$$

where

$$\Gamma_A(R+1) = \int_A^\infty t^R e^{-t} dt$$

This expression for $E_R(A)$ is valid for any positive real number R . It is identical with the original expression for Erlang's formula for integer R , and it interpolates for the original formula when R is not an integer.

This extension of Erlang's formula is useful because in some calculations, especially in optimisation procedures, an intermediate result for the trunks in a group can turn out to be a non-integer. If this happens, we use the extended formula and continue our calculation at the end of which we round up our real R to the next highest integer. If we had rounded up in the middle of the calculation, we would have had rounding-off errors piling up.

(d) Since
$$\Gamma_A(R+1) = e^{-A} A^R + R \Gamma_A(R)$$

it follows that

$$E_R(A) = \frac{A E_{R-1}(A)}{R + A E_{R-1}(A)}$$

This recursive relation can be used in the computation of Erlang's formula.

3.7.2 The Engset (Erlang's Bernoulli) formula

Let us now assume that:

1. The number of subscribers N is finite, and greater than the number of circuits n . If $N \leq n$, we can have no loss.
2. The probability of exactly 1 call arriving in a

time interval dt , given that x circuits are engaged is

$$(N-x)\beta dt + o(dt)$$

where β is the calling rate, i.e. the expected number of calls per unit time per subscriber when the subscriber is idle. The probability of more than one call arriving in dt is $o(dt)$.

3. The probability that a call in progress will terminate during dt is equal to dt . We are here assuming, with no loss in generality, that the mean holding time is unity.
4. Busy sources make no calls

Proceeding as in the derivation of Erlang's formula, let us denote the stationary probability that x circuits are in use at any time t by p_x . We can now write down the statistical equilibrium equations with reference to Fig. 3.8.

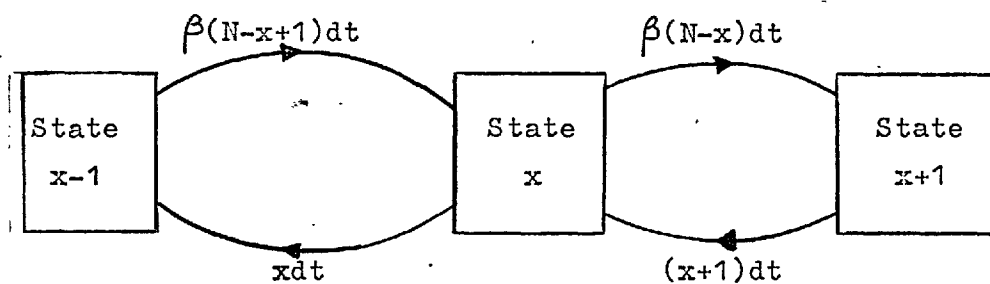


Fig. 3.8 Transitions into and out of state x

$$p_{x-1}\beta(N-x+1)dt + p_{x+1}(x+1)dt + o(dt) = p_x(xdt + \beta(N-x)dt) + o(dt)$$

for $x = 1, \dots, n-1$

and

$$p_0\beta Ndt + o(dt) = p_1dt$$

Dividing through by dt , and then letting $dt \rightarrow 0$,

$$p_{x+1} = \frac{p_x(x + \beta(N-x)) - p_{x-1}\beta(N-x+1)}{x+1}$$

By induction,

$$p_x = \frac{p_0 \beta^x N(N-1)(N-2) \dots (N-x+1)}{x!}$$

Imposing the condition that the p_x sum to unity we find that

$$p_x = \frac{\binom{N}{x} \beta^x}{\sum_{m=0}^n \binom{N}{m} \beta^m} \quad x = 0, 1, \dots, n$$

Let us denote the traffic offered per source by a .

(Note that this is different from β which is the traffic offered per subscriber when the subscriber is idle)

It can be shown that,

$$a = \frac{\beta}{1+\beta} + \left(1 - \frac{n}{N}\right) \frac{\beta^2}{1+\beta} p_n$$

If p_n is small,

$$a \doteq \frac{\beta}{1+\beta}$$

It can be shown that the call congestion is

$$\frac{\beta}{a} \left(1 - \frac{n}{N}\right) p_n$$

and is always less than the time congestion p_n . The total offered traffic is N_a .

3.8 Overflow traffic

There are two major traffic intensity fluctuations observed in a telephone network:

- (a) "primary fluctuations" due entirely to the varying activities of the subscribers,
- (b) "secondary fluctuations" caused by interactions between subscribers and the telephone network, e.g. fluctuations due to overflow traffic, or repeated calls.

It is the overflow traffic that concerns us here.

The (renewal) process of calls overflowing from a full availability group offered Poisson traffic has been considered by Wilkinson (9) leading to his Equivalent Random Theory. (It had already been shown by G. Palm (13) that the interarrival times of the overflowing calls are independent, and equally distributed. ie. the overflow process is a renewal process)

3.8.1 The Equivalent random theory

The Equivalent random theory offers an approximate solution to the congestion problem when a common group of trunks is offered several independent traffic streams, some or all of which have overflowed from first choice trunk groups offered Poisson traffic.

It is assumed that the traffic can be adequately described by the mean m , and the variance v , of its probability distribution. The mean of the traffic offered to the common group is clearly the sum of the individual means. The variance is also equal to the sum of the individual variances since the covariance is equal to zero because of the independence assumption.

Kosten (14) has shown that the total traffic overflowing trunk groups of L_i trunks offered A_i erlangs of random traffic has a mean m and variance v given by

$$m = \sum_i A_i E_{L_i}(A_i)$$

$$v = \sum_i A_i E_{L_i}(A_i) \left[1 - A_i E_{L_i}(A_i) + \frac{A_i}{L_i + 1 - A_i(1 - E_{L_i}(A_i))} \right]$$

Wilkinson (9) found that the composite traffic (m, v) can be approximated by the traffic of equal mean m and variance v overflowing a single full availability group (called the equivalent group) offered Poisson traffic.

We now have an equivalent group of L trunks offered A erlangs, having overflow traffic of mean m , and variance v .

i.e.

$$m = AE_L(A)$$

$$v = AE_L(A) \left[1 - AE_L(A) + \frac{A}{L + 1 - A(1 - E_L(A))} \right]$$

From these implicit equations, A and L can be computed by iteration. The starting values for the iteration

$$A = v + 3 \frac{v}{m} \left[\frac{v}{m} - 1 \right]$$

$$L = \frac{A}{1 - 1/(m + v/m)} - m - 1$$

have been found to be "remarkably accurate" by Rapp(57) and Wallstrom(8).

The equivalent random theory applies to rough traffic only. In the case of smooth traffic, we can use the "carried equivalent theory" developed by Katz(98).

CHAPTER 4. ROUTING IN A SWITCH

- 4.1 Introduction
- 4.2 The 6 to 4 access concentrator problem
 - 4.2.1 State assignment
 - 4.2.2 Probabilistic assumptions and network operation
 - 4.2.3 Mathematical formulation of the problem
 - 4.2.4 The statistical equilibrium equations for the general state x
 - 4.2.5 Solution of the problem
- 4.3 The Clos network of switches
- 4.4 Conclusion

CHAPTER 4. ROUTING IN A SWITCH

4.1 Introduction

The aim of this chapter is to review the methods that have been used in solving routing problems in a switch, the 6 to 4 access concentrator, and in a network of switches, the Clos network. After studying these two examples, we ask ourselves whether we can use similar methods for solving the routing problem in a network of exchanges. The answer is that in general we cannot.

The chapter is based closely on the work that Dr. Benes (25) has done at the Bell Laboratories in the U.S.A. He has formulated and given the formal solution to the 6 to 4 access concentrator routing problem. Although he gives a formal solution, and then some computed results, he gives very little intermediate detail. My contribution to this chapter is, therefore, in working out the intermediate stages within Dr. Benes's framework. This is necessary in order to appreciate whether or not the same methods can be applied to routing in a network of exchanges.

4.2 The 6 to 4 access concentrator problem

The 6 to 4 access concentrator switch is shown in Fig. 4.1

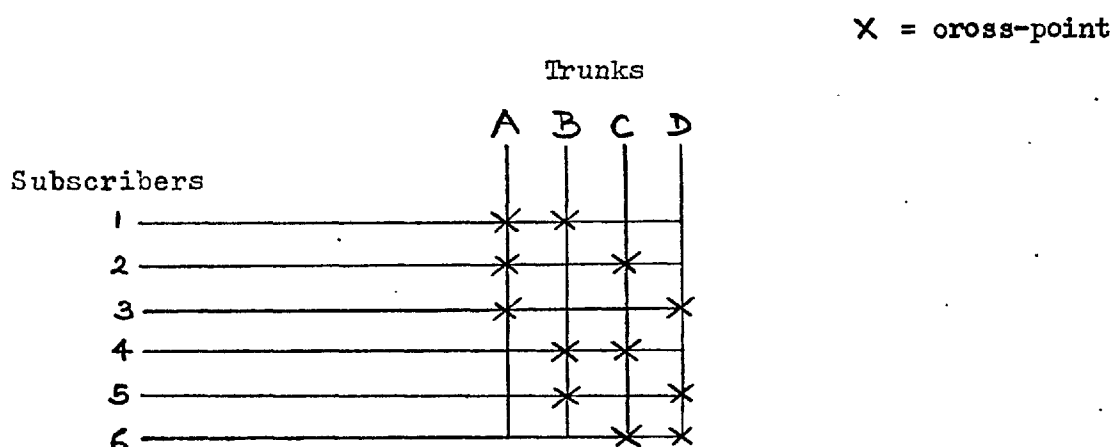


Fig. 4.1 The 6 to 4 access concentrator switch.

Each of the six subscribers has access to two trunks. For example subscriber 1 has access to trunks A and B. The problem is that if at any given instant a subscriber wants a connection to the trunk lines, how should this connection be effected? Different questions arise depending on the state of the network (concentrator):

1. If both trunks to which he has access are available, to which one should he be connected?
2. If there is only one trunk available, should he or shouldn't he be connected to it?

In order to proceed to a general solution to this problem, it is necessary to make the idea of a "state of the network" more precise.

4.2.1 State assignment

Any given cross-point in Fig. 4.1 is either open or closed, i.e. it has only two possible states. Therefore, given all the states of the cross-points, we can define the state of the network.

The problem of assigning states is simplified if we adopt the (mixed)graph representation of Fig. 4.2

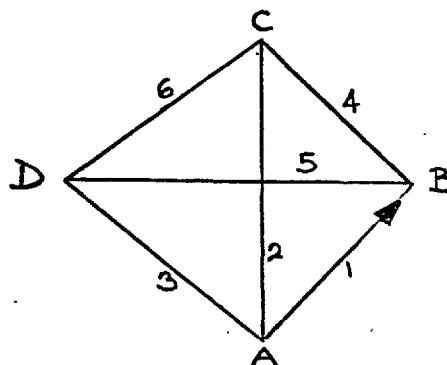


Fig. 4.2 The diagrammatical representation of a state

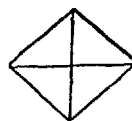
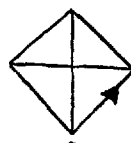
Fig. 4.2 illustrates the state in which subscriber 1 is connected to trunk B. The number on each branch of the graph stands for a subscriber, while each node stands for a particular trunk. The arrow points to the trunk which is being used.

By enumerating the number of different ways of placing 1 call, then 2 calls, then 3 calls, and finally 4 calls, a large number of primitive states is obtained.

To reduce this large number of states, certain equivalences are noted, so that a state now really stands for a class of primitive states. This process has been carried out on the next four pages, in which we define the states starting with state 0, (the ground state in which no call is in progress), and working up to state 11. As each state is defined, we enumerate the states that can be reached from that state by the placing of one more call. If this placement can be done in two ways which lead to different states, then we have a routing choice. The reader will notice that routing choices exist in only states 1, 3 and 4.

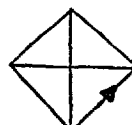
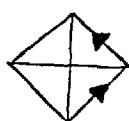
State Assignment

State 0

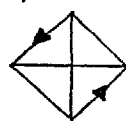
States reachable from
state 0

1

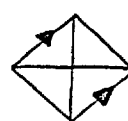
State 1

States
reachable
from state 1

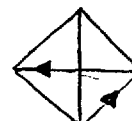
3



2



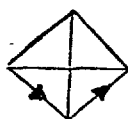
2



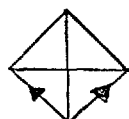
3

Choice

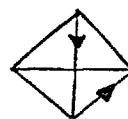
Choice



3



4

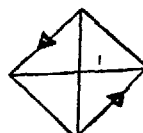
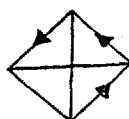


3

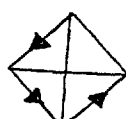


4

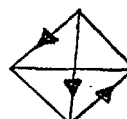
State 2

States
reachable
from state 2

6



6

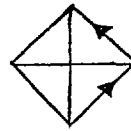


7

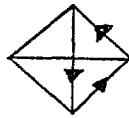


7

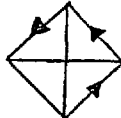
State 3



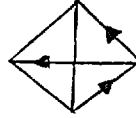
States
reachable
from state 3



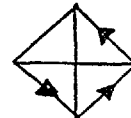
5



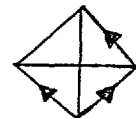
6



8



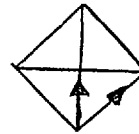
6



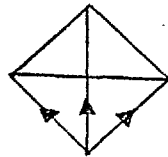
7

Choice

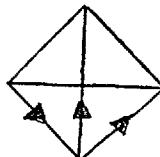
State 4



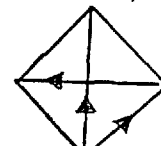
States
reachable
from state 4



9



8



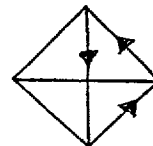
7



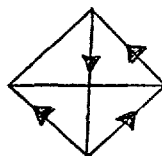
7

Choice

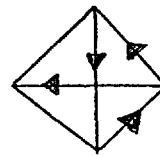
State 5



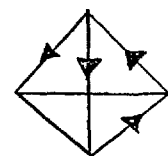
States
reachable
from state 5



11

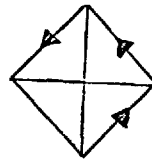


11

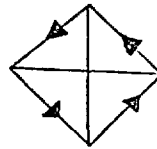


11

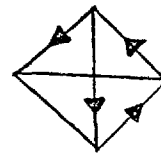
State 6



States reachable
from state 6

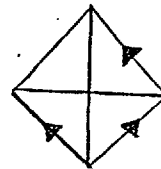


10

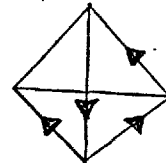


11

State 7

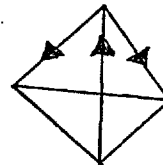


State reachable
from state 7

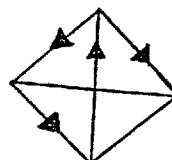


11

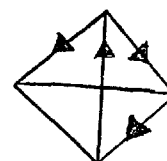
State 8



States reachable
from state 8

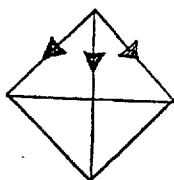


11

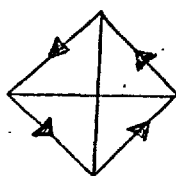


11

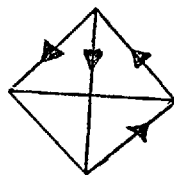
State 9



State 10



State 11



4.2.2 Probabilistic assumptions and network operation

Each of the six subscribers (also called inlets) has a calling rate of λ calls per unit of time. The holding time of each call is a random variable with a negative exponential distribution of unity mean. (The mean holding time being unity, also serves as the unit of time.)

We define three functions on the set of states S :

1. $\alpha(x)$ = number of idle inlets in state x .
2. $\beta(x)$ = number of blocked inlets that are idle in state x .
3. $s(x)$ = number of calls that can be connected in state x .

In fact, $s(x) = \alpha(x) - \beta(x)$.

The values of these functions are given in the table below.

State x	0	1	2	3	4	5	6	7	8	9	10	11
$\alpha(x)$	6	5	4	4	4	3	3	3	3	3	2	2
$\beta(x)$	0	0	1	0	1	0	1	2	1	3	2	2
$s(x)$	6	5	3	4	3	3	2	1	2	0	0	0

The reader can easily verify the entries in the table by referring to the state assignment diagrams.

The operation of the network can be viewed as a particle making a random walk on the states. The routing rules determine the "upward" transition probabilities. The "downward" transition probabilities, which are independent of the routing rules, are shown in Fig. 4.3

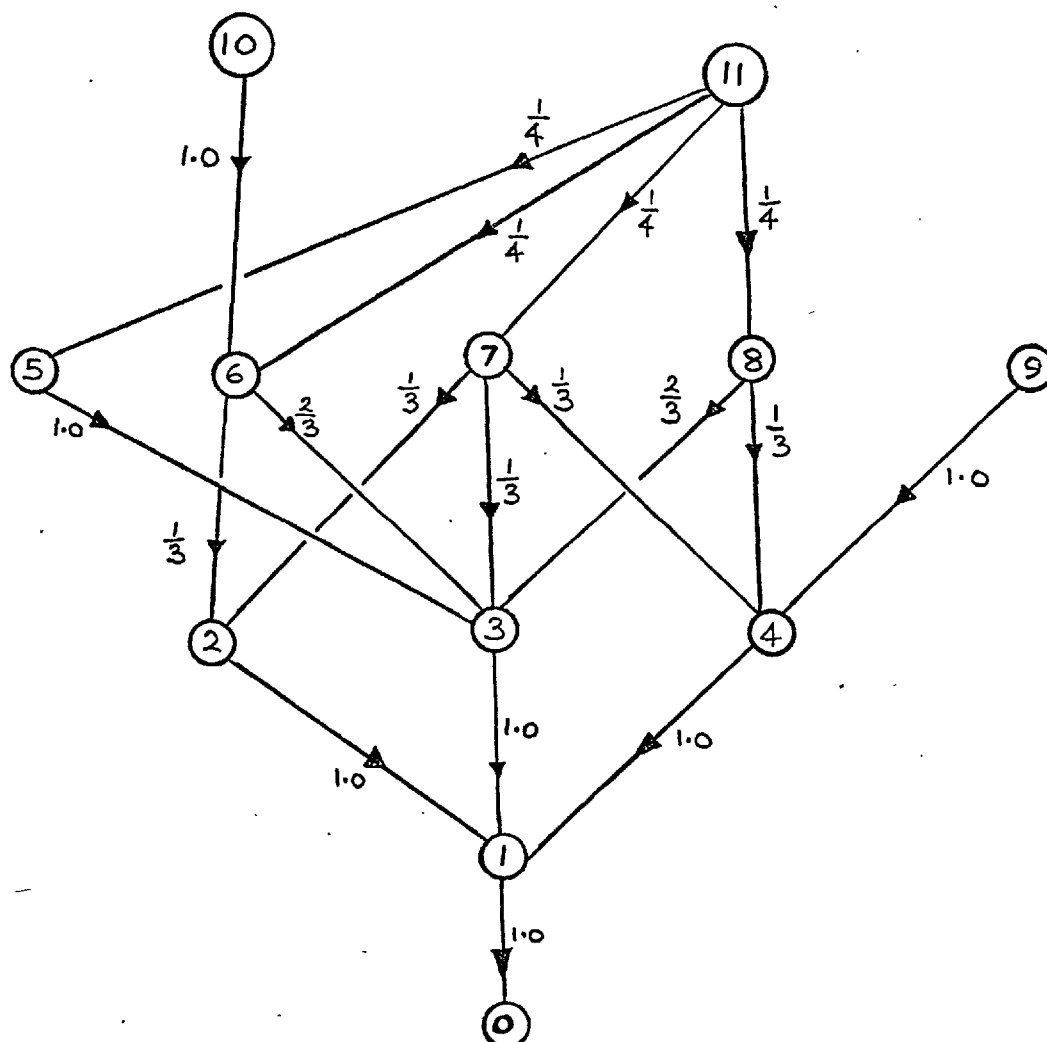


Fig. 4.3 Downward Transition Probabilities

4.2.3 Mathematical formulation of the problem

Let

A_x = Set of states which can be reached from x by adding one call.

B_x = Set of states which can be reached from x by terminating one call.

μ = mean holding time.

$|x|$ = number of calls in progress in state x .

$\gamma(y, x)$ = the probability of a transition from state y to x as a result of a call terminating in y (Fig. 4.3), multiplied by $|y|$.

$\lambda(x, y)$ = the rate at which x calls y ($y \in A_x$), i.e. the rate at which the network moves from state x to state y .

P_x = equilibrium (stationary) probability of state x .

$r_{xy} = r(x, y)$

= the probability of moving into state $y (\in A_x)$ as a result of a call being placed in state x .

Fig. 4.4 shows $\lambda(x, y)$ for our state assignment.

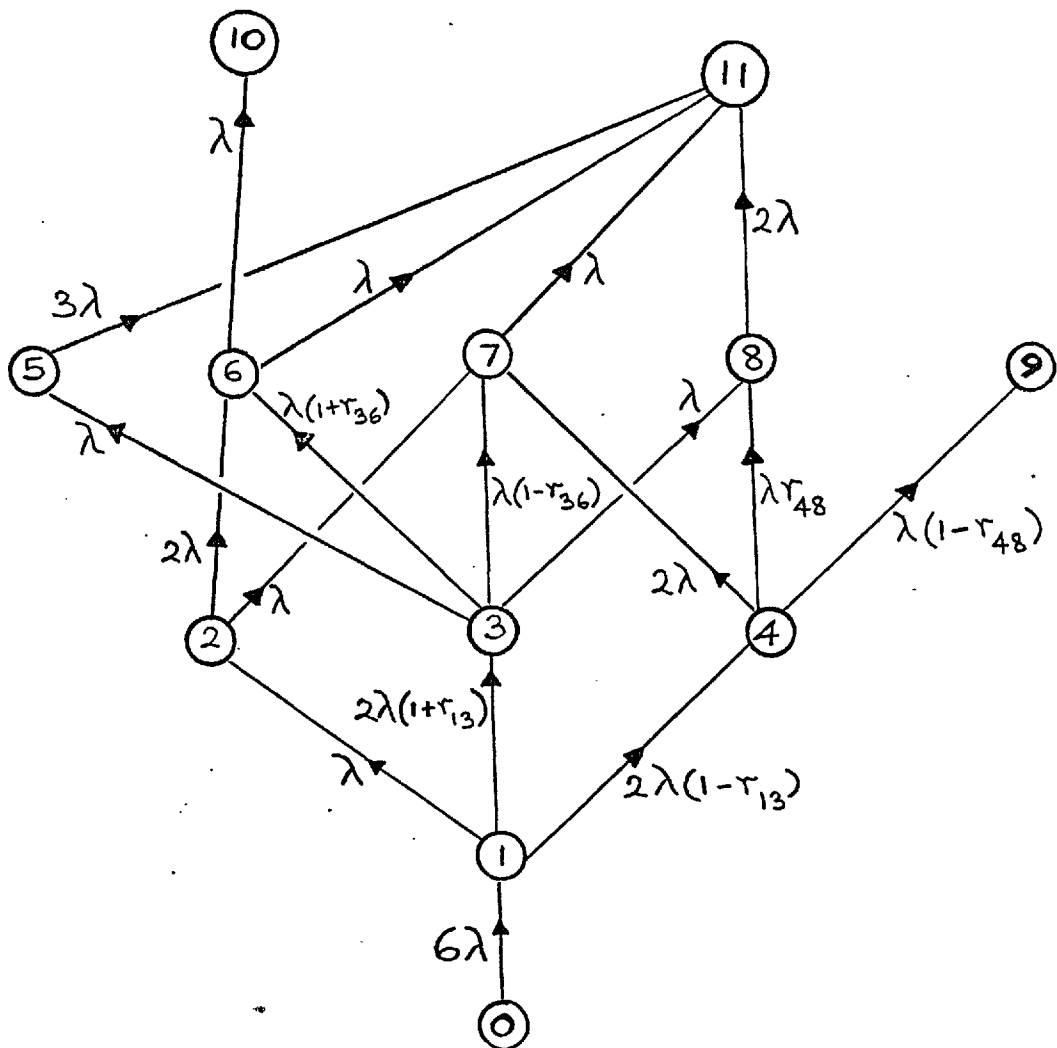


Fig. 4.4 Diagram showing $\lambda(x, y)$

Benes (25) has shown that the probability of congestion B (overall grade of service) of the switch is given by

$$B = \frac{\sum_{x \in S} p_x \beta(x)}{\sum_{x \in S} p_x \alpha(x)}$$

The problem, therefore, is to choose matrix $R(= r_{xy})$, which determines the stationary probabilities p_x , so as to minimise B .

In the Section 4.2.1, we noticed that we only had a routing choice in three states. This is again reflected in Fig. 4.4.

1. In state 1, two calls could arise that we can accommodate in such a way that we make a transition to either state 3 or state 4.
2. In state 3, a call could be offered that can be routed in such a way that we make a transition to either state 6 or state 7.
3. A call can arise in state 4 which can be completed in such a way that a transition to state 8 or state 9 occurs.

Therefore, in the 12 by 12 matrix R , the only elements of real interest are r_{13} , r_{14} , r_{36} , r_{37} , r_{48} , and r_{49} . Since, in fact,

$$r_{14} = 1 - r_{13}$$

$$r_{37} = 1 - r_{36}$$

$$\text{and } r_{49} = 1 - r_{48}$$

we need only concern ourselves with r_{13} , r_{36} , and r_{48} .

There is no need to treat the general matrix R . But if for some reason we wanted to use the whole matrix R , we would proceed as Dr. Benes (25) and make the transformation

$$u_{xy} = p_x r_{xy}$$

which reduces the unknowns to one matrix U . A useful check on the row sums of U is

$$\sum_{y \in A_x} u_{xy} + u_{xx} = p_x \left(r_{xx} + \sum_{y \in A_x} r_{xy} \right) = p_x s(x)$$

4.2.4 The statistical equilibrium equations for the general state x

The statistical equilibrium equations define the relationships between each state x , and the other states adjacent to it.

At equilibrium, the probability of entering a state x is equal to the probability of leaving it.

Now, there are two (mutually exclusive) ways of entering x , or leaving it. State x may be left by either entering $y \in B_x$ or $y \in A_x$, and may be entered by leaving $y \in A_x$ or $y \in B_x$.

The probability of leaving x in the small interval δt

The probability of entering $y \in B_x$ from x equals P_x times the probability that one of the calls in progress will end in the time interval δt , i.e. $P_x |\alpha| \mu \delta t$. The probability of moving to $y \in A_x$, depends on P_x times the rate at which the demand for connection leads to transitions to y i.e. $\lambda(x, y)$, times the probability of routing an incoming call so as to lead to y , i.e.

$$\sum_{y \in A_x} \lambda(x, y) \tau_{xy} P_x \delta t$$

The probability of entering x

To enter x from $y \in A_x$ a call must terminate. But the state $z \in B_y$ that the system goes into is determined by the state assignment. So the probability of entering x from A_x equals $\sum_{y \in A_x} P_y \gamma(y, x) \mu \delta t$. Entering x from $y \in B_x$ is very similar to moving from x to $y \in A_x$. It is

$$\sum_{y \in B_x} \lambda(y, x) P_y \tau_{yx} \delta t$$

Now equating the probability of entering x , and that of leaving it we get

$$\left[|\alpha| + \sum_{y \in A_x} \lambda(x, y) \tau_{xy} \right] P_x = \sum_{y \in A_x} \gamma(y, x) P_y + \sum_{y \in B_x} \lambda(y, x) P_y \tau_{yx}$$

The equilibrium equations can now be written as follows:

$$(0 + 6\lambda)p_0 - p_1 = 0$$

$$(1 + \lambda + 2\lambda(1 + r_{13}) + 2\lambda(1 - r_{13}))p_1 - 2p_2 - 2p_4 - 2p_3 - 6\lambda p_0 = 0$$

$$(2 + 2\lambda + \lambda)p_2 - p_6 - p_7 - \lambda p_1 = 0$$

$$(2 + \lambda + \lambda(1 + r_{36}) + \lambda(1 - r_{36}) + \lambda)p_3 - 3p_5 - 2p_6 - p_7 - 2p_8 - 2\lambda p_1 - 2\lambda p_1 r_{13} = 0$$

$$(2 + 2\lambda + \lambda r_{48} + \lambda(1 - r_{48}))p_4 - p_7 - p_8 - 3p_9 - 2\lambda(1 - r_{13})p_1 = 0$$

$$(3 + 3\lambda)p_5 - p_{11} - \lambda p_3 = 0$$

$$(3 + \lambda + \lambda)p_6 - 4p_{10} - p_{11} - \lambda(1 + r_{36})p_3 - 2\lambda p_2 = 0$$

$$(3 + \lambda)p_7 - p_{11} - \lambda p_2 - \lambda(1 - r_{36})p_3 - 2\lambda p_4 = 0$$

$$(3 + 2\lambda)p_8 - p_{11} - \lambda p_3 - \lambda r_{48}p_4 = 0$$

$$(3 + 0)p_9 - \lambda(1 - r_{48})p_4 = 0$$

$$(4 + 0)p_{10} - \lambda p_6 = 0$$

$$(4 + 0)p_{11} - 3\lambda p_5 - \lambda p_6 - \lambda p_7 - 2\lambda p_8 = 0$$

The following constraints must also be satisfied:

$$\sum_{j=0}^{11} p_j = 1$$

$$p_1 r_{13} - p_1 \leq 0$$

$$p_3 r_{36} - p_3 \leq 0$$

$$p_4 r_{48} - p_4 \leq 0$$

4.2.5 Solution of the problem

The problem is formally solved if we can determine all the p_i such that

$$B = \frac{\sum_{x \in S} p_x A(x)}{\sum_{x \in S} p_x C(x)} \quad \text{is minimised.}$$

From the equilibrium equations, we note that

$$p_1 = 6\lambda p_0$$

$$\text{and } 4p_{10} = \lambda p_6.$$

Two equations can, therefore, be eliminated by substituting for p_0 and p_{10} wherever they occur.

We can now proceed to solve the routing problem by linear programming. The function to be minimised is the numerator of B , with the additional constraint that the denominator is equal to unity. The cost vector is filled in with the values of β , while the cost for the three slack variables is zero. Note that the constraint that the probabilities sum to unity is only imposed as a normalising condition after the linear program has been completed; this constraint would otherwise be inconsistent with the constraint that the denominator of B is equal to unity.

Solution by inspection

If the number of states of a network, similar to the 6 to 4 access concentrator, in which a routing choice exists is not large, we may be able to solve the routing problem by inspection. For the 6 to 4 access concentrator, there are only three states in which alternative routes exist. These are states 1, 3, and 4. As can be seen from the table of page 57, in state 1 we prefer a transition to state 3 to a transition to state 4 because $s(3)$ is greater than $s(4)$. We therefore choose r_{13} to be equal to one. Similarly, we put both r_{36} and r_{48} equal to unity. We now have as many equilibrium equations as we have unknowns; we can therefore solve for the stationary probabilities and hence determine the blocking B .

Computed results

The linear programming results, which agree with those obtained by inspection, are compared to those obtained by Benes and Lloyd, Benes (25), in the table of Fig. 4.5. For the random routing rule, r_{13} , r_{36} , and r_{48} are all equal to 0.5

Calling rate	Overall grade of service			
(Erlangs)	Author (L.P.)	Benes	Lloyd	Random (Benes)
0.1	0.00536	0.00536	0.0049	0.00865
0.2	0.0209	0.0209		0.02972
0.4	0.0717	0.0717	0.0716	0.08856
0.7	0.1627		0.1628	
1.0	0.2478	0.2315	0.2478	0.2494
1.3	0.3207			
1.7	0.4006			
2.0	0.4497	0.4497	0.4498	0.4614

Fig. 4.5 A comparison of computed results

4.3 The Clos network of switches

Consider the Clos network of switches, Fig. 4.6

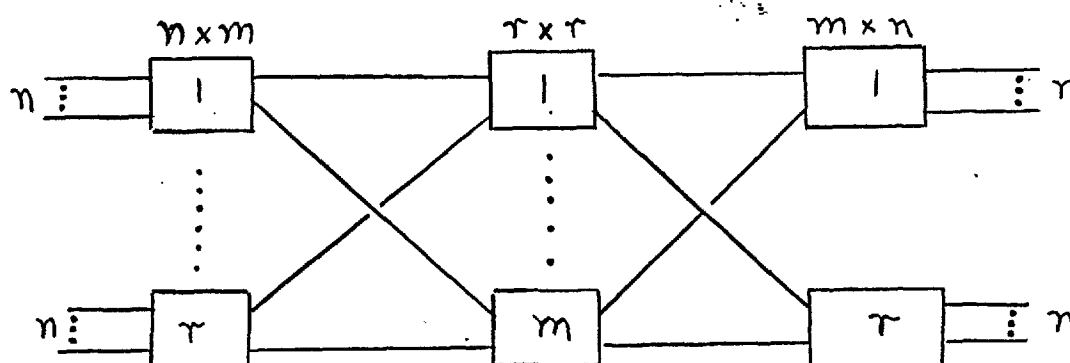


Fig. 4.6 The General Clos network.

There are m middle switches all of which have r , shown equal to two, inlets and outlets, each inlet being able to be connected to any outlet, Fig. 4.7.

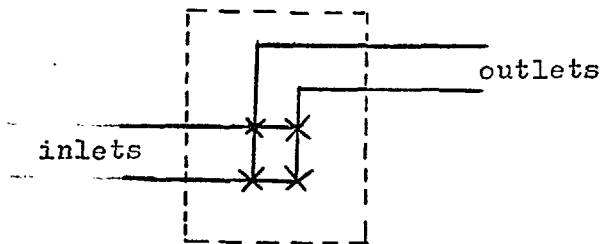


Fig. 4.7 A 2 x 2 switch

Let us define a network to be non-blocking in the strict sense if it is impossible for any call to be blocked however "badly" the calls may be routed. Let us further define a network to be non-blocking in the wide sense if it is possible to find a route for any call provided the calls are routed carefully.

Clos showed that the network of Fig. 4.6 with $m=2n-1$ is strictly non-blocking, Clos (134). Benes then made the important step in showing that the same Clos network with $m = \lceil 3n/2 \rceil$ is non-blocking in the wide sense.

($\lceil x \rceil$ is defined as the largest integer less than or equal to x .)

This step is important because it is an example of how major economies may be realised through the study of routing.

The Clos network with 2 middle switches

Fig. 4.8 is a simple Clos network for which optimal routing has been studied by Benes (25).

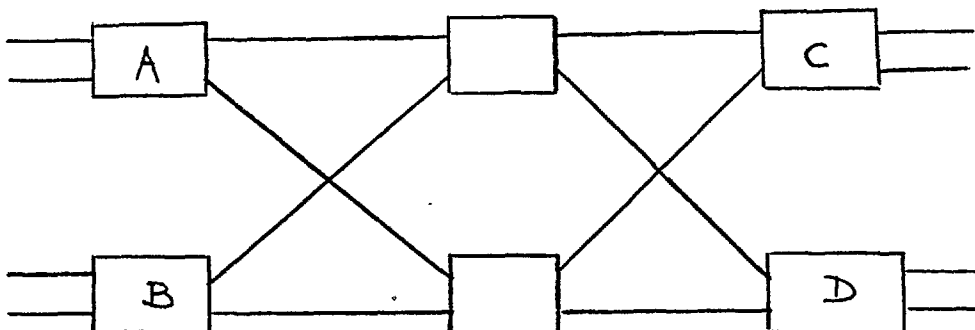


Fig. 4.8 The Clos network with two middle switches

This network can handle a maximum of four calls from the inlets to the outlets.

Consider the case where an inlet on (switch) A requires a connection to an outlet on C, and an inlet on B needs a connection to an outlet on D. These requirements can be satisfied in two different ways as shown in Figs. 4.9a and 4.10a.

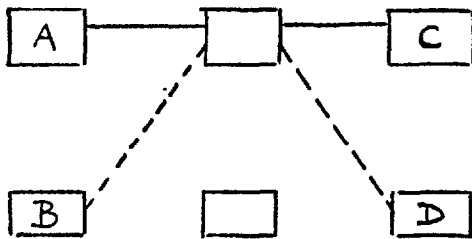


Fig. 4.9a

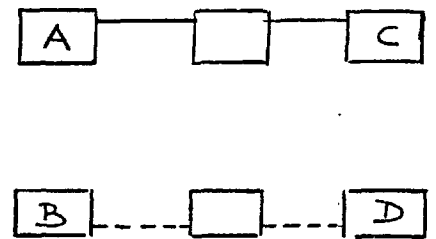


Fig. 4.10a

Suppose now that there is a further request for an inlet on A to be connected to an outlet on C. This can be done starting from either the state in Fig. 4.9a or the state in Fig. 4.10a. leading to, respectively, the states in Fig. 4.9b and fig. 4.10b.

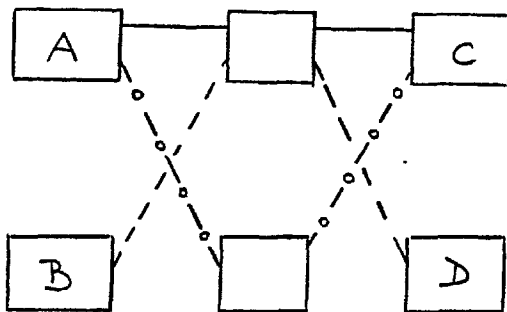


Fig. 4.9b

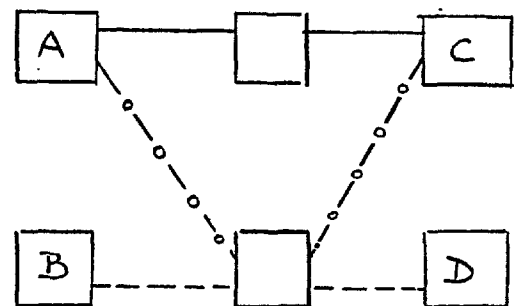


Fig. 4.10b

But if the third request is for an inlet on B to be connected to an outlet on C, it cannot be satisfied in the state shown in Fig. 4.10a

because the only free outlet from switch B leads to switch D. On the other hand, the request can be satisfied starting from the state of Fig. 4.9a. The network of Fig. 4.8 is in fact non-blocking in the wide sense. By always choosing the state in Fig. 4.9a rather than the state in Fig. 4.10a we obtain the optimal result of no blocking.

4.4 Conclusions

In this chapter, we have considered routing problems in the 6 to 4 access concentrator, and in the Clos network.

In each case we have effectively defined a partial ordering on the set of states. (A partial ordering is defined on a set S if a relation R is defined between some pairs of elements in S such that for all these pairs (a,b) , where a and b are elements of S , the relation aRb or bRa is true. If the relation R is defined over the whole cartesian product $S \times S$, we talk of a complete ordering.)

In the case of the access concentrator, we have seen that the problem can be solved either by linear programming or by inspection of the partial order of the states. The Clos network was solved by inspection. The question now arises. Can we use these methods in studying routing problems in a network of exchanges?

The answer turns out to be 'no' for solution by inspection unless the number of states is very small. But there is no clear cut answer if we wish to proceed by linear programming. One obvious limitation is the total number of states; if we have a very large number of states, it may turn out that we cannot solve the linear program. This could be due to a shortage of high speed (core) memory in a digital computer. It may be that the limits of a standard L.P. program may be exceeded by, say, the number of variables or the number of constraints, or the computer run time needed may be ridiculously long. Another disadvantage is that we have to store the whole routing matrix.

In considering the prospects of using linear programming, we have glossed over the state assignment problem. In fact, the process of state assignment, including the problem of equivalencing certain states, is no trivial matter.

Because of the astronomically large number of states that a simple network, of say four exchanges, can take, the methods used in this chapter are quite unsuitable for use in a network of exchanges.

CHAPTER 5. SIMULATION OF A TELEPHONE NETWORK

- 5.1 Introduction
- 5.2 The need for models
- 5.3 Advantages of simulation
- 5.4 Roulette and Event-by-event simulation
- 5.5 Computer simulation programming languages
- 5.6 The generation of uniformly distributed random numbers
- 5.7 Non-uniform continuous random numbers
 - 5.7.1 The exponential distribution
- 5.8 The computer simulation of a telephone network
- 5.9 The verification of simulation results

CHAPTER 5. SIMULATION OF A TELEPHONE NETWORK

5.1 Introduction

The intelligent use of simulation allows us to investigate realistic physical systems, such as telephone networks, or their mathematical representations(models) whose complexity is such that a full analytical treatment is prohibitively difficult. The results of such an investigation enhance our understanding of the physical system, and this often in turn enables us to make appropriate simplifying assumptions about the original system, and may even enable us to carry out the mathematical analysis.

Simulation serves a threefold function:

1. It is a way of verifying theoretically obtained results.
2. It enables us to obtain experimental results in areas where the mathematical analysis is prohibitively difficult.
3. It acts as a feedback mechanism by suggesting new ideas.

The computer as a network simulator is indispensable in an investigation of traffic flow in telecommunication networks. But, perhaps, its single most important contribution is in the study of networks without assuming the independence of link occupancies in multi-link routes. Without this assumption the mathematical analysis of traffic flow bogs down completely. Alternate routing procedures become impossible to investigate.

But computer simulation should not be regarded as an automatic solution for all problems that we find difficult. We must, therefore, guard against "mechanitis" defined by mathematician B.O. Koopman as "the occupational disease of one who is so impressed by modern computing machinery that he believes that a (mathematical) problem which he can neither solve, nor formulate can be readily answered once he has access to a sufficiently expensive (high speed) machine".

5.2 The need for Models

A telephone network is a complex system subject to stochastic behaviour. If the Communications Engineer decides to make a change in the system, he cannot be sure, just by thinking about it, what side effects this change will have. He can, of course, simply make the change and

experiment with the real system. More usually, however, he will make a serious attempt to predict the effect of the change by doing some more or less complicated calculations. When he does this, he is using a model of the system, however mathematically crude it may be.

If we have an adequate model, we have the overriding advantage of being able to experiment with the model rather than the real system. This advantage is not only one of cost and convenience; many decision processes cannot be the subject of experimentation with the real system, if only because the 'real' system has not been built.

Many types of models have been used in system studies, and have been classified in a number of ways. The classification is sometimes made in terms of the nature of the system they model, such as continuous versus discrete, or deterministic versus stochastic. For our purposes, models will be treated as physical models or mathematical models.

Mathematical models have been used to analyse the loss in switching systems by Gosztony and Endrody (94).

But in this study, only physical models have been used. Let us, therefore, examine an example of a physical model of a telecommunication network.

The computer contains an image of the telecommunication network in which there is indicated for each group of circuits, defined by the telephone (or telex) exchanges where they originate and terminate, their capacity and their instantaneous availability. For each connection defined by a pair of exchanges, we note the number of calls attempted, and the number of calls which were unsuccessful due to the congestion of the circuits. We can thus evaluate the Grade of Service, i.e. the proportion of unsuccessful calls.

As input to the computer are presented, spaced out in time, calls between the different exchanges which attempt to reproduce the variations of real traffic. When a call is made, we try to route it taking into account the actual operating rules of the network, and the availabilities of the groups of circuits of the simulated network. If we can complete the call, we note that one more call has been successfully placed between the two given exchanges, and prepare to release the circuits used by this call at the end of a time corresponding to the duration of the call.

5.3 Advantages of Simulation

Three advantages of simulation will be emphasised. First, even if an analytical or relational model is available, it may still be more advantageous to use simulation particularly as analytic models for stochastic systems often deal with long run expectations and rarely provide much insight into transient behaviour. A simulation model, on the other hand, would draw explicit attention to the transient behaviour, while in the analytic model this behaviour may only be implicit.

Secondly, it is less tempting in constructing a simulation model to make unjustified assumptions in order that the problem may be made to fit a solution technique. For example, in a teletraffic simulation there is no difficulty in letting the inter-arrival distribution of calls have a distribution other than the usual negative exponential.

Thirdly, simulation is easy to use. The major drawbacks of simulation are that it is expensive, and that it is not an optimising technique.

5.4 Roulette and Event-by-event Simulation

There are two main methods of implementing a computer telephone traffic simulation in current use; the Roulette and the event-by-event methods.

In the event-by-event method, right from the beginning of the simulation, we keep a "calendar" of the time certain events are to occur. If the events are caused by occurrences that are external to the system, they are called exogenous events. An example of an exogenous event is the origination of a call by a subscriber. Events originated in the system, like clearing down the connections of a terminated call, are called endogenous events. The simulation program stores in its time calendar, the times at which all known events are to occur, and as further events are foreseen they are scheduled in the calendar. The calendar is scanned every time an event is completed so that further events arising from that event can be scheduled. ~~The most imminent event is placed at the top of the list, and the least imminent is placed at the bottom.~~

If an event takes place, like the successful connection of a call, at

simulated time t_0 , a random variable is drawn from the appropriate distribution to determine the holding time of that call, equal to say t_1 . We then immediately schedule the call to be terminated and the relevant circuits released at simulated time $t_0 + t_1$. Notice that it is easy to choose the random variable representing the holding time from any distribution function.

This technique of event-by-event simulation is also called direct call-timing simulation.

Roulette simulation is the name given to the more popular indirect method of simulation due to Kosten (110).

In this method, the system being simulated is examined regularly at equally spaced intervals of simulated time, and changes are made to the state of the system in accordance with probabilities that depend only on the state of the system at that particular epoch of time. The model so formulated is, therefore, a discrete Markov process (with its total lack of memory), and can be used to simulate a telephone system having Poisson arrivals, and holding times described by the negative exponential distribution.

To implement the roulette method, rules are needed for deciding from a random number whether a call arrival, or a call termination is required. If it is a call termination, we must use another rule to decide which trunks to release. For a call arrival, yet another rule is needed to decide which group of trunks must be examined for a free path.

This method, while being easy to use, suffers from the severe disadvantage of being able to cope with only the negative exponential distribution. This is a very severe limitation since most of the telephone systems we may wish to simulate are at one point or another offered non-random traffic. ~~Adaptation of this method to simulate delay working, as opposed to demand working, presents so many difficulties that it is better to change to the event-by-event simulation method.~~

A further disadvantage arises in connection with calls using a multi-link route, in that we cannot release all the links connecting a particular call at the same instant in simulated time. What is worse, it would seem that the information about which links were used to connected which call is not available.

The event-by-event method of simulation was adopted for the present

study because it can easily handle the following situations:

1. Smooth and Rough traffic.
2. Non-exponential distributions of call arrival and call termination.
3. Queuing situations, which may arise if a more detailed model is used.

The Roulette method cannot handle these situations easily.

5.5 Computer Simulation programming languages

Any programming language used for the implementation of a simulation model can qualify as a simulation language. But we shall confine ourselves to the languages that have been specifically written for simulation, such as SIMSCRIPT, SIMULA and GPSS III.

Early attempts at simulation revealed that the task of developing and implementing a workable representation of a system could be extremely arduous. It was apparent that some form of generalised approach was necessary to make simulation a practical technique. The simulation languages which developed serve the following two vital functions:

1. They provide a conceptual structure or framework on which a representation or model of a particular system can be built. This way of approaching the problem both reduces the intellectual task, and assures an easy transition from the modeling to the computer programming.
2. Simulation languages provide the user with "Status Descriptors" by means of which the essential elements of the system to be modelled may be described. They also provide the ability to initiate extensive modifications to the simulated system with a minimum of computer instructions, but at the same time enable modifications to be made on a highly selective and detailed basis. Most of the languages supply some kind of internal timing and control mechanism. Thus the necessary 'housekeeping' is done automatically. In some languages, e.g. GPSS III, much relevant data is collected automatically.

5.6 The generation of uniformly distributed random numbers

Simulation, and a number of other techniques in applied mathematics and statistics, depends on having available sequences of numbers which

appear as if they are drawn at random from a particular probability distribution. For convenience, these numbers will be referred to simply as random numbers. This section is devoted to outlining some methods by which these pseudo-random numbers are generated on a digital computer.

The computing procedure is as follows:

1. Choose a positive integer x_0 called the seed, or starting value.
2. Choose an integer, a , called the multiplier.
3. Choose another integer c .
4. Choose yet another positive integer m called the modulus, such that $m > x_0$, $m > |a|$, $m > |c|$

We then define the sequence x_i of non-negative integers, each x_i less than m , by means of the congruence relation.

$$x_i = ax_{i-1} + c \quad (\text{modulo } m)$$

To obtain numbers in the interval $(0,1)$ we form the sequence x_i / m

Such an arithmetic sequence is completely determined, and could hardly be called random. It turns out, however, that the statistical behaviour of our sequences is quite good, with only a few exceptions, as long as the sequences do not repeat themselves too soon.

Any sequence must repeat itself eventually, because it can contain at most m different numbers, each number in a particular sequence being determined solely by its predecessor. It is possible by imposing conditions on x_0 , a , m and c , that the maximum period m is obtained. On a binary machine, it is desirable for ease of computation to choose m to be a power of 2, but other choices are possible.

With m a power of 2, and c an odd integer, with $a \equiv 1 \pmod{4}$, the sequence defined by the congruence relation above has the full period m , Greenberger (113) and Rotenberg (112). This method of generating random numbers is known as the mixed congruential method. The special case, $c=0$ is called the multiplicative congruential method, or the power residue method. In this case, while we can no longer choose the integer a so that our sequence has the full period, we can obtain a period that is still quite large by choosing integer a such that the smallest value of n satisfying

$$a^n \equiv 1 \pmod{m}$$

is a maximum. This value of n is called the indicator of m . (see IBM(114)).

The mixed congruential methods have longer periods than the multiplicative congruential methods, but in either case, the periods are so long that this is not usually a real advantage. The disadvantage of mixed methods is in their statistical behaviour. Actually, this is generally good, but a small fraction of cases are completely unacceptable. There are no such cases amongst the multiplicative methods.

For the computer used in most of the simulations, the CDC 6600, three pseudo-random number generators are available:

$$\begin{aligned} \text{(a)} \quad x_i &= (2^{31} + 3) x_{i-1} \quad (\text{modulo } 2^{59}) \\ \text{(b)} \quad y_i &= (2^{29} + 3) y_{i-1} \quad (\text{modulo } 2^{59}) \\ \text{and (c)} \quad u_i &= (2^{29} + 3) u_{i-1} \quad (\text{modulo } 2^{57}) \end{aligned}$$

In (a) and (b) the starting integers x_0 and y_0 can be any odd positive integers less than 2^{59} . The multiplicative congruential method as has been discussed possesses a long cycle, good statistical properties, and is easy to implement on a digital computer. The statistical properties are claimed by the Control Data Corporation to be improved further by combining (a) and (b) in one random number generator.

5.7 Non-uniform continuous random numbers

A general requirement in simulations, is a sequence of random numbers drawn from a distribution that is both continuous and non-uniform.

If $f(x)$ is the given probability density function, consider the cumulative distribution $F(x)$.

$$F(x) = \int_{-\infty}^x f(x) dx$$

$F(x)$ is a monotone increasing function whose values lie in the range 0 to 1, Fig. 5.1 .

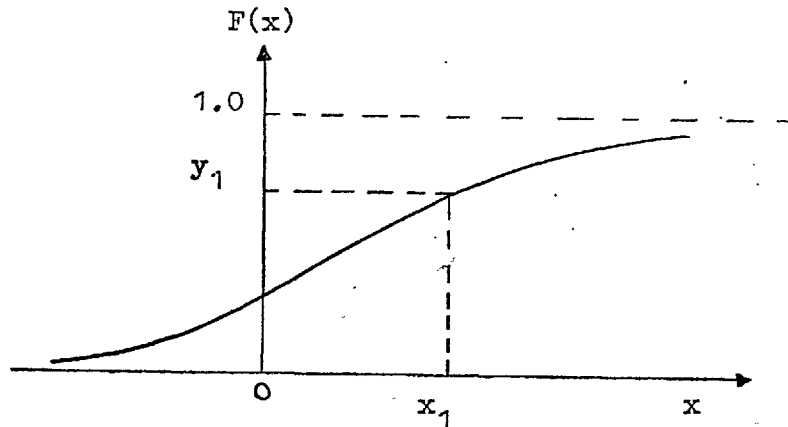


Fig. 5.1 The generation of non-uniform random numbers

Assume that we have a uniform random number generator and that we draw a number say y_1 . From the distribution function this corresponds to the value x_1 . If we draw a large number of y_1 the corresponding x_1 will be distributed according to the probability distribution function $F(x)$

5.7.1 The Exponential Distribution

The probability density function (p.d.f.) in this case has the form

$$p(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \quad t \geq 0$$

$$= 0 \quad t < 0$$

and the distribution function $F(t)$ is given by

$$F(t) = \int_0^t p(t) dt = 1 - e^{-\frac{t}{\tau}}$$

We can obtain the exponentially distributed random numbers by inversion as

$$t = -\tau \log_e (1 - F(t))$$

If each particular value of $F(t)$ is replaced by the random variable called RANDM, which is uniformly distributed in the open interval $(0,1)$, we can obtain exponentially distributed random numbers from RANDM by using

$$t = -\tau \log_e (\text{RANDM})$$

Because the exponential distributed is used so frequently in telephone traffic simulations, (it is not unusual to need of the order of 1,000,000 exponentially distributed random variables), we try to avoid evaluating the log function every time we need an exponentially distributed random

variable. The solution is to use table look-up procedures. We choose values a_i , $i = 1, 2 \dots N$ where N is arbitrary but fixed, say 1000, in the open interval $(0,1)$, work out $-\log_e(a_i)$, and tabulate as in Fig. 5.2

a_i	$-\log_e(a_i)$
0.001	
$\vdots$	
1.0	

Fig. 5.2 A random look-up table.

Now whenever we draw a value from the uniform random number generator, we look up the table and find that it lies between, say, a_j and a_{j+1} . The corresponding exponentially distributed random variable is then obtained by linear interpolation between $-\log_e(a_j)$ and $-\log_e(a_{j+1})$.

5.8 The Computer simulation of a telephone network.

The computer contains an image (a capacitated topological graph) of the network in which there is indicated for each group of circuits, their capacity and their instantaneous availability. Statistics are collected during the run. For example, the number of calls attempted between each pair of exchanges, and the number of calls which were unsuccessful due to the congestion of the circuits are noted. Thus we can obtain the point-to-point, as well as the overall (network) grade of service. Similarly we can obtain the trunk group utilities (erlangs per circuit).

As input to the computer, we have the offered traffic between each pair of exchanges, the mean holding time, the structure of the network, and a routing table that reflects the routing strategy being used. Fictitious calls are generated, spaced out in simulated time, between each pair of exchanges. A call has associated with it two attributes representing the originating and destination exchanges, and the holding time. As each call is made, we try to route it taking into account the routing rules and the instantaneous availabilities of the groups of circuits of the simulated network. If we succeed, the call acquires as many attributes as are required to specify the route used so that the connection can be cleared down at the end of the holding time. The network status is then updated. But if we are unsuccessful in placing the call, we note that a call has encountered congestion, update the statistics of unsuccessful calls, "destroy" the blocked call, and then proceed to the next most imminent event in the simulated time calendar.

Before ending this chapter, we present the method that has been used to verify the simulation results.

5.9 The verification of simulation results

The grades of service and link utilities obtained as simulation results are, in fact, just samples used to estimate the parameters of the distribution from which they are drawn. We must, therefore, test the significance of our simulation results.

The method of establishing the confidence intervals used in this study is based on two assumptions. First, that the distribution from which the observations are made is stationary, and secondly, that the observations are independent.

Typically 100,000 calls were processed in each single run. This run would be divided up into 10 intervals of 10,000 calls each, and for each interval i the grade of service b_i would be estimated. The average grade of service b and its variance S^2 are then estimated, for n intervals, by

$$b = \frac{1}{n} \sum_{i=1}^n b_i$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (b_i - b)^2$$

If the true mean and variance are given respectively by μ and σ^2 , the normalized random variable z given by

$$z = \frac{(b - \mu)\sqrt{n}}{\sigma}$$

has a Student-t distribution. To find the confidence level, usually expressed as a percentage, the Student-t distribution is integrated between $\mu - \alpha$ and $\mu + \alpha$ where $\pm \alpha$ is the confidence interval. In practice, the integration need not be done because the Student coefficients can be easily found in statistical tables.

CHAPTER 6. ROUTING IN A NETWORK OF EXCHANGES
AND RELATED PROBLEMS

6.1 Introduction

6.2 Routing in a network of exchanges

6.3 Some problems that are similar to the routing problem.

6.3.1 Network flow problems.

6.3.2 The inventory problem.

6.3.3 Markov processes with rewards.

CHAPTER 6. ROUTING IN A NETWORK OF EXCHANGES AND RELATED PROBLEMS

6.1 Introduction

Very little research work has been published on the routing of telephone traffic in a network of exchanges. This is in direct contrast to the voluminous literature on routing in a switch or network of switches, and the routing of store-and-forward (message) traffic.

Most telephone administrations are apparently content with their routing plans, and are therefore not researching into the construction and use of more efficient routing rules. The fact that the majority of the present telephone exchanges cannot be used in implementing complex routing strategies has also been a hindrance. But with the advent of Register Translators (RTs), it has become possible to divorce the routing plan from the numbering plan; the digits dialled by the subscriber need not be used directly in determining the routing of the call. But even with RTs, the information processing ability of exchanges has not been enough to handle complex routing plans. It is only with the introduction of Stored Program Controlled (SPC) exchanges that we have the data processing capacity that we need. The time has come to re-examine the efficiency of our routing plans.

6.2 Routing in a network of exchanges.

Some of the major research work on the routing of telephone traffic in a network of exchanges has been done by Wilkinson (9), Weber (30,31), Wright (29), Grandjean (1) and Wallstom (8).

The present study was, in fact, inspired by the analysis of the network of Fig. 6.1. by Grandjean (1).

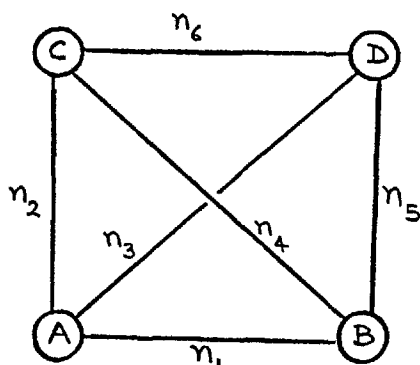


Fig. 6.1 A fully connected four-exchange network

This is a completely connected network of four exchanges A,B,C, and D, with trunk groups having $n_1, n_2, n_3, n_4, n_5, n_6$ two-way trunks. The (symmetrical) Connection matrix C is therefore.

$$C = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & n_1 & n_2 & n_3 \\ n_1 & 0 & n_4 & n_5 \\ n_2 & n_4 & 0 & n_6 \\ n_3 & n_5 & n_6 & 0 \end{bmatrix} \end{matrix}$$

And the offered traffic matrix T is given as

$$T = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & t_{AB} & t_{AC} & t_{AD} \\ t_{BA} & 0 & t_{BC} & t_{BD} \\ t_{CA} & t_{CB} & 0 & t_{CD} \\ t_{DA} & t_{DB} & t_{DC} & 0 \end{bmatrix} \end{matrix}$$

Where t_{ij} , $i, j = A, B, C, D$ means that exchange i offers exchange j t_{ij} erlangs. The following four assumptions were made in Grandjean's study:

1. All trunk groups have an equal number n of trunks, i.e. Fig. 6.1 is a "symmetrical" network.
2. Each exchange offers every other exchange an equal amount of traffic a erlangs (The network is "symmetrically" loaded).
3. The network has no hierarchy, and each exchange is fully informed about the state of the network as defined by the instantaneous number of engaged circuits in each trunk group.
4. All exchanges can act as transit centres.

The problem then was to devise and investigate various alternate routing rules. The network performances corresponding to the various routing rules were then compared to the performance obtained when no alternate routing was permitted.

The results obtained were of the form shown in the graph of Fig. 6.2

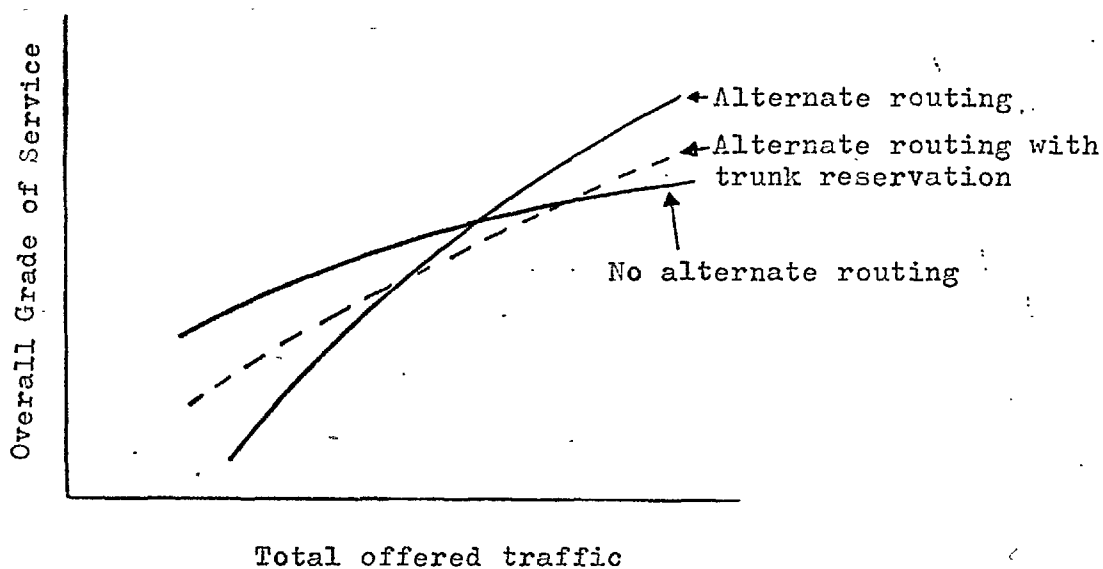


Fig. 6.2 A comparison of routing strategies

The important observation to be drawn from Fig. 6.2 is that the alternate routing strategy produced results that were better than those obtained with no alternate routing up to a certain critical value of offered load. But above that the alternate routing strategy gave results that were worse than those obtained without using alternate routing. The use of the technique of trunk reservation reduced this grade of service deterioration considerably.

The three-exchange network of Fig. 6.3 behaves in a manner that is similar to that of the four-exchange network.

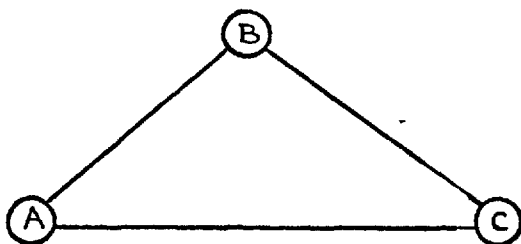


Fig. 6.3 The three-exchange network

Both these networks will be reviewed in more detail in Chapter 7. The important problem that is outstanding at this stage is whether we can devise routing strategies that give a performance that is always better than that obtained without the use of alternate routing regardless of the level of offered traffic.

Weber (30,31) has studied extensively the traffic characteristics of telephone networks using automatic alternate routing. Using computer simulation, Weber and Gimpelson (84), he observed the performance of the sequential alternate routing strategy. His main conclusion was that several critical problems in the design and utilization of telephone networks are still outstanding. One of these is "our lack of understanding of the behaviour of alternate routing strategies under overload conditions, both local and system-wide". This lack of understanding is revealed by the two problems:

1. How can we design networks that are efficient during normal operation, but which do not deteriorate catastrophically under abnormal loads?
2. If we are given a network design, how can we best alter the switching pattern, for the duration of an overload, so as to improve the performance?

In this thesis, we re-formulate and outline a possible solution to these problems.

Wilkinson (9) made an important contribution to routing studies by establishing his Equivalent random theory, and by studying the characteristics of overflow traffic. His studies showed that overflow traffic is rough, and that it can be adequately described using the negative Binomial distribution. Wallstrom (8) has followed in Wilkinson's footsteps.

We conclude this very brief survey by emphasizing how much routing studies owe to such papers as that of Wright (29) for motivation, guidance and perspective.

6.3. Some problems that are similar to the routing problem

In this section I propose to draw the reader's attention to three classes of problem that have a close similarity to the routing problem:

1. Network flow problems.

2. The inventory (or storage) problem.
3. The problem of Markov processes with rewards.

6.3.1 Network flow problems

The methods of network flow, Hu(4), Ford and Fulkerson (5), have been used by several authors, Silk (6), Ling (7), Gomory and Hu(15), Onaga(16), Frank and Hakimi(17), Kleinrock(75), and Hakimi(83), to mention only a few, in the analysis of the routing of message traffic.

The problem of immediate interest in this thesis, is the design of routing rules that would enable us to use the telephone line plant most effectively, thereby enhancing the network's telephone traffic carrying capabilities. Similar problems have engaged students of network flow theory for a long time. The results and insights obtained from these studies have enabled students of message switching to apply network flow methods to message routing problems with great success. I have, therefore, examined these methods to find out if they provide a suitable model or framework for our studies. If they do not, then we must set down clearly why not, and point out the modifications required to make them more suitable.

It is important at the outset to make a clear distinction between the natures of telephone and message traffic. The vital difference between them is that telephone traffic is 'live' whereas message traffic is not. We say 'live' in the sense that a dialled call must be set up on demand (we are excluding manual connection or delay working). This means that if a call requires three links in tandem, it must have them all (almost at once) before the call is completed. In contrast, in a message switching network, a message can be held at the originating exchange, or a transit exchange if there are no free available links, and then sent on later when free paths in the required direction are available.

This vital difference between 'live' and 'dead' traffic leads to a difference in the service criterion to be applied in each network. For telephone traffic we judge the quality of the service by, among other things, the "grade of service", i.e. the fraction of attempted calls which get the 'busy signal' as a result of congestion in the network. For message traffic it is the expected delay between initiating a message and its being received at its destination that defines the quality of service.

Let us examine the multi-commodity network flow problem to see if we have a suitable framework for formulating our telephone traffic problem.

The multi-commodity flow problem

We are given a topological graph $G = (N; A)$ of nodes or exchanges N_i , and arcs or trunk groups A_{ij} . A commodity flow from N_i to N_j is defined as the distinct flow from node N_i to N_j . The problem is to maximize the sum of the commodities flowing between all pairs of nodes.

This problem has been solved using the arc-chain formulation, Hu(4,49) Ford and Fulkerson(80), Tomlin (81), and Grinold(87,88). So let us try and set our problem in the framework provided by this problem. The nodes of the graph represent telephone exchanges, and the arc capacities represent as before the number of circuits in a link.

The point of departure between the multi-commodity problem set up and the telephone traffic problem is in the initiation of flows. The multi-commodity problem tacitly assumes that we have a very large (ideally infinite) amount of commodity that we wish to transfer from any node N_i to any other node N_j , and the demand is constant and deterministic. The situation is completely different for the telephone traffic problem for which the demand is stochastic.

Another important difference is that while in the multi-commodity flow problem we are interested in maximizing the total commodity flow in the network, this is not the case in a telephone network where service considerations require that every parcel of traffic should be given an equitable service.

In the feasibility problem of multi-commodity flows, the flow values of each of the commodities is prescribed, and the question then is whether or not all these flow values can be realised simultaneously in the given network without violating the arc capacities. Hu(4) has given a solution to this problem. Because this is a feasibility problem, the solution does not necessarily give us any guidance on how the flows should be set up. Nevertheless, this feasibility problem is important because we can now divide up the time axis into the intervals $(0, t_1)$, (t_1, t_2) during which we have a fixed set of requirements and we want to know if they can all be satisfied simultaneously. In telephone systems, we do not set out to satisfy each and every one of the subscriber's demands. We are content if we are able to satisfy a prescribed percentage of these demands. As we have already noticed, this idea of a grade of service is conspicuously absent in network flow methods.

It is now clear that before we can use the network flow methods, we must incorporate into them the concept of grade of service, and a method for dealing with stochastic demands.

6.3.2. The inventory problem

Consider the problem of routing a call between exchanges A and D in the four-exchange network of Fig. 6.4

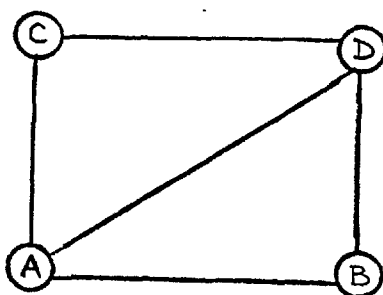


Fig. 6.4

The obvious choice of route is the direct link AD. But if this route is congested, we may try routes ACD and ABD. Now it may be useful to consider each trunk group, AB, BD, etc. as an inventory. Each of these inventories is subjected to a stochastic demand because of the stochastic arrival of telephone calls. Each inventory has a "home" demand, and a "foreign" demand. Take the trunk group AD. The home demand consists of the A to D and D to A calls. And the foreign demand consists of all other calls that use AD as an overflow trunk group. The "re-ordering policy" of the inventory depends on the release of the circuits as the calls terminate. If a home demand is not satisfied, the inventory incurs a penalty. When a foreign or home demand is satisfied, the inventory makes a gain. But if a foreign demand is not satisfied no penalty is incurred.

We now wish to determine the policy that will maximise the average gain per demand. Inventory problems of this kind are unusual, Moran (149), because there is a home as well as a foreign demand. This type of inventory problem, as far as the author knows, has not been solved.

6.3.3 Markov process with rewards

Problems involving Markov process with rewards are similar to inventory

problems in that they use the idea of average gain per state transition. Let us re-examine the routing of the A to D call. Routing the call via exchange C leads to a different state from that obtained by routing the call via B. If for every state transition we define a reward, we may well ask for a policy of making transitions that will result in the maximum gain per transition. Howard (152) has worked out an elegant method, called policy iteration, for dealing with this problem for the case of all demands being home demands. If only we had a method for dealing with both a home and foreign demand?

CHAPTER 7. ROUTING IN A FOUR-EXCHANGE NETWORK

- 7.1 Introduction
- 7.2 The direct routing strategy
- 7.3 Alternative routing strategies
 - 7.3.1 The three-exchange network
- 7.4 The four-exchange network
 - 7.4.1 Dynamic routing strategies
 - 7.4.2 The hyper-diagonal routing strategy
 - 7.4.3 The min-max routing strategy
 - 7.4.4 The random routing strategy
- 7.5 Conclusion

CHAPTER 7. ROUTING IN A FOUR-EXCHANGE NETWORK

7.1 Introduction

Up to now, we have considered routing strategies in a network of switches. We have given as examples the Clos network and the 6 to 4 access concentrator. It is important to notice that these routing problems have been solved by writing down and resolving the state equilibrium equations. For the access concentrator, a solution by linear programming was also included. The number of states in these examples was small thus simplifying the process of state assignment and state reduction. The process of writing down and solving the equilibrium equations was also similarly simplified.

In turning our attention to networks of exchanges, as opposed to networks of switches, we find that even the simplest networks have a very large number of states making it extremely difficult to apply the methods we have used so far. (For some state variable solutions with a large number of states, see Kibble (135)). In a search for alternative methods of solution, let us start with a simple three-exchange network, and then consider the four-exchange network in more detail.

7.2 The direct routing strategy.

The diagram of Fig. 7.1 shows a three-exchange network, in which the exchanges A, B, and C are connected by the two-way trunk groups AB, AC, and BC having n_1 , n_3 and n_2 trunks respectively.

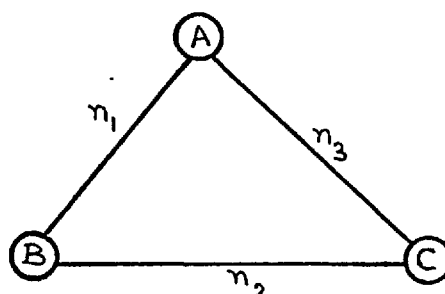


Fig. 7.1 The three-exchange network.

Let us designate as a direct routing strategy a routing rule that prohibits any alternate routing. In our three-exchange network of Fig.7.1,

the direct routing strategy requires that only the direct one link route be used for any call. In a network in which some pair of exchanges has no direct link, a basic, multi-link route is assigned to the traffic between the two exchanges. That multi-link route, some links of which may be shared by other parcels of traffic, is then considered as the direct route between the pair of exchanges. This is the routing plan, for example, in a hierarchical network in which no alternate routing is used.

Assume that the sum of the traffic offered to B by A, and the traffic offered to A by B equals a_1 erlangs. Thus a_1 erlangs are offered to the n_1 trunks of AB. Similarly a_2 erlangs are offered to BC, and a_3 erlangs offered to AC. The probabilities of congestion encountered by these three (isolated) parcels of traffic are, by definition, given by Erlang's formula

$$E_n(a) = \frac{\frac{a^n}{n!}}{\sum_{k=0}^n \frac{a^k}{k!}} \quad \begin{array}{l} n = \text{number of trunks} \\ a = \text{offered traffic} \end{array}$$

The overall (network) grade of service B for the three-exchange network using the direct routing strategy is given by

$$B = \frac{\sum_{i=1}^3 a_i E_{n_i}(a_i)}{\sum_{i=1}^3 a_i}$$

7.3 Alternative routing strategies

This kind of routing that attempts to place a call by trying out two or more possible paths in a given order until either a free path is found, or all the alternatives are exhausted is known as a sequential alternative routing strategy. This strategy can be described in more detail as follows. When a call originates at an exchange, an immediate attempt is made to establish a communication path to the required exchange along a direct route if one is available. If the direct route is busy or otherwise not available, the route translators step through a pre-determined sequence of alternate routes to try and complete the call. If the direct, and a limited number of alternate routes are all occupied, a busy tone is

then returned. An exchange that can choose an alternate route when the most preferred route is not available, is said to have automatic alternate routing capabilities. In the three-exchange network, a call from A to C will be offered to route AC first. If AC is congested, an attempt will be made to use the alternate route ABC provided trunk lines are available on both AB and BC. A similar procedure is used for a call between any other two exchanges. Because the network has no high usage routes, we can say that we are using mutual alternate routing.

It is difficult to determine analytically the exchange-to-exchange and the overall grade of service of the network when the sequential alternate routing strategy is used. A very rough approximation can be obtained by using Erlang's formula and the Lee-le Gall theorem, Grandjean (1), and Syski (22). After comparing the results obtained using this analytical method with those obtained by simulation, I came to the conclusion that the grade of service is better estimated using simulation than analytical methods. Lately, S.S. Katz has proposed a much more accurate method for computing overall and exchange-to-exchange grades of service, Katz (98). This method is applicable to sequential alternate routing networks using step-by-step (progressive) routing. Katz has claimed that the results obtained by his method are just as accurate as those obtained by simulation, and that they are computed in a fraction of the time.

In this thesis, we are using right-through (end-to-end) routing, as opposed to step-by-step routing. Katz's method has therefore not been used.

The overall (network) grade of service B is determined from the exchange-to-exchange grades of service B_i and offered traffic a_i as

$$B = \frac{\sum_i a_i B_i}{\sum_i a_i}$$

7.3.1 The three-exchange network

The three-exchange network of Fig. 7.1, with $n_1 = n_2 = n_3 = 50$ and $a_1 = a_2 = a_3$ in the range 38 to 48 erlangs, using the sequential alternate routing strategy was simulated on a computer. The results,

together with those of the direct routing strategy, are plotted in Fig. 7.2 .

Two important observations can be drawn from the graph of Fig. 7.2:

1. The use of sequential alternate routing improves the overall grade of service for all traffic levels up to an overload of 25 per cent, from what it is when the direct routing strategy is used. (The nominal rated load is defined as that load which gives an overall grade of service of 0.002 under the direct routing strategy).
2. For overloads above 25 per cent, the performance of the alternative routing strategy is inferior to that of the direct routing strategy.

As far as the routing problem is concerned, this observation makes it clear that above 25 per cent overload it is better to reject a call rather than place it on a two-link route even if that route is available. The explanation of this poor overload performance is not difficult to find. The use of a two-link route for a single call is, in general, inefficient. At low traffic loads, this inefficiency is preferable to rejecting a call. This is so because at this low traffic level the average trunk utilities (carried erlangs per trunk) are very low, with most of the trunks being free most of the time. At the high traffic loads, on the other hand, there are much fewer trunks that are free for any sizeable proportion of the time. At this traffic level it does not pay to engage two trunks for a single call when each of these trunks is likely to be needed for carrying a call in its own direct group.

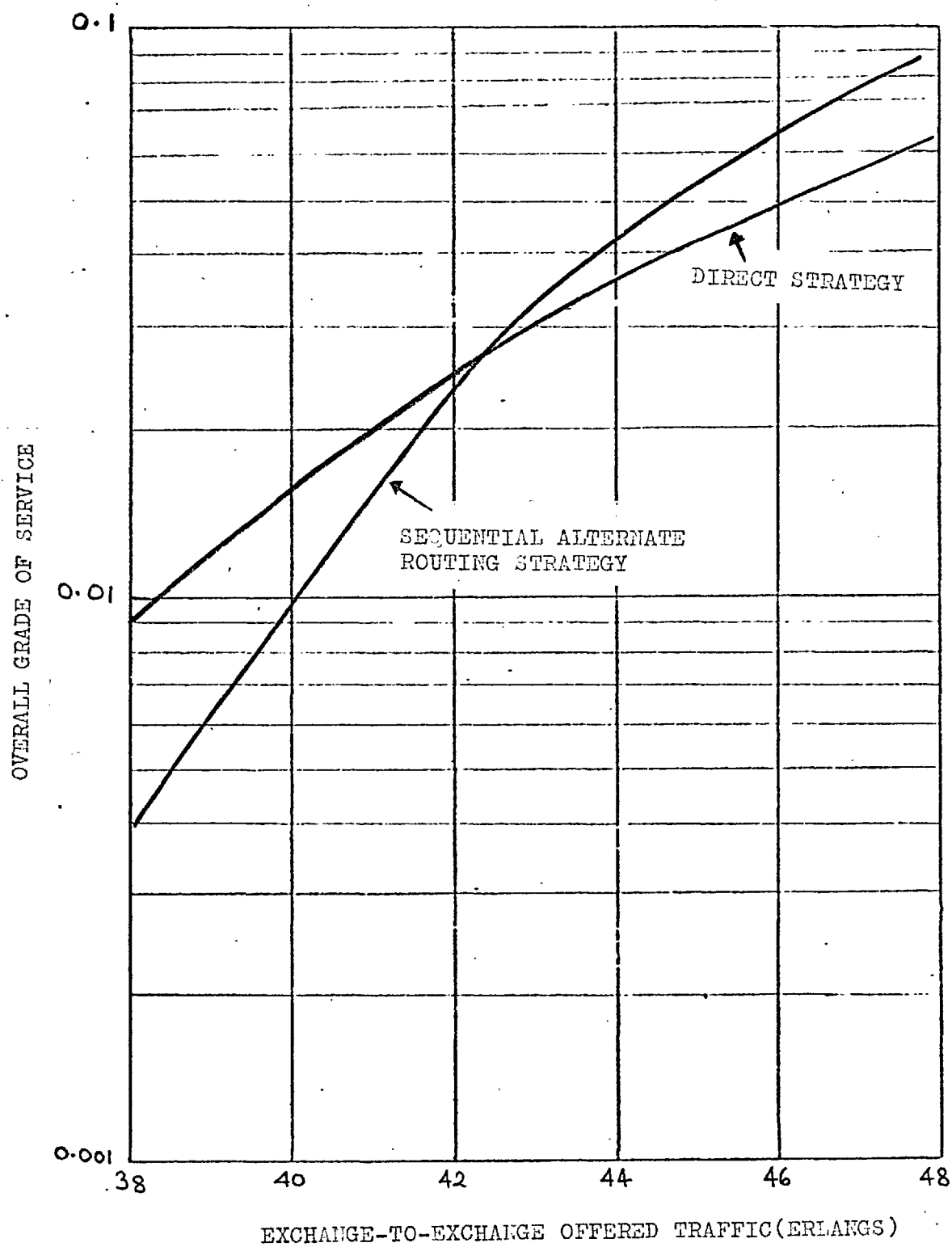


Fig.7.2 Overall grade of service for the three-exchange network

Fig. 7.2 was drawn for the case when all trunk groups had 50 circuits. In that case, the alternative routing strategy started giving a performance inferior to that given by the direct routing strategy at an overload of 25 per cent. Similar simulations were done for the three-exchange network with trunk groups having a number of circuits in the range 10 to 100. The overloads at which this cross-over happens were observed. Fig. 7.3 shows this "threshold" overload plotted against the number of circuits in each trunk group. For example, deterioration sets in at about 14 per cent overload for a group of 100 circuits.

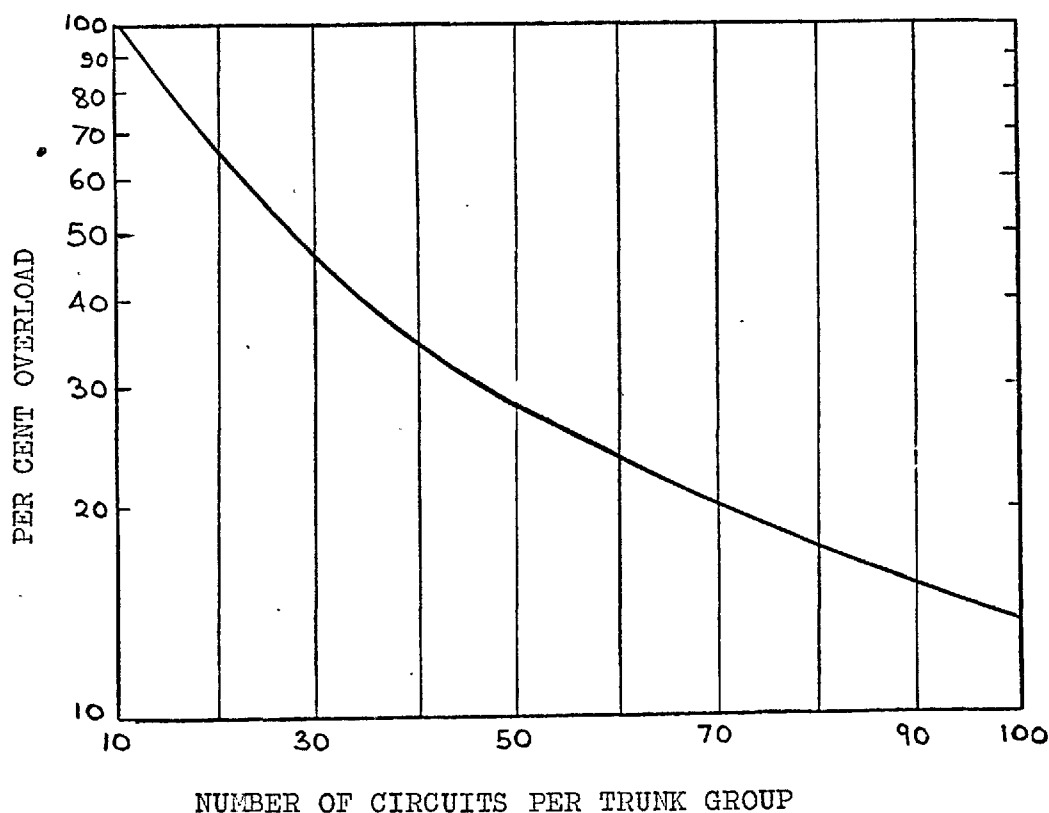


Fig. 7.3 Per cent overload at the cross-over

7.4 The four-exchange network

In the last section, we considered routing in the three-exchange network. This is the simplest network for which any routing problem exists. But because of its simplicity, it allows only two possible routes for any

call, and it is quite clear from the outset that it is preferable to use the direct route, if it is not congested, than to use a two-link route.

What emerged as being very interesting was the fact that at high overloads it is better to reject a call rather than place it over a two-link route. We now move on naturally to investigate the four-exchange network.

The four-exchange network of Fig. 7.4 provides many more possible routes for each call than the three-exchange network.

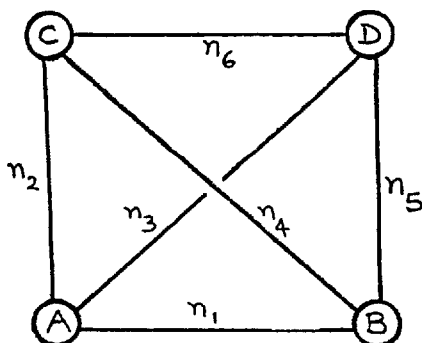


Fig. 7.4 The four-exchange network.

Thus a call from A to B has five possible routes:

1. A - B
2. A - C - B
3. A - D - B
4. A - C - D - B
5. A - D - C - B

Previous studies of this and similar networks, Grandjean (1), have shown that choices (4) and (5) lead to routes that are too circuitous, engaging too many links for a single call. By the same reasoning used in Section 7.3.1, this would lead to very severe deterioration in the grade of service. For this reason choices (4) and (5) were disallowed without any further investigation. This leaves the direct route and the two two-link routes as the possible choices.

Again we observe that the direct route is the best route to use if it is not congested. But it is not at all obvious in general which of the two two-link routes one should choose if they are both available and the direct route is engaged.

Before examining various routing strategies, let us make the assumption

that each exchange offers every other exchange $a/2$ erlangs, and that each trunk group has n circuits.

7.4.1 Dynamic routing strategies

The sequential strategy used for routing in the three-exchange network is a static strategy in the sense that it does not take into account the number of circuits in use in all trunk groups at the instant when the choice of a route is made.

A dynamic routing strategy, on the other hand, pays particular attention to the number of busy circuits in each one of the groups that might be used for routing a given call.

7.4.2 The hyper-diagonal routing strategy

From the diagram of Fig. 7.4 let us abstract the two routes ABC and ADB as in Fig. 7.6

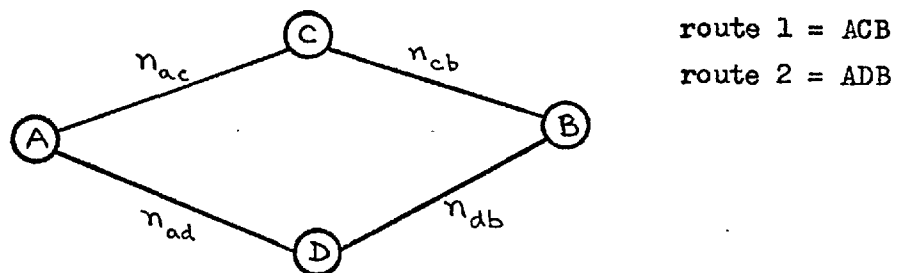


Fig. 7.6

Assume that the following information is given:

1. The number of busy circuits in the links AC, CB, AD, BD, is given by n_{ac} , n_{cb} , n_{ad} , and n_{bd} respectively.
2. The number of calls in progress, at a given instant, between exchanges A and C, C and B, A and D, and D and B are given by l_{ac} , l_{cb} , l_{ad} , and l_{db} .

Since the network structure is symmetrical, i.e. all trunk groups have n circuits, and since the loading is also symmetrical, each exchange offers every other exchange $a/2$ erlangs, it is reasonable to expect that the state of the network defined by the six dimensional state vector $(n_{ab}, n_{ac}, n_{ad}, n_{bc}, n_{bd}, n_{cd})$ should lie in the neighbourhood of the hyperdiagonal connecting the points $(0,0,0,0,0,0)$ and (n,n,n,n,n,n) .

An heuristic routing strategy was therefore constructed in which a call would be routed in such a way that the distance between the resulting points $(n_{ab}, n_{ac}, n_{ad}, n_{bc}, n_{bd}, n_{cd})$ and $(l_{ab}, l_{ac}, l_{ad}, l_{bc}, l_{bd}, l_{cd})$ would be a minimum, i.e.

$$\min \sum (n_{ij} - l_{ij})^2 \quad i, j = a, b, c, d \text{ and } i \neq j$$

We would then keep as close to the leading hyper-diagonal as possible and try to even out any temporary traffic imbalances.

No computation procedure was found for evaluating this routing strategy even approximately, so an event-by-event simulation was set up on a digital computer. The simulation results together with those of the direct strategy are plotted on Fig. 7.5

Fig. 7.5 bears a very close resemblance to Fig. 7.2, and the general observations made on Fig. 7.2 hold for Fig. 7.5

7.4.3 The Min-max routing strategy

Consider the two possible two-link routes of Fig. 7.6 Let,

$$m_1 = \max (n_{ac}, n_{cb})$$

$$m_2 = \max (n_{ad}, n_{db})$$

$$\text{and } k = \min (m_1, m_2)$$

If $k = m_1$, we route the AB call via exchange C. If, on the other hand, $k = m_2$ we route the call via D. Should $m_1 = m_2$, we would choose between route ACB and ADB at random. This is the min-max routing strategy. Simulation results using this strategy in a four-exchange network with $n = 10$ are compared to those of the direct routing strategy in Fig. 7.7

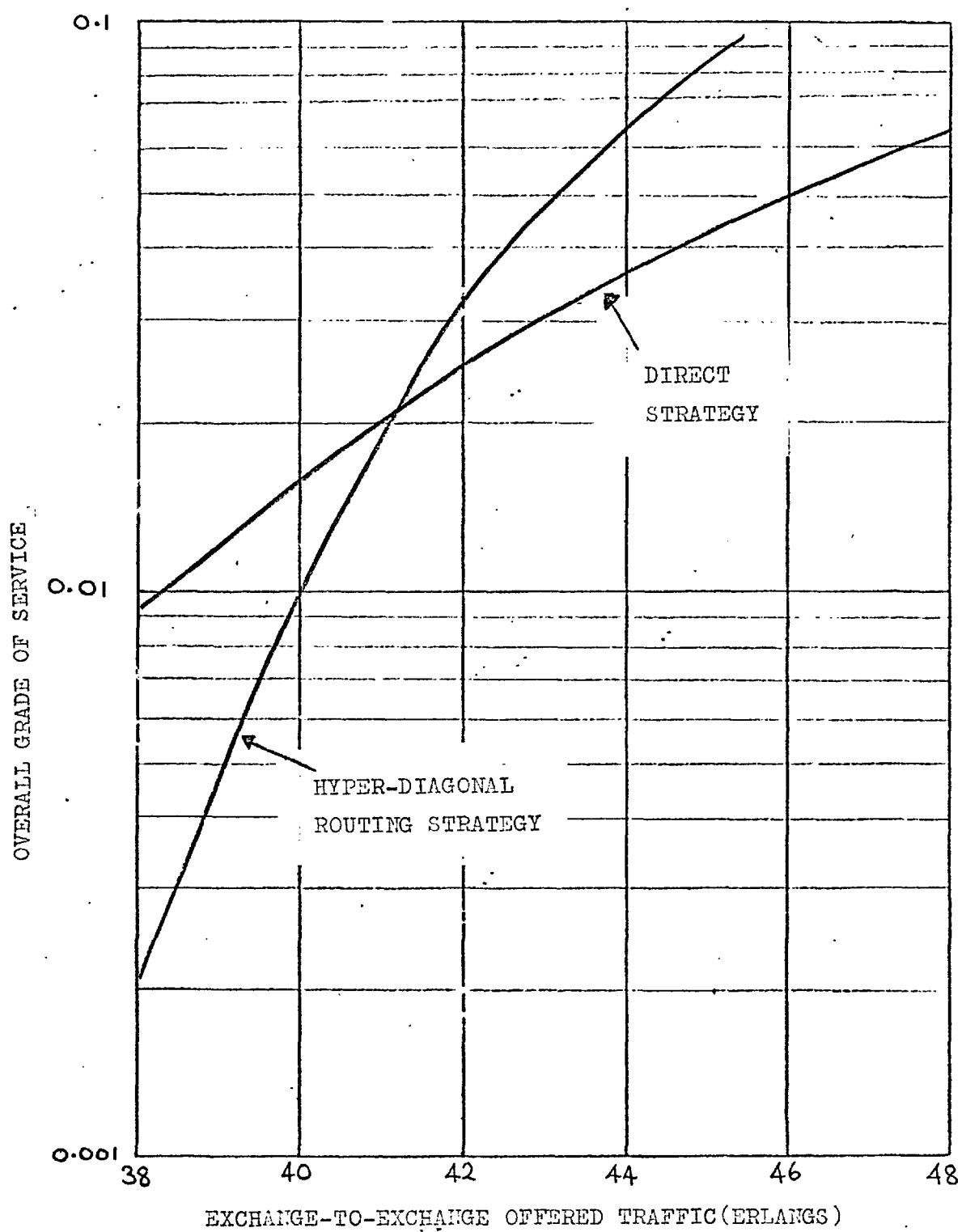


Fig.7.5 Overall grade of service for the four-exchange network

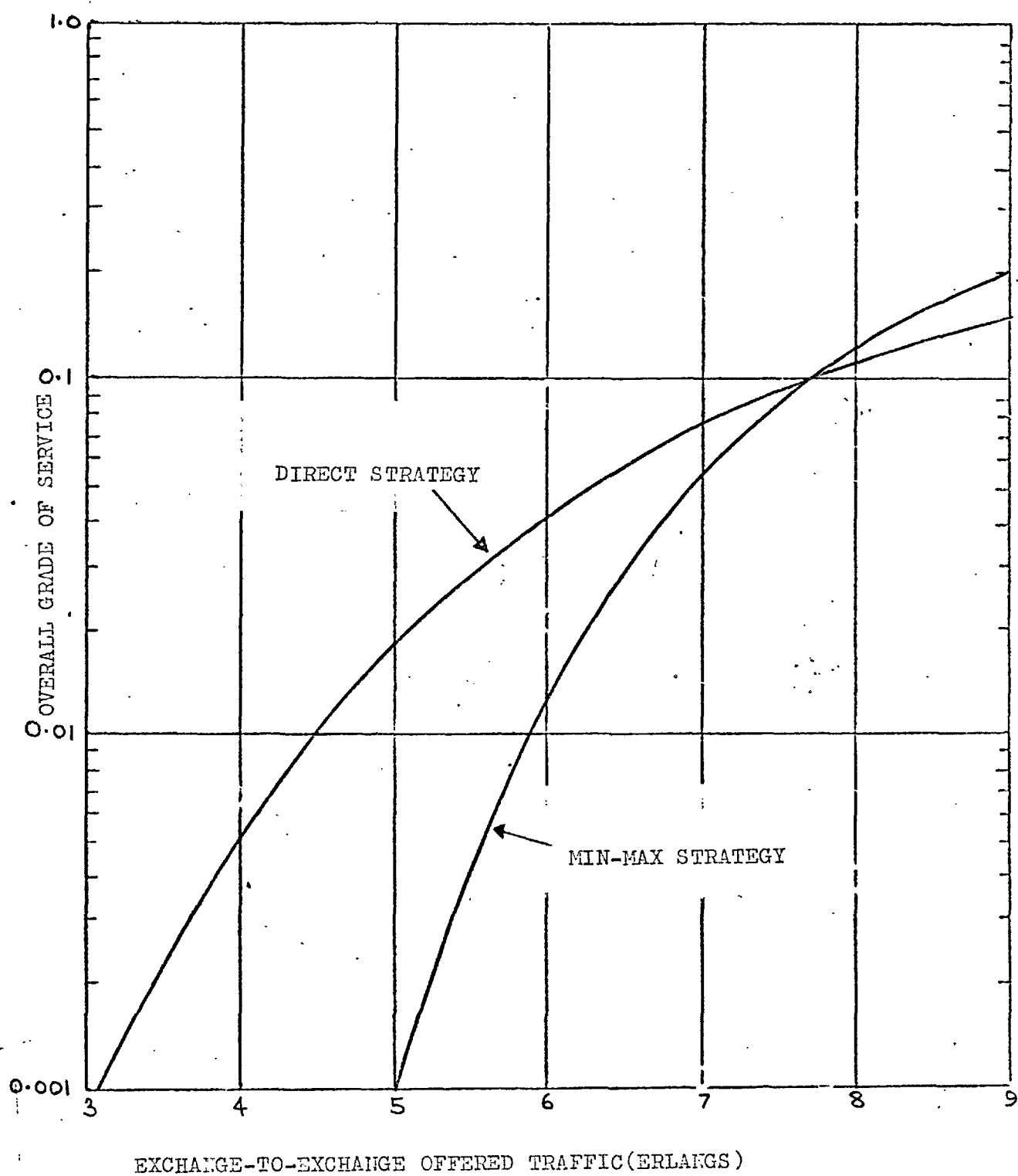


Fig. 7.7 Overall grade of service for the four-exchange network
(10 circuits per trunk group)

One interesting observation is that the results of the min-max routing strategy are almost exactly the same as those obtained by using the hyper-diagonal routing strategy as shown in the table below.

Exchange-to-exchange Offered Traffic (erlangs)	Overall grade of service	
	Min-max Strategy	Hyper-diagonal Strategy
4.0	0.000	0.000
5.0	0.001	0.001
5.5	0.004	0.004
6.0	0.012	0.013
6.5	0.029	0.031
7.0	0.056	0.060
8.0	0.133	0.134
9.0	0.209	0.213

N = 10

This near equality is to some extent expected, since both the routing strategies aim at getting a uniform circuit usage throughout the network.

Trunk Reservation

A persistent feature of the alternative routing strategies that we have looked at so far, is their rapid deterioration in performance at high overloads. This phenomenon has been observed by other workers, e.g. Weber (31) and Grandjean (1). One method of dealing with it, is to cancel any alternative routing at a pre-determined traffic level. But this has the disadvantage that it would give rise to unacceptably high probabilities of congestion for those calls which depend on alternative routing to overflow from their high usage routes on to other (perhaps final) routes. A better method of dealing with this service deterioration is to use the technique known as trunk reservation.

Let us apply trunk reservation to our four-exchange network with $n = 50$. If a call cannot be accommodated on the direct route, it is only allowed to use a two-link route if there are at least m free circuits on each one of the two links. The value for m may be 2, 3 or more; in general m has to be optimised. By using trunk reservation, we are, in fact, reserving the last $m - 1$ circuits for direct traffic only. The case $m = 1$ means that there is no trunk reservation.

Simulation results obtained using trunk reservation with $m = 4$ are shown in Fig. 7.8. The traffic level at which degradation sets in has now been raised from 25 per cent to 35 per cent overload, and the performance once degradation has set in is now much better than it was without trunk

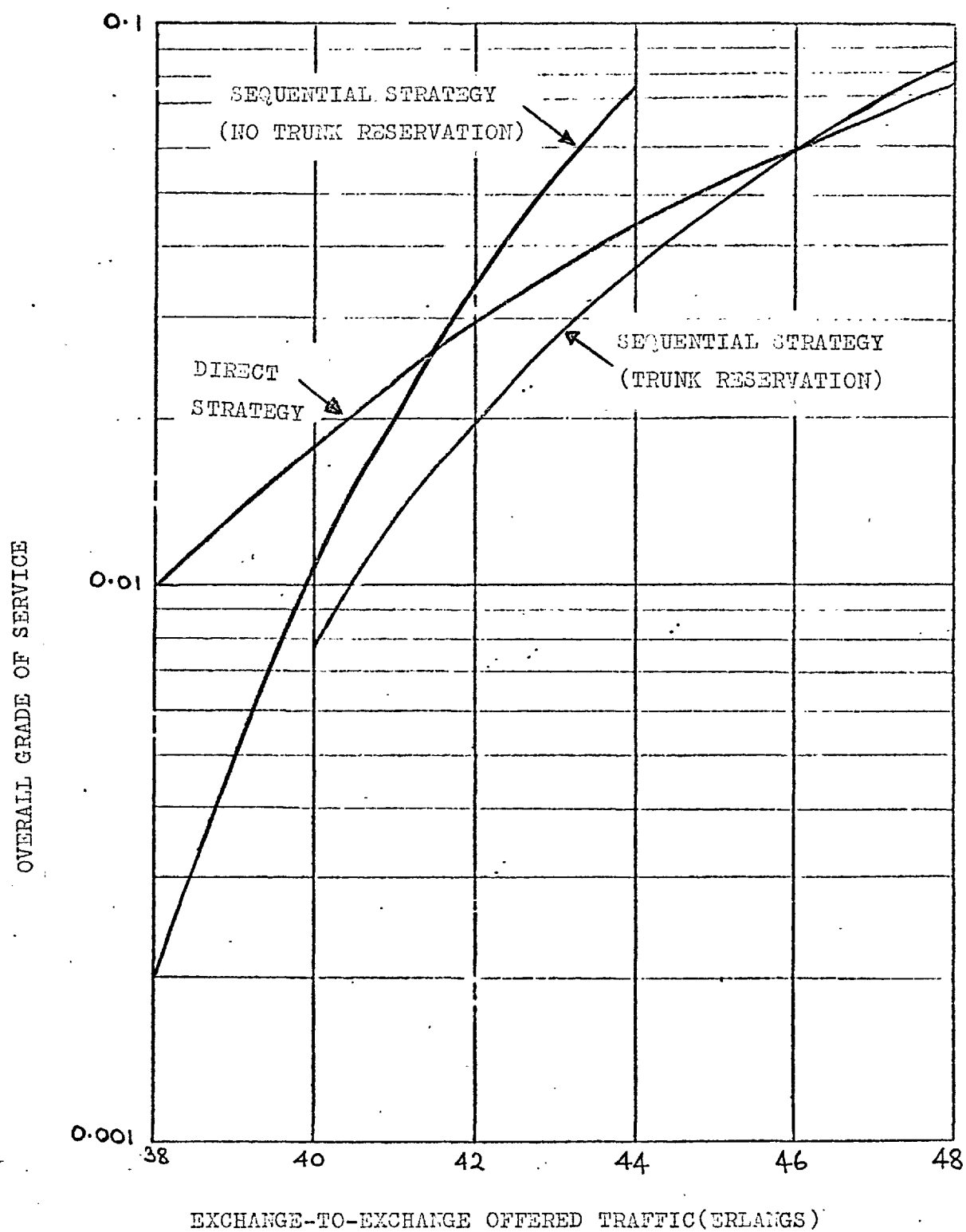


FIG. 7.8 Overall grade of service for the four-exchange network
(50 circuits per trunk group)

reservation.

7.4.4 The random routing strategy

In a symmetrical network that is symmetrically loaded, the state vector tends by itself to lie fairly close to the hyper-diagonal. This means that the hyper-diagonal routing strategy is given little opportunity to show its usefulness, and the degenerate case of $m_1 = m_2$ is presumably a frequent occurrence. This conjecture was tested by setting up a simulation in which the probability of choosing route ADB was made equal to the probability of choosing route ACB if they were both simultaneously available. The results obtained are, as one would expect, almost exactly the same as those obtained by the hyper-diagonal and min-max routing strategies.

7.5 Conclusion

It is clear that both the min-max and the hyper-diagonal routing strategies as so far presented, depend for their success entirely on the network's structural and loading symmetry.

But there is no reason why they cannot be generalised to networks which are both unsymmetrical in their structure and loading. Nevertheless, it seems that more fruitful results can be obtained by studying more carefully the stochastic behaviour of a group of circuits from a routing point of view.

CHAPTER 8 STOCHASTIC BEHAVIOUR OF A GROUP OF CIRCUITS:
THE TIME TO CONGESTION CONCEPT.

- 8.1 Introduction.
- 8.2 Transition probabilities in the occupancy of a group of trunks.
- 8.3 The computation of the stationary probabilities.
- 8.4 The fundamental matrix.
- 8.5 The first passage time.
 - 8.5.1 The mean first passage time.
 - 8.5.2 The variance of the first passage time.
 - 8.5.3 The probability distribution of the first passage time.
- 8.6 The stochastic behaviour of a group of circuits offered non-random traffic.
 - 8.6.1 The transition matrix for smooth traffic.
 - 8.6.2 The transition matrix for rough traffic.
- 8.7 The first passage time distribution in continuous time.
- 8.8 The probability density function of $Z = X - Y$
 - 8.8.1 The probability of a routing error.
- 8.9 Conclusion.

8.1 Introduction

In Chapter 3 we determined the probability that a call offered to a group of trunks will find all the trunks engaged. This is the probability that a call is lost; and for random traffic it is given by Erlang's formula. In this chapter we shift our emphasis. Suppose x trunks are busy at time t . What is the average time interval before y trunks are busy for the first time after t ? This is, in fact, the first passage problem that has been extensively studied in the theory of Markov processes. In this chapter we concentrate on the average passage time, and the probability distribution of the passage time. We look at the problem first in discrete time, and then in continuous time.

But first let us be quite clear as to why we need to solve the first passage time problem. Let us take the simple example of the three-exchange network of Fig. 8.1.

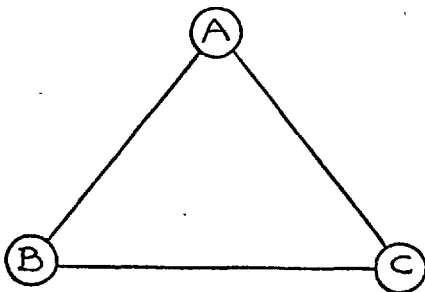


Fig. 8.1 The three-exchange network

Consider the call from exchange B to C. If the direct route BC is congested, we may try the two link route B-A-C. In Chapter 7, Section 7.3.1 we have seen that this is advantageous if the traffic level in the network is low, and disastrous if it is high. We have given the qualitative argument that at high traffic levels it is unwise to use up two links for a single call because the demand for circuits is very high. It is therefore better to reject the BC call even if it is possible to place it via exchange A. Yes, but at what traffic level should we start rejecting calls? The answer to this question can only be found re-examining the stochastic behaviour of a group of trunks.

Suppose each of the trunk groups in Fig. 8.1 has N trunks. We are

then vitally interested in the first passage time from any state where j ($j < N$) trunks are engaged to the congestion state when all trunks are busy. This we have called the "time to congestion". It is this time to congestion that is of prime importance in routing. If we use the route B-A-C, what is the time to congestion for B-A and A-C? How likely is it that either B-A or A-C or both will pass through the congestion state while the call we are considering is in progress? To my knowledge these questions have not been answered up to now. This Chapter provides some answers.

8.2 Transition probabilities in the occupancy of a group of trunks

Consider a group of N trunks offered λ erlangs. We let, without any loss of generality, the mean holding time be our unit of time. For the case of random (Poisson) traffic, the mean of the offered traffic is equal to its variance. The probability of a call being offered to the group in a small interval of time dt is given by

$$\text{Prob}(1 \text{ call offered in time interval } dt) = \lambda dt + o(dt)$$

$$\text{where } \lim_{dt \rightarrow 0} \frac{o(dt)}{dt} = 0$$

$$\text{Prob}(2 \text{ or more call offered in } dt) = o(dt)$$

We define the state of the group of trunks as the number of trunks that are simultaneously in use at a particular epoch of time t . There are thus a total of $N + 1$ states ranging from state 0 to state N .

If the trunk group is in state j , $j \neq 0$, $j \neq N$, only three events can take place during a small interval of time dt :

1. A call may terminate.
2. A call may arrive.
3. The state may remain unchanged.

Assume that the holding time is characterised by the negative exponential distribution with mean $\mu = 1$. Then the probability of one call among n calls in progress terminating during dt is

$$\text{Prob}(1 \text{ call out of } n \text{ calls terminates in } dt) = n \mu dt + o(dt)$$

Using the standard notation of stochastic processes, the probability of a state transition from n to $n+1$, from n to $n-1$, and the probability of no transition are given by

$$p_{n, n+1}(dt) = \lambda dt + O(dt)$$

$$p_{n, n-1}(dt) = n\mu dt + O(dt)$$

$$p_{n, n}(dt) = 1 - \lambda dt - n\mu dt + O(dt)$$

For the boundary states 0 and N , we have the following transition probabilities.

$$p_{0,1}(dt) = \lambda dt + O(dt)$$

$$p_{0,0}(dt) = 1 - \lambda dt + O(dt)$$

and

$$p_{N, N-1}(dt) = N\mu dt + O(dt)$$

$$p_{N, N}(dt) = 1 - N\mu dt + O(dt)$$

For convenience we can arrange these state transition probabilities in an $N+1$ by $N+1$ square matrix P .

$$P(dt) = (p_{ij}(dt)) \quad i, j, = 0, 1, \dots, N$$

The stochastic behaviour of the group of trunks can now be studied in terms of the transition matrix P .

The Matrix P represents a regular Markov process with discrete states in continuous time, i.e. it is possible to be in any state after some time interval T no matter what the starting state was. The imbedded Markov Chain can be revealed by giving dt a small numerical value (the value of dt depends on λ) say $dt = 0.01$, thus turning the continuous time variable into a discrete time variable.

Having so reduced the problem, we can now proceed to use the standard (finite) Markov Chain methods for the rest of the investigation, Feller (2,3) Chung (36,37,38), Cohen (122), and Kemeny and Snell (35).

8.3 The computation of the stationary probabilities

By definition, a Markov Chain is regular if and only if (iff)

it is possible to be in any state after some number of steps N , no matter what the starting state is.

It is a well known result that P^n has a limit as $n \rightarrow \infty$

$$A = \lim_{n \rightarrow \infty} P^n$$

where each row of A is the same probability row vector

$$\mathcal{A} = (a_0, a_1, \dots, a_N)$$

Therefore,

$$A = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} (a_0, a_1, \dots, a_N)$$

The matrix A can be evaluated by three distinct methods :

1. By multiplying P by itself n times, for sufficiently large n , we can obtain the matrix A to any desired accuracy. This method is not very practicable if it is applied directly, but we can raise P to the powers 2, 4, 8, and obtain

$$P^{(2^n)}$$

and for reasonably small values like $n = 10$, or even less we can approach A as closely as desired. Some theorems are available that indicate the rate of convergence of P^n , Kemeny and Snell (35).

2. We can determine P^n by diagonalising P first, or reducing it to the Jordan canonical form.
3. Telephone traffic theory approach

The row vector $\mathcal{A} = (a_0, a_1, \dots, a_N)$ is the stationary equilibrium probability distribution of the Markov Chain. In telephone traffic terms, the a_i are the probabilities of finding i trunks simultaneously engaged at any time t . From the derivation of Erlang's loss formula, the a_i are given by

$$a_i = \frac{\frac{\lambda^i}{i!}}{\sum_{j=0}^N \frac{\lambda^j}{j!}}$$

where as before

λ = calling rate in erlangs

N = number of trunks in the group.

Method (3) was used to compute A for the case of random offered traffic. Note that the following properties of A can be used to check the computed value of A ,

$$PA = AP = A$$

$$AA = A$$

8.4 The fundamental matrix

The fundamental matrix Z , also known as the characteristic matrix, plays a very important role in the analysis of Markov chains. It is defined by

$$Z = (I - (P - A))^{-1}$$

By realising that the row vector $\mathcal{L} = (a_0, \dots, a_N)$ is in fact the eigenvector of matrix P corresponding to the eigenvalue $\lambda_1 = 1$ (which P must always have) it can be shown that the inverse matrix defining Z always exists.

The matrix Z can also be expanded (using the binomial theorem) into the infinite series

$$Z = I + \sum_{n=1}^{\infty} (P - A)^n$$

Using the properties $AP = PA = A$, and $AA = A$, this reduces to

$$Z = I + \sum_{n=1}^{\infty} (P^n - A)$$

By invoking the matrix theorem which states :

If B^n tends to 0 (the zero matrix) as n tends to infinity,
then $(I - B)$ has an inverse

$$(I - B)^{-1} = I + B + B^2 + \dots$$

$$= I + \sum_{k=1}^{\infty} B^k$$

we see that the infinite series converges since $\lim (P^n - A) = 0$ as $n \rightarrow \infty$.

The following properties of Z can serve as useful computation checks:

1. $PZ = ZP$
2. $(Z) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$
3. $AZ = A$
4. $I - Z = A - PZ$

Note that while Z has several properties in common with a transition matrix, e.g. row sums equal 1, it does not necessarily have non-negative elements.

8.5 The first passage time

Define $f_{ij}^{(n)}$ as the probability of entering state j at the n -th step for the first time after leaving the initial state i

Thus for the Markov chain x_n , $n = 0, 1, \dots$ (with stationary transition probabilities) taking on values $(\dots i, j, \dots)$ in state space S , we can define $f_{ij}^{(n)}$ as

$$f_{ij}^{(n)} = \begin{cases} 0 & , n = 0 \\ \text{Prob}(x_{m+1} = j \mid x_m = i), & n = 1 \\ \text{Prob}(x_{m+n} = j, x_{m+k} \neq j, k = 1, \dots, n-1 \mid x_m = i) & \\ & \text{for } n = 2, 3, \dots \end{cases}$$

The transition from i to j can evidently take place in two mutually exclusive ways. Either the transition takes place in one step directly from i to j , or it can take place through an intermediate state k , and then from k to j in some number of steps n , $n = 1, 2, \dots$

8.5.1 The mean first passage time

The mean first passage time m_{ij} is defined by

$$m_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}$$

and is given recursively by

$$m_{ij} = 1 \cdot p_{ij} + \sum_{k \neq j} (1 + m_{kj}) p_{ik} \\ (1 \text{ step or } (1 + m_{kj}) \text{ steps})$$

We can write this equation for the matrix $M = (m_{ij})$ as

$$M = P + P(M - M_{dg} + E - E_{dg})$$

Where E is the square matrix with all elements equal to 1,

B_{dg} is the diagonal matrix $(b_{ij} \delta_{ij})$

δ_{ij} is the Kronecker delta.

It follows that

$$M = P(M - M_{dg}) + P(I + E - E_{dg}) \\ = P(M - M_{dg}) + PE \\ M = P(M - M_{dg}) + E$$

This implicit equation determines the mean first passage time matrix.

Multiplying through by A

$$AM = A(M - M_{dg}) + AE \\ AM_{dg} = AE$$

Therefore,

$$E = AM_{dg}$$

$$\text{i.e. } m_{ii} = \frac{1}{a_i} \quad i \in S$$

Let us define the diagonal matrix D by

$$D = (d_{ij}) = \frac{\delta_{ij}}{a_i}$$

It can then be shown, Kemeny and Snell (35), that the mean passage time matrix M is given by

$$M = (I - Z + EZ_{dg}) D$$

8.5.2. The variance of the first passage time

By using arguments similar to those used for the mean, it can be shown, Kemeny and Snell (35), that the variance of the first passage time matrix M_2 is given by

$$M_2 = M (2Z_{dg}D - I) + 2(ZM - E(ZM)_{dg}) - M_{sq}$$

$$\text{where } M_{sq} = (m_{ij}^2)$$

8.5.3. The probability distribution of the first passage times

Using the matrix methods outlined above, we can obtain all the mean first passage times for $i, j \in S$, and all the variances of the first passage times. It is most unlikely though that we are interested in all of these. In fact, in the routing problem we are principally only concerned with the passage times to the state of congestion N from the internal states 0, 1, 2, ... N-1. It is, therefore, wasteful of both computation time and memory space used for storing the square matrices, to consider the whole mean passage matrix.

The alternative is to consider the disjoint possibilities of reaching state j from state i in exactly n steps. By applying some renewal theory, Feller (2,3), it is seen that

$$\begin{aligned} f_{ij}^{(n)} &= p_{ij}^{(n)} - f_{ij}^{(n-1)} p_{jj} - f_{ij}^{(n-2)} p_{jj}^{(2)} - \dots - f_{ij}^{(1)} p_{jj}^{(n-1)} \\ &= p_{ij}^{(n)} - \sum_{k=1}^{n-1} f_{ij}^{(n-k)} p_{jj}^{(k)} \end{aligned}$$

where $p_{ij}^{(n)}$ is the n -step transition probability from i to j

Note that $P^n = (p_{ij}^{(n)})$.

We can therefore compute $f_{ij}^{(n)}$ $i, j \in S$ for $n = 1, 2, \dots$

using the recursive relation above. By using a column vector F_{ij} containing a sufficiently large number of elements, we can calculate the mean, variance, and all the other moments of the first passage time. Similarly, we can compute the cumulative distribution function $F_{ij}(n)$ by

$$F_{ij}(n) = \sum_{m=0}^n f_{ij}^{(m)}$$

8.6 The stochastic behaviour of a group of circuits offered non-random traffic

The difference between the stochastic behaviour of a group of circuits offered non-random traffic from the behaviour of the same group offered random traffic is easily studied by making a suitable modification to the transition matrix P . The rest of the analysis is then exactly the same as in the case of random traffic. We shall assume that the non-random traffic can be adequately described by its mean m , and its variance v .

8.6.1 The transition matrix for smooth traffic

The Binomial process has been used to describe smooth traffic, i.e. traffic whose variance is less than its mean, with remarkably good success, Mina (40). It is proposed to write the matrix P using the Binomial distribution as the call arrival process. The holding time is again assumed to have a negative exponential distribution with unit mean.

Let us assume that the number of subscribers having access to the group of circuits is finite, and equal to n . The probability that an idle subscriber makes a call is equal to p . Then the probability that x calls are originated, $p(x)$, is given by.

$$\begin{aligned} p(x) &= \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ &= 0 & x = n+1, \dots \end{aligned}$$

The mean m , and the variance v of x are found to be

$$\begin{aligned} m &= np \\ v &= np(1-p) \end{aligned}$$

We define a to be the variance to mean ratio, i.e. $a = v/m$.

We can express p and n in terms of a and m as

$$p = 1-a$$

$$n = \frac{m}{1-a}$$

If we let the average calling rate per subscriber be B , (the time when the subscriber is busy is also included in working out this average) we find, Ellidin and Lind (12), that B is given by

$$B = \frac{1-a}{a}$$

Now if the circuit group is in a state with k calls in progress, the probability that an additional call arrives in the time interval dt is

$$\text{Prob (1 call arrives in } dt) = (n-k) Bdt + O(dt)$$

Therefore,

$$p_{k,k+1} = (n-k) Bdt + O(dt)$$

$$p_{k,k-1} = kdt + O(dt)$$

$$p_{k,k} = 1 - p_{k,k-1} - p_{k,k+1}$$

Let p_i be the stationary probability for state i . If we are given a trunk group with N trunks, then the mean carried traffic m_c is given by

$$m_c = \sum_{i=0}^N i p_i$$

If the mean offered traffic is m , then we should find that

$$m = m_c + mp_N$$

Where mp_N is the mean overflowing traffic. If this equality is not satisfied, then B should be adjusted (slightly) until the relationship is satisfied. If n is very much greater than N , then no adjustment should be necessary, but for values of n of the same order as N , an adjustment in B is necessary.

8.6.2 The transition matrix for rough traffic

We have already remarked that the arrival distribution for rough traffic, whose variance is greater than its mean, can be very closely approximated by the negative Binomial distribution, Wilkinson (9). We are, therefore going to write down the transition matrix P in terms of an arrival process governed by the negative Binomial distributions.

The probability of having x simultaneous calls from $x + r$ subscribers is given by

$$p_r(x) = \binom{-r}{x} (-q)^x (1-q)^r \quad x = 0, 1, \dots$$

Now the mean m and the variance v of x are

$$m = \frac{kq}{1-q}$$

$$v = \frac{kq}{(1-q)(1-q)}$$

and

$$a = \frac{v}{m} = \frac{1}{1-q}$$

$$r = \frac{m}{a-1}$$

$$q = 1 - \frac{1}{a}$$

In the positive Binomial description of smooth traffic, the calling rate was state dependent, decreasing linearly with the number of busy subscribers. This gave a variance that was less than the mean. In the negative Binomial description of rough traffic, the calling rate is also state-dependent, increasing linearly with the number of busy subscribers.

Let the average calling rate per subscriber be B (the number of subscribers depends on the number of circuits in the group and equals $r + N$). Then

$$B = q = 1 - \frac{1}{a}$$

If the trunk group is in a state with k calls in progress, the probability that another call arrives in a time interval dt is

$$\text{Prob (1 call arrives in } dt) = (k+r) B dt + O(dt)$$

The transition matrix P is, therefore given by

$$P_{k,k+1} = (k+r) B dt + O(dt)$$

$$P_{k,k-1} = k dt + O(dt)$$

$$P_{k,k} = 1 - P_{k,k-1} - P_{k,k+1}$$

An adjustment may be made to B as was done in the smooth traffic case to make sure that the offered traffic is indeed equal to the prescribed value m .

8.7 The first passage time distribution in continuous time

We consider a trunk group of n circuits which is offered λ erlangs. In particular, we are interested in the Markov process $X(t)$, $0 \leq t$ in continuous time with discrete states j , $j = 0, 1, \dots, n$.

Let
$$p_{ij}(t, u) = \text{prob} (X(t) = j \mid X(u) = i)$$

Since the process is time-homogenous, we can write

$$p_{ij}(t-u) = \text{prob} (X(t) = j \mid X(u) = i)$$

In the case when $t - u$ is small, say dt , we can write the probabilities as

$$p_{ij}(dt) = q_{ij} dt + O(dt) \quad i \neq j$$

$$p_{ii}(dt) = 1 + q_{ii} dt + O(dt)$$

Let us define

$$P(t, u) = (p_{ij}(t, u))$$

and
$$Q(t) = (q_{ij}(t))$$

If the row sums of matrix Q are all equal to zero, the process is said to be conservative and satisfies Kolmogorov's forward and backward equations,

$$\frac{\partial P(t, u)}{\partial u} = P(t, u)Q(t)$$

$$\frac{\partial P(t, u)}{\partial t} = Q(t)P(t, u)$$

These Kolmogorov equations enable us to calculate the matrix P when we are given the matrix Q .

For the trunk group,

$$q_{i, i+1} = \lambda \quad i = 0, \dots, n-1$$

$$q_{i, i-1} = i \quad i = 1, \dots, n$$

$$q_{i, i} = -(\lambda + i) \quad i = 0, \dots, n$$

and all other elements of Q equal to zero. Thus,

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 1 & -(\lambda+1) & \lambda & & & & & & \\ \vdots & & & & & & & & \\ \vdots & & & & & & & & \\ \vdots & & & & & & & & \\ 0 & 0 & & & n-1 & -(\lambda+n-1) & \lambda & & \\ & & & & & -n & -(\lambda+n) & & \end{bmatrix}$$

For our case of constant Q , the formal solution to both the forward and backward equations is

$$P(t) = \exp(Qt)$$

The first passage time distributions of particular interest in a routing context are $f_{in}(t)$, $i = 0, \dots, n-1$,

where $f_{ij}(t) = \text{prob}(X(t) = j, X(u) \neq j, 0 \leq u < t \mid X(0) = i)$.

Thus $f_{ij}(t)$ is the distribution of the smallest value of t at which state j is reached given that $X(0) = i$.

The simplest way to study the first passage time distributions is to modify the process and turn state n into an absorbing state. The Q matrix is then

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & & 0 & 0 \\ 1 & -(\lambda+1) & \lambda & & & \\ & & & n-1 & -(\lambda+n-1) & \lambda \\ 0 & & & 0 & 0 & 0 \end{bmatrix}$$

By re-naming the states, and letting

$$\begin{array}{llll} \text{new state } 0 & = & \text{old state } n \\ " & 1 & = & " \quad n-1 \\ " & i & = & " \quad n-i \\ " & n & = & " \quad 0 \end{array}$$

we get the new matrix Q_{new}

$$Q_{\text{new}} = \begin{bmatrix} 0 & 0 & 0 & & 0 \\ \lambda & -(\lambda+n-1) & n-1 & & \\ & & & & \\ 0 & & & -\lambda & \lambda \end{bmatrix}$$

Ledermann and Reuter (137) have shown that if matrix Q_{new} has eigenvalues $B_0, B_1, B_2, \dots, B_n$, we have $0=B_0 > B_1 > B_2 > \dots > B_n$.

That is, all the eigenvalues are distinct, and negative with the exception of B_0 which equals zero.

Since the eigenvalues of Q are obviously the same as those of Q_{new} , we can always diagonalise Q , and obtain a compact solution for P .

Then starting from an initial state i defined by the column vector p_0 , with zeros everywhere except the i -th row which has a 1, the n -th element in the column vector Pp_0 gives the probability of being in state n at time t in the modified process. This is equal to the probability, in the first process, that the first transition to state n has occurred at or before the time t . So that the n -th element of $P(t)p_0$ is the cumulative distribution of the first passage to state n . Now $f_{\text{in}}(t)$ is the derivative of this cumulative distribution. From the solution it turns out that $f_{\text{in}}(t)$ has the form,

$$f_{\text{in}}(t) = \sum_{i=1}^n b_i \exp(-a_i t)$$

In the study of routing, an event of paramount interest is that of the first passage time to the congestion state being less than the holding time. Suppose we wish to place an alternatively routed call using trunk group A - B connecting exchanges A and B. If we place the call on this group, and if during its holding time the congestion state on A-B is attained, any calls that may be originated between A and B will be lost. The first passage time to the congestion state reflects the probability of such calls being lost.

Let X be a random variable representing the time to congestion, and Y a random variable representing the holding time of a telephone call. We are interested in the random variable Z defined as $Z = X - Y$.

8.8 The probability density function of $Z = X - Y$

Let the p.d.f.s. of the random variables Z, X, Y , be $f_z(z)$, $f_x(x)$, and $f_y(y)$ respectively.

$$f_y(y) = \begin{cases} p \exp(py) & y \geq 0 \\ 0 & y < 0 \end{cases}$$

and

$$f_x(x) = \begin{cases} \sum_{i=1}^n b_i \exp(-a_i x) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$= 0$$

$$x < 0$$

Then

$$f_z(z) = \int_{-\infty}^{+\infty} f_z(z-u) f_y(-u) du$$

$$= p \int_{-\infty}^{u=z} \sum_i b_i \exp(-a_i(z-u)) \cdot \exp(-pu) du$$

The case $z \leq 0$.

$$f_z(z) = \sum_i p b_i \exp(-a_i z) \int_{-\infty}^{u=z} \exp(u(p + a_i)) du$$

$$= \sum_i \frac{p b_i}{p + a_i} \cdot \exp(pz)$$

The Case $z > 0$

$$f_z(z) = \sum_i p b_i \exp(-a_i z) \int_{-\infty}^{u=0} \exp(u(p + a_i)) du$$

$$= \sum_i \frac{p b_i}{p + a_i} \exp(-a_i z)$$

It is easily verified that the integrals of $f_z(z)$ and $f_x(x)$ from $-\infty$ to $+\infty$ are both equal to

$$\sum_i (b_i/a_i)$$

8.8.1 The probability of a routing error

The probability that the time to congestion is less than the holding time is found by integrating $f_z(z)$ from $-\infty$ to 0 giving the result

$$\text{Prob (routing error)} = \sum_i \frac{b_i}{a_i + p}$$

This probability of a routing error will form the basis of a class of routing strategies to be used in Chapters 10 and 11.

8.9 Conclusion

In this chapter we have introduced the concept of "time to congestion". We have shown how the probability distribution of the time to congestion can be computed and how it can be used in a routing

context. It is my belief that this probability distribution of the time to congestion will in time assume an importance in routing studies similar to that assumed by Erlang's formula in trunk group dimensioning.

One computational difficulty must be mentioned, however, before we pass on to the next chapter. In the computation of the distribution of the time to congestion, we have used the eigenvalues of a matrix. It is clear that for large trunk groups, of say over 100 circuits, the computation of the eigenvalues can become a very serious problem, especially if they are close to each other. I hope that a better method of determining the distribution of the time to congestion will soon be found.

I must emphasize that the first passage time to congestion plays a vital role in indicating the probability that a direct call may be lost as a direct result of placing an alternately routed call. This important point has been discussed on page 118.

CHAPTER 9. ROUTING STRATEGIES USING THE MEAN TIME TO THE CONGESTION STATE

9.1 Introduction

9.2 The four-exchange network

9.3 The non-symmetrical four-exchange network

9.3.1 The protection of the point-to-point grades of service.

9.3.2 The effect of the variance to mean ratio on the mean time to congestion

9.4 A non-symmetrical network with large trunk groups.

9.5 Conclusion

CHAPTER 9. ROUTING STRATEGIES USING THE MEAN TIME TO THE CONGESTION STATE

9.1 Introduction

In this Chapter we use the mean time to the congestion state (MTC) in devising routing strategies. Consider the three-exchange network of Fig. 9.1.

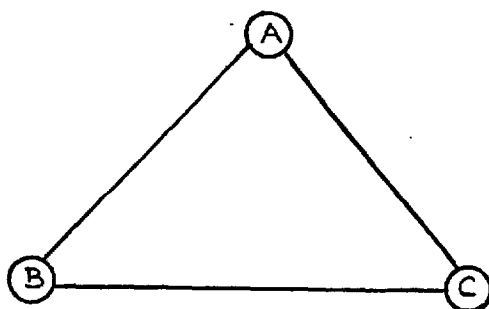


Fig. 9.1 The three-exchange network

A call from exchange B to C can either use the direct route BC or be routed indirectly via exchange A.

We have already seen that this overflow capability can be used to improve the overall grade of service. Unfortunately, the use of this capability at high overloads leads to a severe degradation of service. Qualitatively, this is due to the "uneconomical" occupation of circuits by some calls; some calls are using two circuits at a time when circuits are in severe demand. This leads to a severe loss in efficiency.

For a quantitative explanation, we turn to the mean time to the congestion state (MTC). Using the methods of Chapter 8, we have computed the mean time to congestion for a ten-circuit group for various levels of offered traffic in the range 5 to 11 erlangs and plotted the family of curves in Fig. 9.2. In Section 9.2, we make use of these curves.

9.2 The four-exchange network

Let us re-consider the four-exchange network of Fig. 9.3a, and the A to B call in particular.

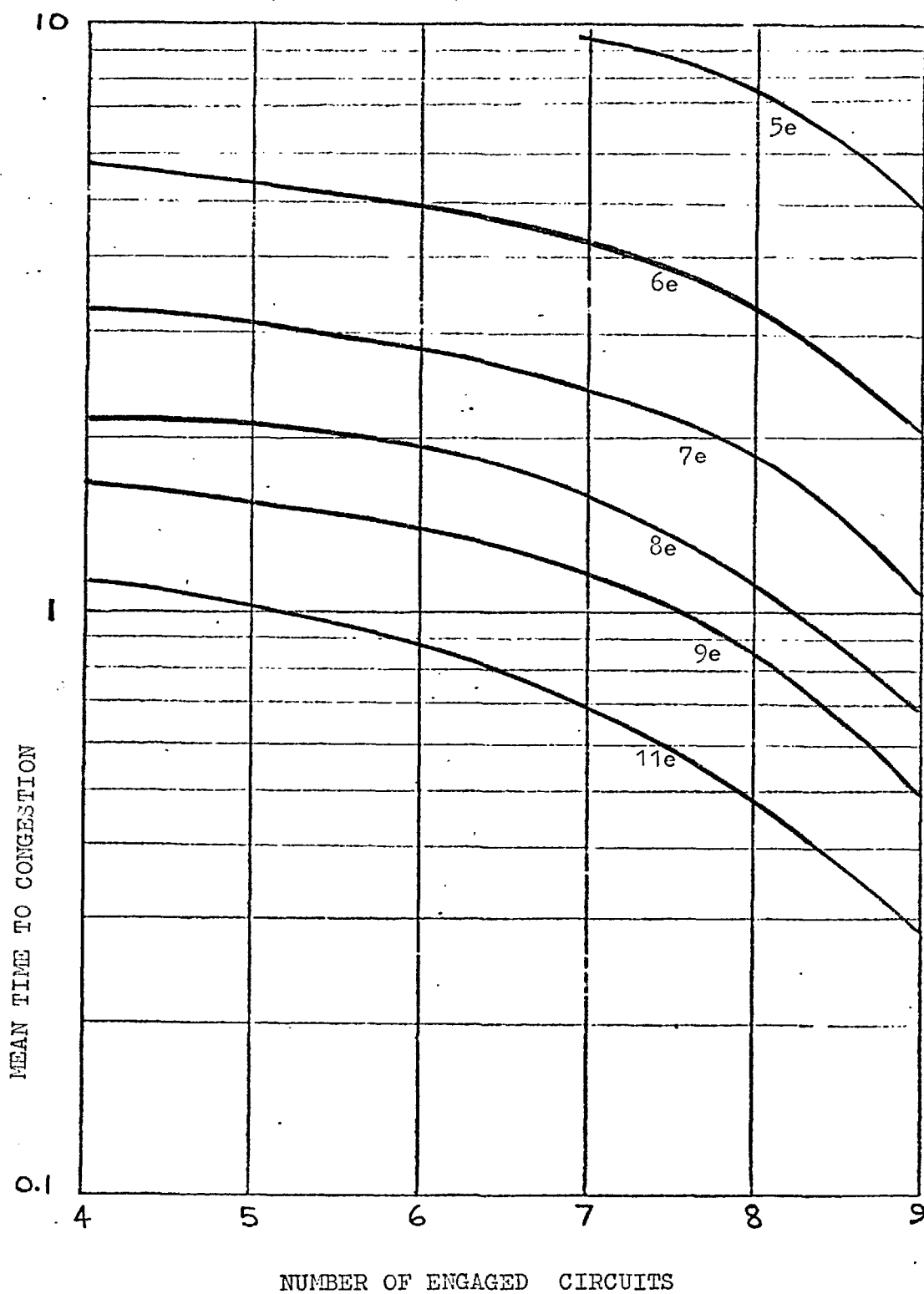


Fig. 9.2 The mean time to congestion for a group of ten circuits for various levels of offered traffic

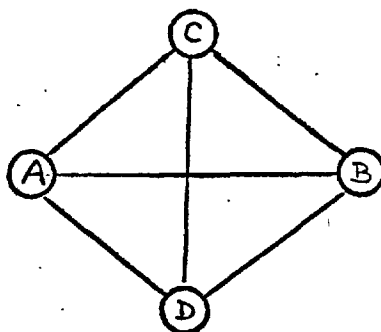


Fig. 9.3a The four-exchange network.

The first choice route for an AB call is the direct link AB. If the AB group of circuits is congested, we may choose between the two routes of Fig. 9.3b as follows.

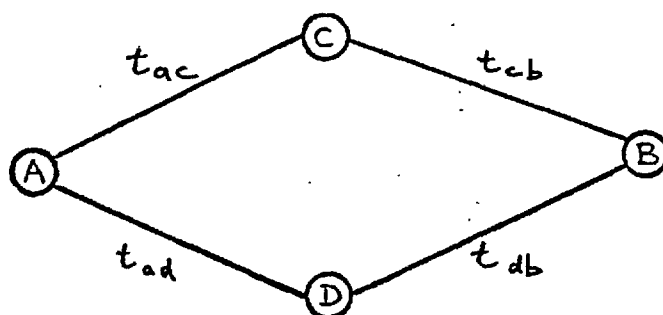


Fig. 9.3b The two alternative routes for an AB call with the mean time to congestion for each link.

Denote routes ACB and ADB by Route 1 and Route 2 respectively. Then let

$$y_1 = \min (t_{ac} - 1, t_{cb} - 1)$$

$$\text{and } y_2 = \min (t_{ad} - 1, t_{db} - 1) .$$

Route ACB is disallowed if y_1 is less than or equal to zero. Similarly route ADB is not allowed if y_2 is negative or equal to zero. This disqualification means that each link of an alternative route must have a mean time to congestion greater than unity (the mean holding time). With this constraint we hope to reduce considerably the uneconomical occupation of circuits. From Fig. 9.2, for example, we see that if the offered traffic is 9 erlangs, three circuits are reserved for direct traffic. If the offered traffic is 8 erlangs, on the other hand, then only two circuits are reserved for the direct traffic. The mean time to congestion has thus enabled us to employ the selective reservation of trunks.

If both routes are still in the running, we use the min-max method (Chapter.7) on the reciprocals of the mean times to congestion to choose the better of the two routes. (This is equivalent to a max-min choice on the MTC. The change from min-max to max-min is due to the fact that the MTC decreases monotonically with an increase in the number of busy circuits).

This MTC routing strategy was applied to the four-exchange network of Fig. 9.3a in which each trunk group has ten trunks. The offered traffic between each pair of exchanges was varied in the range 4 to 10 erlangs.

The performance of the network using this MTC routing strategy is almost exactly the same as that obtained in Chapter 7, using the min-max strategy, at the low levels of offered traffic. But at the higher overloads, the MTC strategy is definitely superior, see Fig.9.4. In fact, the overall grade of service is always better than that obtained with the direct routing strategy.

We have tested the MTC routing strategy for the four-exchange network whose trunk groups have all got ten circuits. This network has special properties; our next step must therefore be to investigate this strategy for networks which have trunk groups with an unequal number of circuits, and which have unequal exchange-to-exchange offered traffic levels. But before doing this, let us make three observations on our MTC routing strategy:

1. In computing the mean time to congestion we assume that we know the exact (nominal) amount of traffic offered to each link. In fact, the offered traffic is greater than its nominal value because we are re-offering the overflow traffic to the alternative routes.

2. The holding time is a random variable with unity mean. It also has a variance σ^2 which is non-zero, and which should be taken into account. It is more appropriate, therefore, to require that the mean time to congestion should be greater than $1 + f(\sigma^2)$, where $f(\sigma^2)$, is some function of the variance, before alternative routing is allowed.

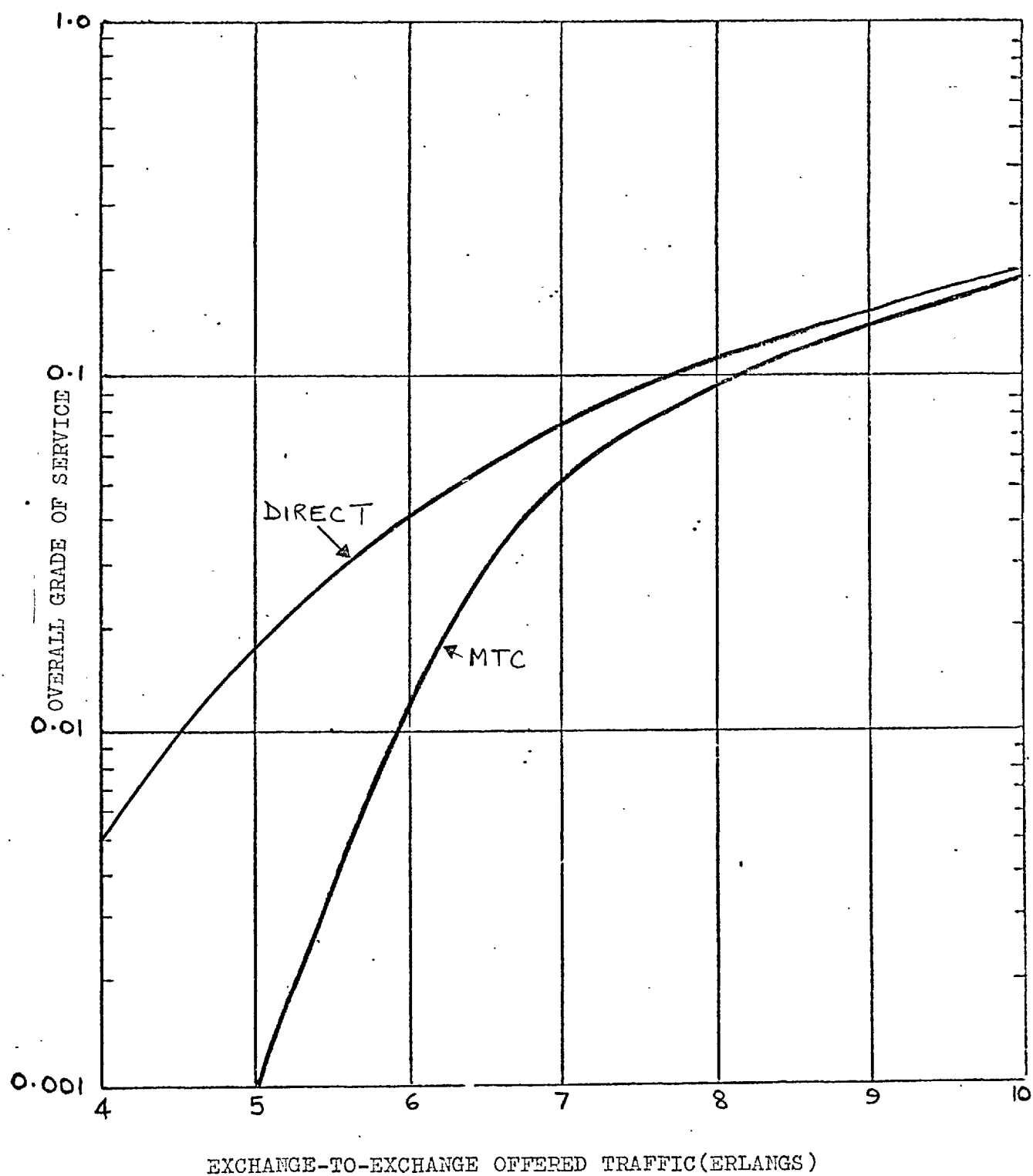


Fig. 9.4 A comparison of the MTC and the direct routing strategies

3. We assumed in computing the mean time to congestion that the call arrival process was Poisson, with the mean of the offered traffic being equal to its variance. This is no longer the case because of the overflow traffic component, which, in general, has a variance that is greater than its mean.

If the overflow traffic is small compared to the nominal offered traffic, as is the case for the very low levels of offered traffic or very high overloads, observations (1) and (3) diminish considerably in importance. Similarly, if many parcels of overflow traffic are offered to the same link, the total offered overflow traffic tends to have a variance that approaches its mean in value.

9.3 The non-symmetrical four-exchange network

The four-exchange networks considered so far were ^{made} up of trunk groups that had an equal number of circuits. We now take a four-exchange network whose trunk groups have an unequal number of circuits, and whose inter-exchange offered traffics are in general different. The network is shown in Fig. 9.5.

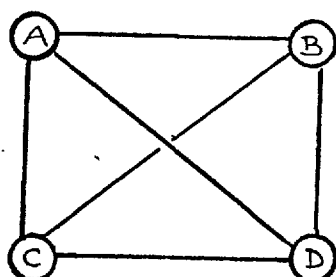


Fig. 9.5 The four-exchange network

The circuits per trunk group, and the amounts of directly offered ^{traffic} for each trunk group are shown in the table below.

Trunk group	Number of circuits	Directly offered traffic (erlangs)
A - B	4	1
A - C	8	4
A - D	5	2
B - C	6	3
B - D	9	5
C - D	10	6

It is now required to evaluate the performance of the network for loads ranging from the nominal (rated) load as defined in the table above to 100 per cent overload. This performance was evaluated by simulation using the MTC routing strategy. The overall grade of service obtained using this strategy is plotted on Fig. 9.6, and for comparison the overall grade of service using the direct routing strategy is plotted on the same graph.

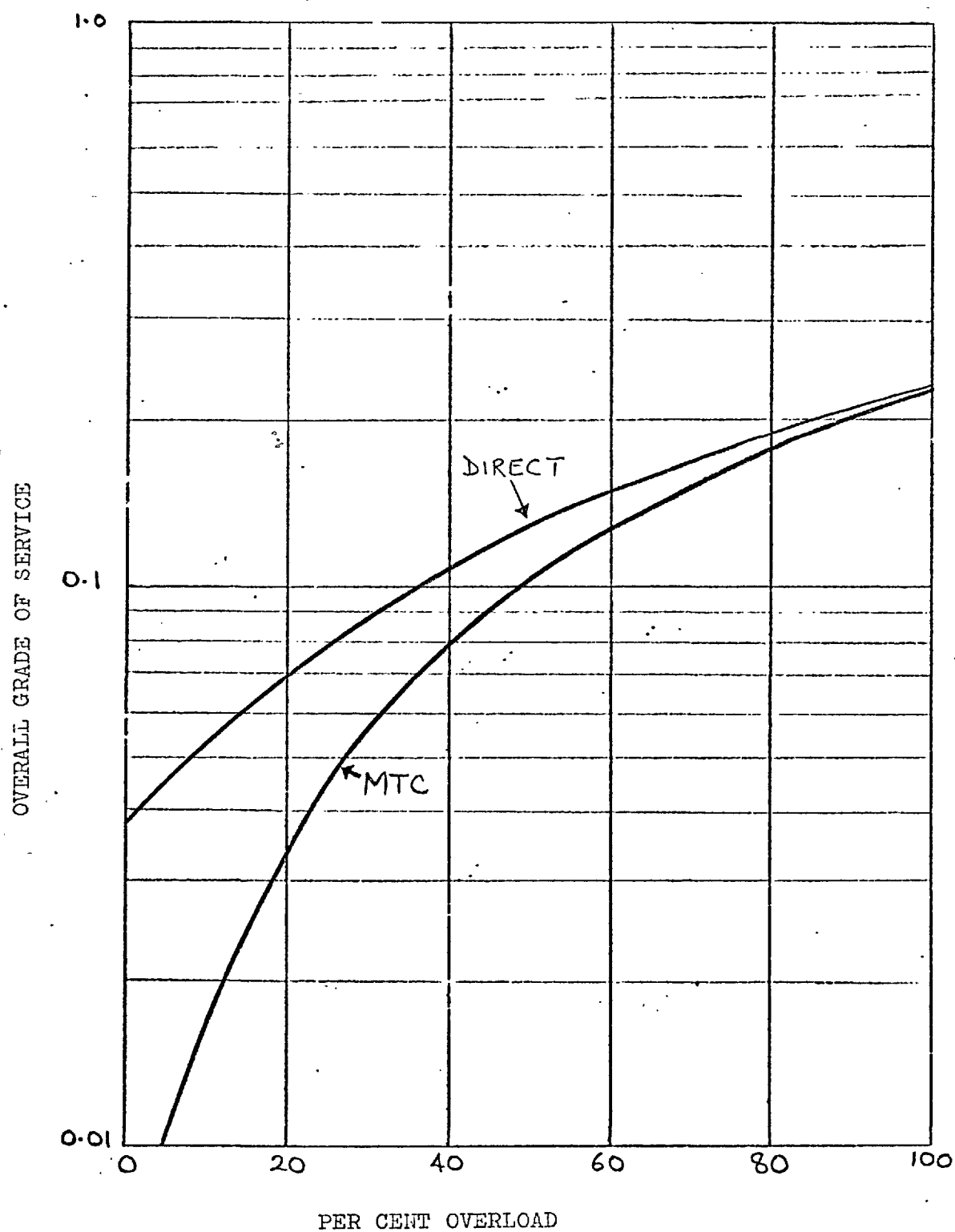


Fig. 9.6 The overall grade of service for the non-symmetrical four-exchange network

It is evident from Fig. 9.6 and Fig. 9.7 that the overall grade of service using the MTC routing strategy is always better than that given by the direct strategy. This reinforces the same conclusions we came to in the case where all the trunk groups had the same number of circuits.

9.3.1 The protection of the point-to-point grades of service

One of the fears voiced about optimising the overall grade of service of a network has been the danger that some parcels of traffic might become isolated in this effort to maximise the traffic carrying capacity of the network. This fear of providing an inequitable service is entirely justified.

In the example where all the trunk groups had an equal number of circuits, each trunk group had an individual grade of service that was better than that obtained by evaluating Erlang's formula for each trunk group. Thus not only did the system as a whole perform better than the same network with no alternate routing, but the individual point-to-point grades of service were always better than would have been obtained with no alternate routing.

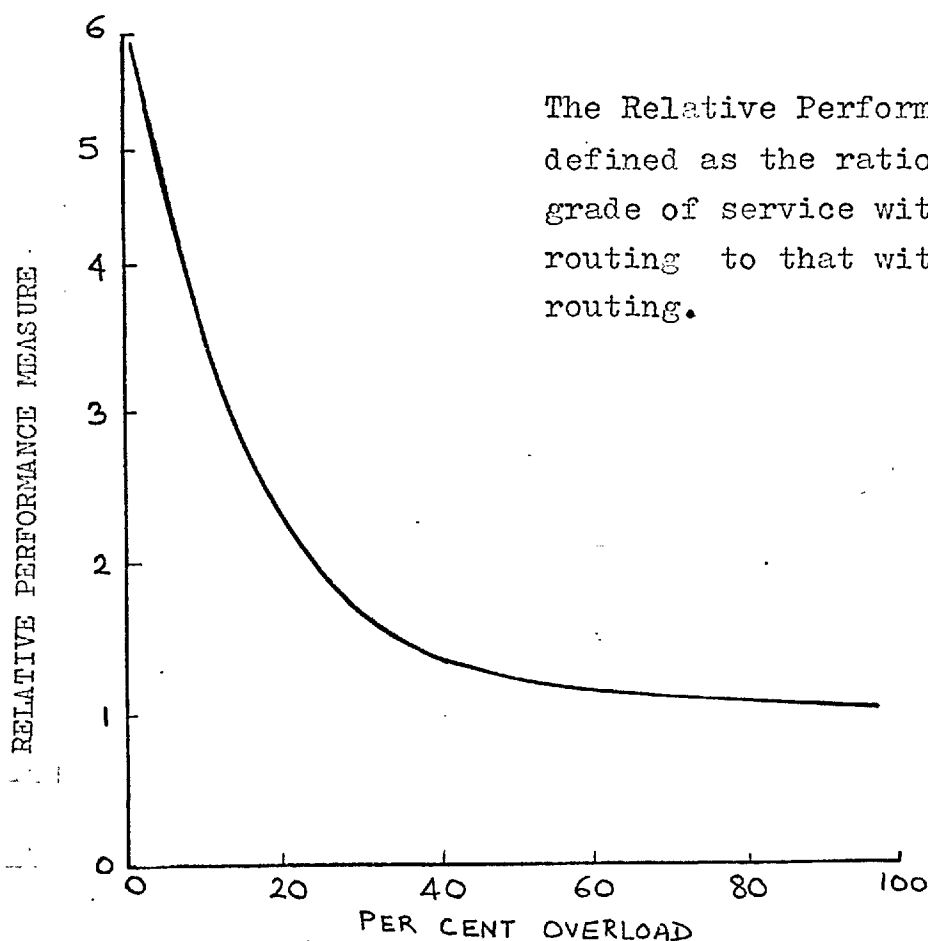


Fig. 9.7 Relative Performance Measure against overload

But in the case of unequal numbers of circuits per trunk group, the smaller groups suffer to some extent at relatively high overloads because of the alternate routing. It would seem therefore that while the decision rule to use alternate routing if the mean time to congestion is greater than unity produces some very acceptable results, it does not safe-guard the small groups sufficiently.

To get further insight into the behaviour of the mean time to congestion (MTC) strategy, we have investigated the MTC from the penultimate state to the ultimate (congestion) state; in an n -trunk group, this is from the state in which $n-1$ circuits are engaged to the state in which all n circuits are occupied. This MTC is plotted against overload for the six trunk groups in Fig. 9.8. From this graph it can be seen that for a five-trunk group, the MTC is only less than one for overloads

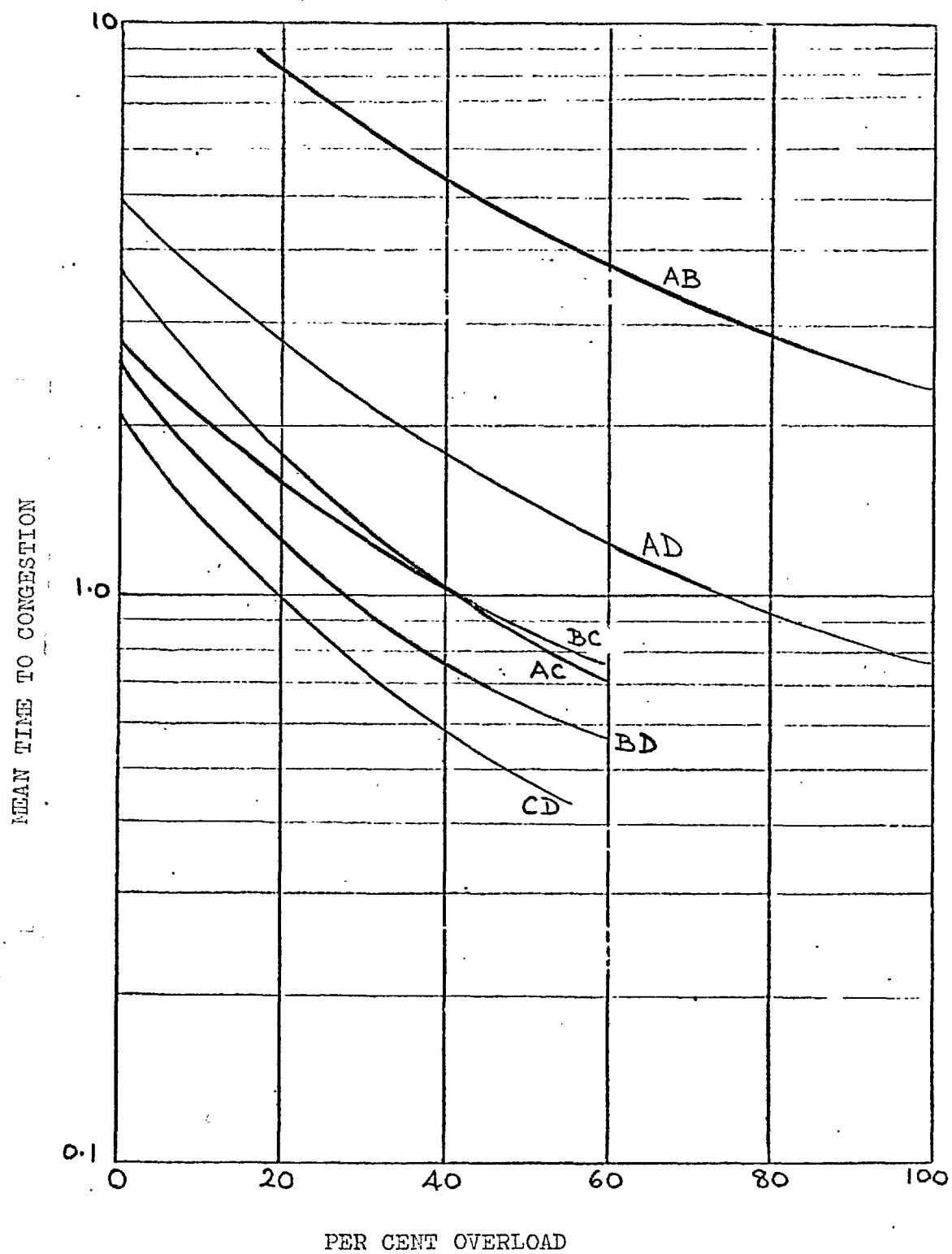


Fig. 9.8 The effect of overloads on the mean time to congestion for the trunk groups of Fig. 9.5 .

exceeding 80 per cent. In the case of a four-trunk group, the overload must exceed 200 per cent for the MTC to fall below one. This reveals that the small trunk groups are rather insensitive in their MTC to overloads. (The usual graceful deterioration of the grade of service of small trunk groups with overload is also reflected in the insensitivity of the MTC to overloads).

9.3.2 The effect of the variance to mean ratio on the mean time to congestion

In computing the mean time to congestion to be used in the MTC routing ^{strategy} in the four-exchange network, we assumed that the variance to mean ratio of the offered traffic stays constant at unity. We know, in fact, that this ratio is not equal to unity in practice, but rather tends to be greater than unity because of the rough component of traffic overflowing some other trunk group.

The mean time to congestion has been computed as a function of both the mean, and the variance to mean ratio of the offered traffic. The graphs of Fig. 9.9 and Fig. 9.10 show the case for a six trunk group, the mean time to congestion being computed from the state in which five of the six trunks are engaged. In Fig. 9.9, if the mean of the offered traffic is 4.3 erlangs, the mean time to congestion is greater than unity if λ is less than 1.1. But for $\lambda = 1.3$ the mean time to congestion is less than unity. This example is designed to show that decision errors will occur either because the mean offered traffic or the variance to mean ratio of the offered traffic is not known sufficiently accurately.

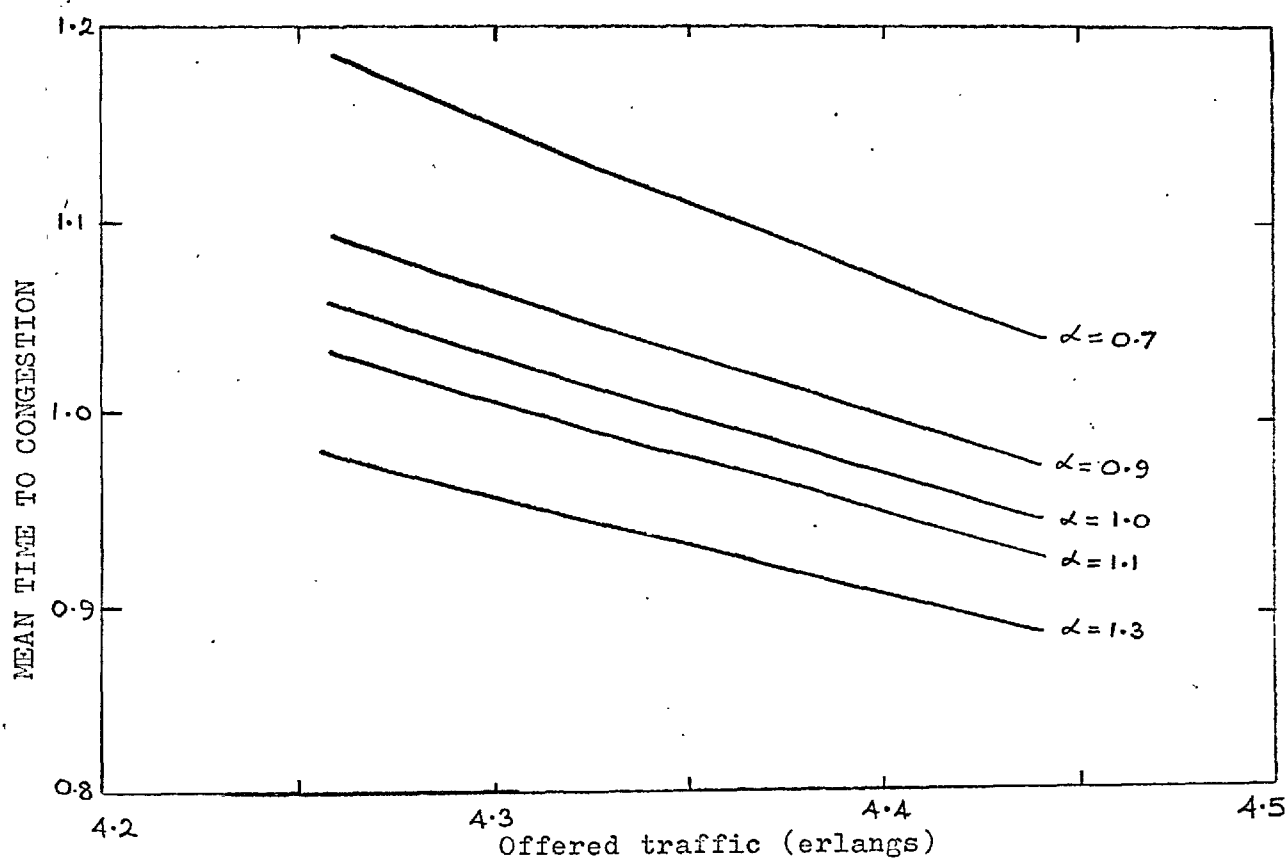


Fig. 9.9 Mean time to congestion against offered traffic with constant variance to mean ratio α .

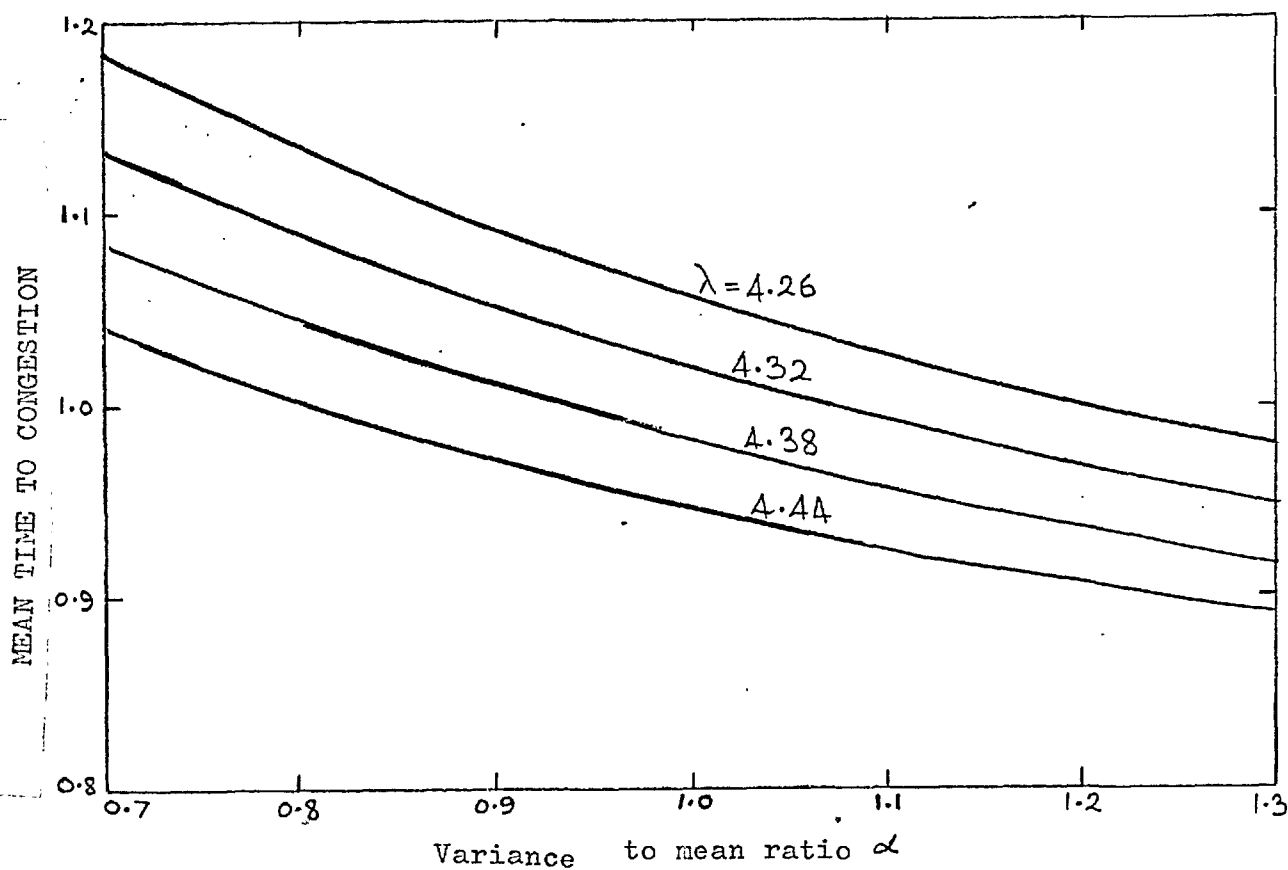


Fig. 9.10 Mean time to congestion against variance to mean ratio with constant offered traffic λ .

9.4 A non-symmetrical network with large trunk groups

The example in Section 9.3 treated a non-symmetrical network with rather small trunk groups. In this section we consider a four-exchange network with large trunk groups and correspondingly heavy point-to-point offered traffic, Figs. 9.11a and 9.11b.

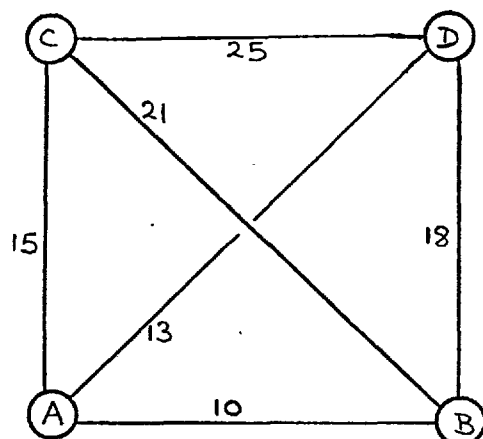


Fig.9.11a Network structure showing the number of exchange-to exchange trunks.

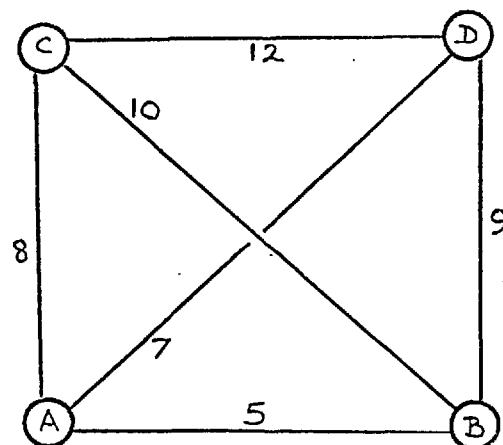


Fig.9.11b Exchange to-exchange offered traffic in erlangs.

Fig.9.11a shows the size of the trunk groups, while Fig.9.11b shows the nominal (zero per cent overload) offered traffic.

The performance of the network was evaluated for the MTC routing strategy. For comparison the network performance using the direct routing strategy was computed using Erlang's loss formula. In Fig.9.12 we have tabulated the overall grade of service. Simulation results of the MTC strategy, the computed results of the direct routing strategy, and the per cent increase in the traffic carrying capacity of the network.

OVERLOAD	TOTAL OFFERED TRAFFIC (ERLANGS)	OVERALL GRADE OF SERVICE		PERCENT MORE CARRIED TRAFFIC
		MTC STRATEGY	DIRECT STRATEGY	
0	51.0	-	0.006	0.6
10	56.1	0.0004	0.011	1.1
20	61.2	0.002	0.018	1.6
30	66.3	0.006	0.029	2.3
40	71.4	0.019	0.042	2.3
50	76.5	0.042	0.058	1.6
60	81.6	0.064	0.077	1.3
70	86.7	0.092	0.099	0.7
80	91.8	0.117	0.122	0.5
90	96.9	0.145	0.145	0.0
100	102.0	0.171	0.170	- 0.1

Fig. 9.12 Tabulation of the performance results of the non-symmetrical four-exchange network with large trunk groups.

Again as in the case of the non-symmetrical network with small trunk groups, the MTC strategy is almost always better than the direct strategy. The exchange-to-exchange grades of service for this network seem to have been adequately protected.

9.5 Conclusion

The MTC strategy is promising. It has given overall grades of service that are almost always better than those given by the direct strategy. Its main shortcoming has been the poor protection given to some exchange-to-exchange grades of service in the non-symmetrical network with small trunk groups.

In Chapter 10 we investigate some routing strategies that make use of the probability distribution of the time to congestion.

CHAPTER 10. ROUTING RULES DEPENDING ON THE DISTRIBUTION OF THE TIME TO CONGESTION.

10.1 Introduction

10.2 The probability that the time to congestion exceeds unity.

10.3 A comparison of routing strategies used in the four-exchange network.

10.3.1 Routing strategies to be compared

10.3.2 Comparison of overall (network) grade of service

10.3.3 Comparison of the point-to-point grades of service.

10.4 Conclusions.

10.5 Appendix : Tables of the point-to-point and overall grades of service for the network of Figs. 10.3a and 10.3b.

CHAPTER 10. ROUTING RULES DEPENDING ON THE DISTRIBUTION OF THE TIME TO CONGESTION

10.1 Introduction

Up to now we have only used the mean time to congestion in devising routing rules. It is natural to expect that we can devise better routing schemes if we are given more information about the random variable X representing the time to congestion.

If we have the first four central moments of X , we can approximate the probability density function of X using the Gram-Charlier or Edgeworth series, Barton and Dennis (136), and Fry (153). As it happens, we know from Chapter 8 the probability distribution of X . In this chapter we examine routing strategies that make use of this distribution of X , and apply them to the network of Fig.9.5 in Chapter 9.

10.2 The probability that the time to congestion exceeds unity

The event that the time to congestion exceeds the mean holding time (which is unity) is of great interest. We can easily determine the probability of this event using the distribution of X . Fig.10.1 and Fig.10.2 show how this probability reacts to overloads, the time to congestion being computed from the state where two and three circuits respectively are free.

We have devised a routing plan, named RP1, that uses the probability of the event that the time to congestion exceeds the mean holding time. If the probability of this event is greater than 0.5 for a given trunk group in a given state, then that trunk group may be used to carry alternately routed traffic. This condition must be satisfied by each one of the trunk groups forming the alternate route. If two or more routes satisfy these requirements, then we make a min-max decision using the reciprocals of the probabilities. The evaluation of RP1 was done by simulation. The results of this and other routing strategies will now be compared.

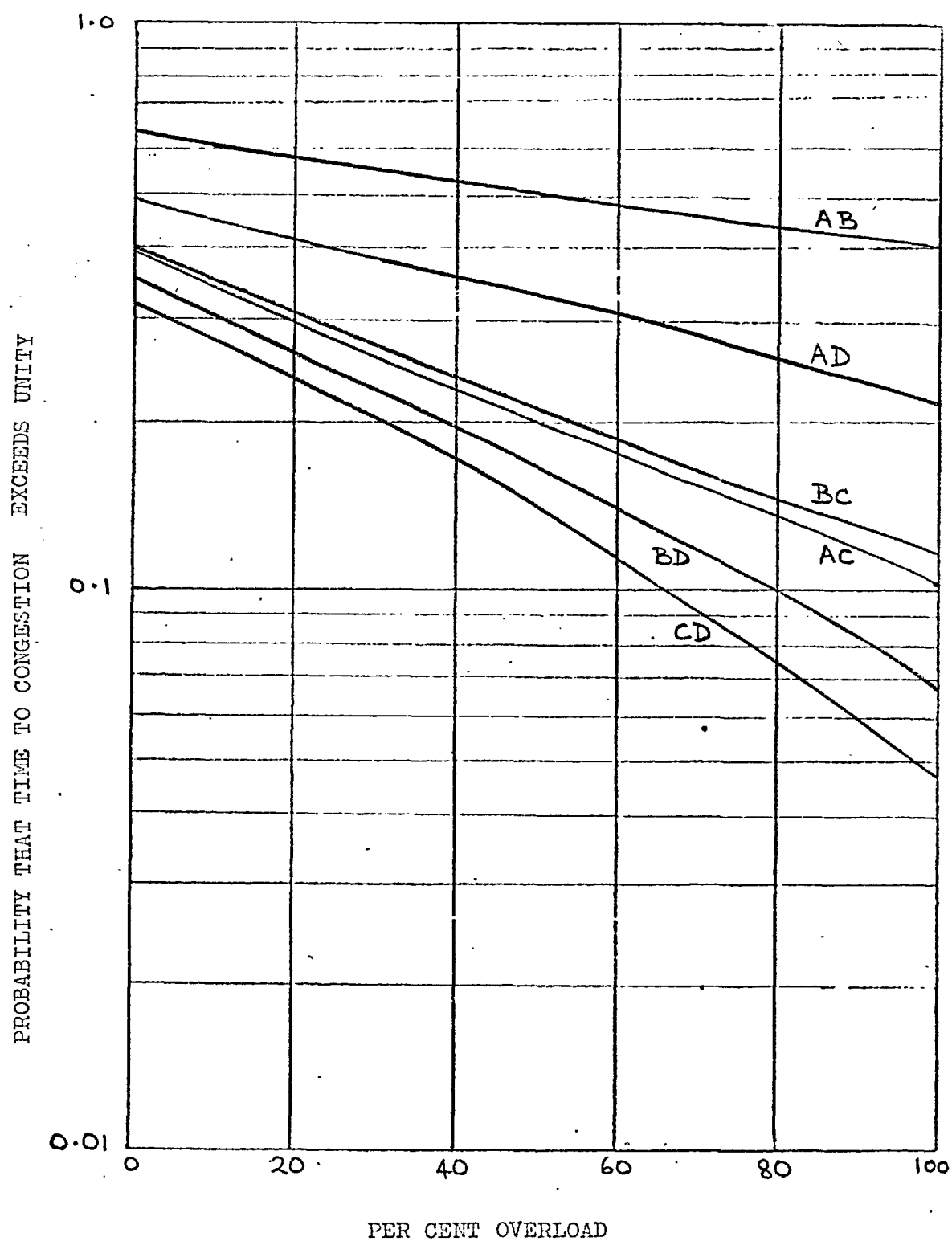


Fig. 10.1 The probability that the time to congestion, from the state where there are two free circuits, is greater than unity.

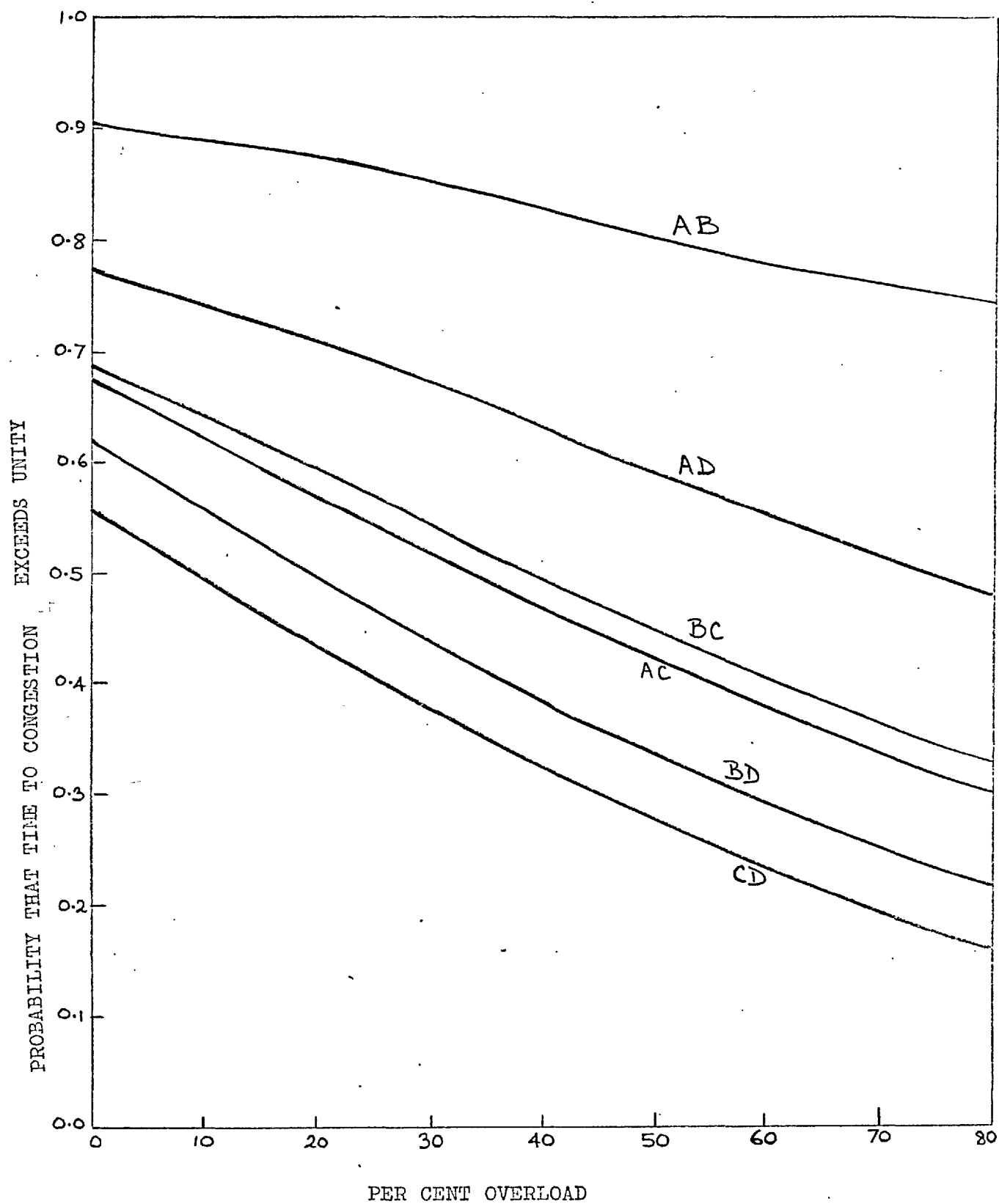


Fig. 10.2 The probability that the time to congestion, from the state where there are three free circuits, is greater than unity

10.3 A comparison of routing strategies used in the four-exchange network

At this stage it is important that we pause and compare the various routing strategies that we have experimented with.

I must confess that I am not aware of any unique yardstick by which I can determine the goodness of a routing plan; there is no generally agreed performance measure to optimize. In this study we have restricted ourselves to three main considerations: (1) we want a routing plan that will give the "best" network grade of service for any traffic pattern, (2) in attempting to approach the optimum network grade of service, we want to make sure that no parcel of traffic suffers a disproportionately lower grade of service than other parcels, (3) we are looking for a routing plan that is efficient, i.e. that leads to maximum carried traffic at all levels of offered traffic.

It is in the light of these considerations that we wish to compare the routing rules. Let us now list the six strategies to be compared.

10.3.1 Routing strategies to be compared

1. MTC (Mean Time to Congestion) Strategy

In this strategy, a link in a multi-link route is considered available for alternate routing at any given time if the mean time to congestion at that time is greater than unity. (Remember that the mean holding time is the unit of time). For example, in a two-link route, that route can only be used if for both links the mean time to congestion is greater than unity. We use the min-max method on the reciprocals of the mean times to congestion for choosing between two available routes as explained in Chapter 9.

2. RP1 (for Routing Plan 1) Strategy

In RP1 we use the cumulative distribution of the time to congestion. In particular if for a given link the probability that the time to congestion exceeds unity is greater than 0.5, then that link is considered available for use in an alternate route. Again the choice between two available routes was made Using the min-max method.

3. PERR1 Strategy

In PERR1, the probability of a routing decision error was used, see page 119. A link is considered available if the probability of error is less than 0.5. Again for two available routes, the min-max method was used.

4. PERR2 Strategy

PERR2 is similar to PERR1. The only difference is that in addition to requiring that every link of a multi-link route has a probability of error less than 0.5, we also require that the sum of the probabilities of error for all the links in the route be less than 0.5. We again use the min-max method to choose between two available routes.

5. PERR3 Strategy

PERR3 is very similar to PERR2. The only difference is in the way we choose between two available routes. In PERR3 we choose the route whose sum of probabilities of error is minimum.

6. PERR4 Strategy

This strategy is designed to determine the importance of the method we use for choosing between two available routes. PERR4 is similar to PERR3, the significant difference being that out of two or more available routes, we choose that route which is least preferred in routing strategy PERR3.

10.3.2 Comparison of Overall (network) grade of service

The nominal offered traffic and the number of circuits between the exchanges are shown in Fig.10.3a and Fig.10.3b.

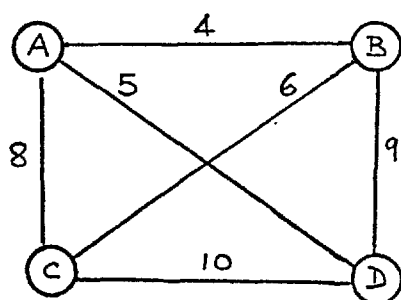


Fig.10.3a The number of circuits in each trunk group.

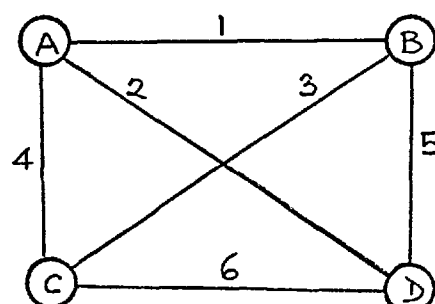


Fig.10.3b The inter-exchange offered traffic (erlangs)

The graph of Fig.10.4 shows the overall grade of service of the network for uniform overloads ranging from 0 (rated load) to 100 per cent overload.

We notice that all state-dependent alternative routing strategies show a better routing performance than the direct strategy, and this performance is maintained for the whole range of overloads. The reader will recall from Chapter 7 that this ability to give a performance that is always superior to that given by the direct routing strategy is not in general possessed by the alternative routing strategies that are not state-dependent.

For clarity we have redrawn in Fig.10.5 the overall grade of service performance for PERR1 to compare it to the performance of PERR2. It is evident from the graph that PERR1 gives a better overall grade of service than PERR2.

PERR3 and PERR4 give overall and point-to-point grades of service that are very nearly the same as those given by PERR2. This would suggest that very few occasions arise when both alternate routes are "available".

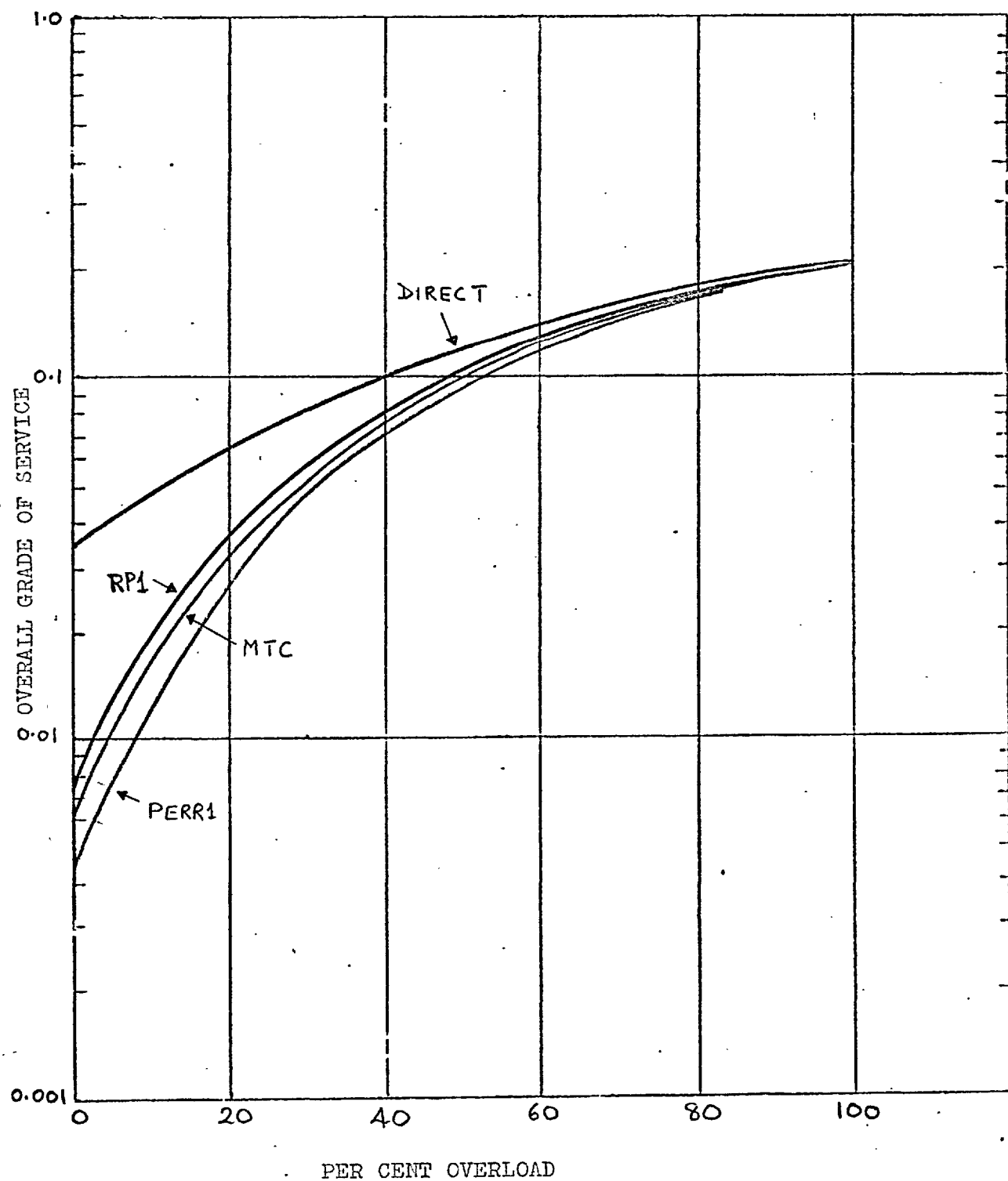


Fig.10.4 A comparison of the overall grades of service given by the different routing strategies

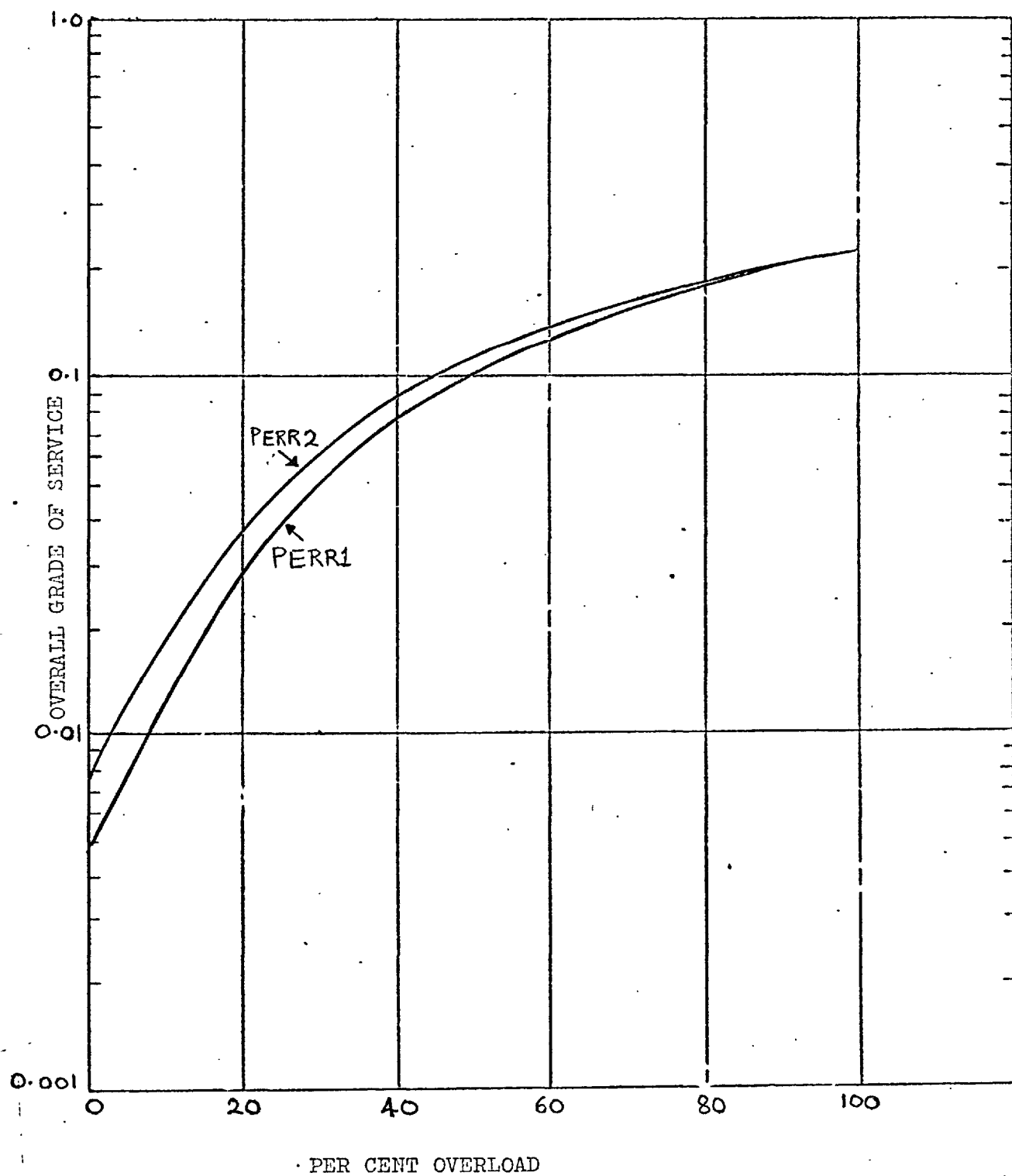


Fig.10.5 A comparison of the overall grades of service given by the PERR1 and PERR2 strategies

10.3.3 Comparison of the point-to-point grades of service

We have already made the point that we should not try to optimize the overall grade of service at the expense of some parcels of traffic. We now show how we can give some protection to these parcels of traffic. Let us focus our attention on the traffic between exchanges A and B which needs, by far, the greatest protection in our four-exchange example.

We have plotted the grade of service for this parcel of traffic in Fig.10.6 for the direct routing strategy, PERR1 and PERR2. Notice that PERR1 which gave the best overall grade of service performance gives the best results for the very small overloads, but deteriorates catastrophically for the higher overloads. PERR2, on the other hand, almost always gives worse results than the direct strategy, but for high overloads performs much better than PERR1.

10.4 Conclusions

The choice between PERR1 and PERR2 must remain open. If the overloads are unlikely to exceed 15 per cent, then PERR1 is better than PERR2. But in order to provide more effective protection for parcels of traffic like AB, especially if high overloads are likely, we would prefer PERR2.

Appendix 10.5 gives the detailed results of the strategies that we have compared. The reader may have noticed that the four-exchange network we have taken as an example has been investigated by Weber(31). The results in the appendix to this chapter will enable the reader to make detailed comparisons if he so wishes.

Before concluding this comparison, I wish to make one final point about the protection of parcels of traffic. For both PERR1 and PERR2 we have used the value 0.5 as the threshold for our decision; only if the probability of error is less than 0.5, do we allow the pertinent trunk group to be used for alternate routing.

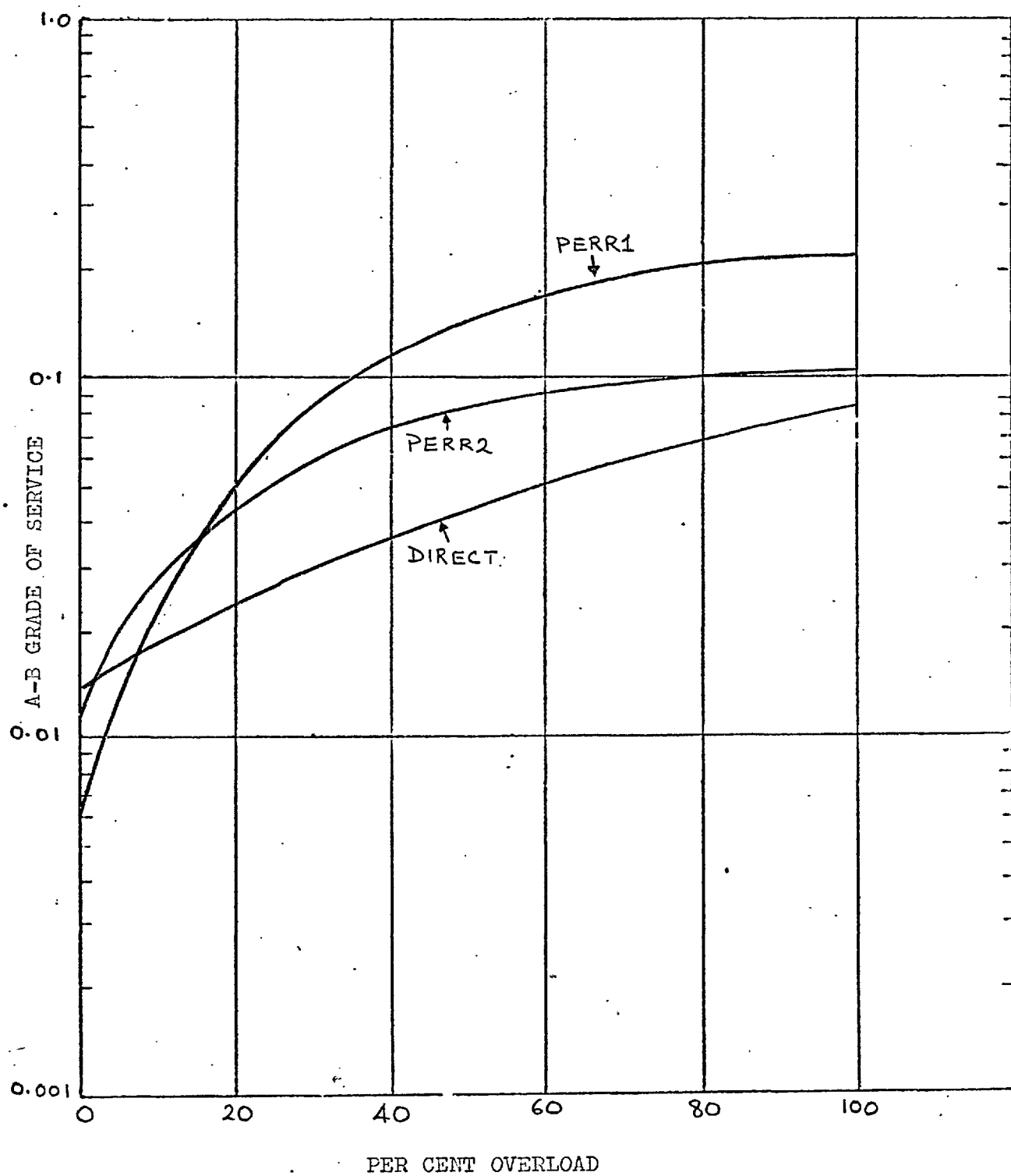


Fig.10.6 A comparison of the A-B point-to-point grades of service

By decreasing this threshold slightly to, say 0.45 or some empirically or otherwise decided value, we may be able to give greater protection to parcels of traffic. Of course the threshold need not be the same for all trunk groups in the network. Further investigation of traffic parcel protection by optimizing the threshold is needed, but we shall pursue this problem no further in this thesis.

COMPARISON OF ROUTING RESULTS

10.5 Appendix: Tables of the point-to-point and overall grades of service for the network of Figs. 10.3a and 10.3b

All the results in this appendix are approximately within ± 0.0004 of the stated values with a 95 per cent confidence level. A dash instead of a result means that that particular result was not obtained.

<u>OVERLOAD = 0%</u>					
<u>LINK</u>	<u>MTG</u>	<u>RP1</u>	<u>PERR1</u>	<u>PERR2</u>	<u>DIRECT</u>
AB	0.010	0.011	0.006	0.012	0.0154
AC	0.007	0.006	0.007	0.007	0.0304
AD	0.006	0.010	0.009	0.010	0.0367
BC	0.007	0.009	0.005	0.010	0.0522
BD	0.005	0.006	0.003	0.005	0.0375
CD	0.006	0.009	0.004	0.010	0.0431
OVERALL	0.007	0.008	0.005	0.008	0.039

OVERLOAD = 10%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	0.019	0.025	-	0.030	0.0204
AC	0.016	0.018	-	0.017	0.0444
AD	0.018	0.017	-	0.021	0.0488
BC	0.018	0.021	-	0.020	0.0697
BD	0.015	0.021	-	0.015	0.0548
CD	0.016	0.024	-	0.028	0.0634
OVERALL	0.017	0.021	-	0.021	0.055

OVERLOAD = 20%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	-	0.052	0.057	0.044	0.0262
AC	-	0.033	0.029	0.034	0.0609
AD	-	0.038	0.054	0.043	0.0624
BC	-	0.040	0.033	0.042	0.0891
BD	-	0.038	0.015	0.037	0.0751
CD	-	0.047	0.035	0.051	0.0869
OVERALL		0.040	0.031	0.042	0.0742

OVERLOAD = 30%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	0.063	0.077	0.094	0.070	0.0328
AC	0.061	0.054	0.057	0.053	0.0797
AD	0.075	0.058	0.100	0.068	0.0773
BC	0.071	0.063	0.062	0.076	0.1100
BD	0.062	0.056	0.026	0.061	0.0978
CD	0.042	0.072	0.052	0.084	0.1127
OVERALL	0.059	0.062	0.055	0.069	0.0953

OVERLOAD = 40%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	-	0.100	0.123	0.081	0.0400
AC	-	0.083	0.079	0.083	0.1002
AD	-	0.086	0.061	0.096	0.0933
BC	-	0.094	0.077	0.101	0.1318
BD	-	0.085	0.087	0.091	0.1221
CD	-	0.107	0.093	0.117	0.1400
OVERALL		0.093	0.085	0.098	0.1178

OVERLOAD = 50%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	0.151	0.133	0.167	0.093	0.0480
AC	0.106	0.113	0.100	0.112	0.1219
AD	0.160	0.133	0.107	0.115	0.1101
BC	0.117	0.120	0.105	0.127	0.1542
BD	0.104	0.107	0.103	0.117	0.1474
CD	0.108	0.139	0.130	0.157	0.1680
OVERALL	0.115	0.123	0.114	0.128	0.1411

OVERLOAD = 60%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	-	0.166	0.198	0.099	0.0565
AC	-	0.136	0.127	0.139	0.1444
AD	-	0.158	0.1411	0.138	0.1274
BC	-	0.145	0.136	0.150	0.1768
BD	-	0.131	0.135	0.150	0.1731
CD	-	0.170	0.154	0.192	0.1960
OVERALL		0.149	0.143	0.156	0.1648

OVERLOAD = 70%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	-	0.178	0.217	0.112	0.0655
AC	-	0.150	0.134	0.157	0.1673
AD	-	0.196	0.170	0.165	0.1452
BC	-	0.184	0.190	0.186	0.1994
BD	-	0.148	0.154	0.172	0.1989
CD	-	0.209	0.192	0.221	0.2238
OVERALL		0.177	0.171	0.181	0.1886

OVERLOAD 80%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	-	0.116	0.227	0.114	0.0750
AC	-	0.190	0.180	0.188	0.1903
AD	-	0.183	0.215	0.186	0.1631
BC	-	0.210	0.192	0.211	0.2217
BD	-	0.207	0.173	0.202	0.2243
CD	-	0.244	0.233	0.253	0.2508
OVERALL		0.208	0.200	0.209	0.2121

OVERLOAD 90%

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	0.228	0.130	0.252	0.113	0.0850
AC	0.214	0.212	0.210	0.215	0.2131
AD	0.235	0.204	0.235	0.194	0.1811
BC	0.221	0.237	0.216	0.236	0.2436
BD	0.207	0.231	0.199	0.237	0.2491
CD	0.254	0.274	0.257	0.277	0.2769
OVERALL	0.227	0.233	0.226	0.234	0.2351

OVERLOAD = 100

LINK	MTC	RP1	PERR1	PERR2	DIRECT
AB	0.247	0.134	-	0.117	0.0952
AC	0.226	0.237	-	0.241	0.2356
AD	0.259	0.219	-	0.210	0.1991
BC	0.257	0.257	-	0.257	0.2649
CD	0.293	0.302	-	0.304	0.3019
OVERALL	0.253	0.256	-	0.257	0.2575

CHAPTER 11 THE SIX-EXCHANGE NETWORK:
FURTHER ROUTING TECHNIQUES.

- 11.1 Introduction.
- 11.2 Routing in the fully connected six-exchange network.
 - 11.2.1 Performance comparisons.
 - 11.2.2 Overall grades of service.
 - 11.2.3 Point-to-point grades of service.
- 11.3 Networks in which some exchanges have no direct links to other exchanges.
 - 11.3.1 Exchange and route partitioning.
 - 11.3.2 Exchange lumping.
 - 11.3.3 Exchange lumping: an example.

CHAPTER 11. THE SIX-EXCHANGE NETWORK:: FURTHER ROUTING TECHNIQUES.

11.1 Introduction

We have up to now been devising routing strategies for the three- and four-exchange networks. We have compared these strategies in Chapter 10, and found that PERR1 and PERR2 possess many of the characteristics that we desire. In this chapter we move on to the six-exchange network.

The reader recalls that we started this study by devising routing rules for networks that were not only fully connected, i .e . every exchange had a direct link to every other exchange, but which also had an equal number of circuits in each trunk group. In Chapter 9 we removed the assumption that all trunk groups had an equal number of circuits. Now, in Section 11.3, we remove the remaining assumption that all exchanges are interconnected by direct links.

11.2 Routing in the fully connected six-exchange network.

The six-exchange network of Fig.11.1 that we are about to analyze was designed and used by Weber (31) to study alternate routing in communication networks.

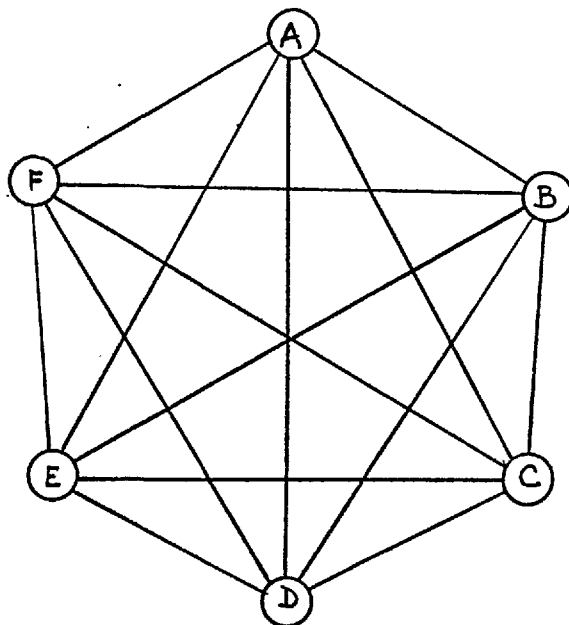


Fig.11.1 The fully connected six-exchange network.

The directly offered traffic, and the number of circuits per trunk group are given in Fig. 11.2.

Link	Number of circuits	Offered traffics (erlangs)
AB	4	1
AC	8	5
AD	4	1
AE	7	3
AF	5	2
BC	5	2
ED	9	6
BE	6	3
BF	7	4
CD	7	4
CE	9	6
CF	7	4
DE	8	5
DF	11	7
EF	11	7

Fig. 11.2 The number of circuits and the offered traffic for the network of Fig. 11.1.

A routing table which allows alternate routes of up to three links in tandem was constructed, and the network simulated for each of the following strategies:

1. PERR1
2. PERR2
3. MTC
4. Static sequential alternate routing (SEQ. ALT. ROUT.)
5. Sequential alternate routing with one trunk reserved (1 TRNK RES.) .

The computed results of the direct routing strategy, and the simulation results of the other strategies, including the 95 per cent confidence interval, are tabulated in Fig. 11.3.

OVER- LOAD %	PERR2	PERR1	MTC	SEQ. ALT. ROUT.	1 TRNK RES.	DIRECT
0	0.010 ± 0.0002	0.006 ± 0.0007	0.006 ± 0.0006	0.005 ± 0.0004	0.007 ± 0.0009	0.057
10	0.032 ± 0.0012	0.017 ± 0.0038	0.027 ± 0.0012	0.020 ± 0.0009	0.021 ± 0.0016	0.079
20	0.063 ± 0.0022	0.044 ± 0.0074	0.052 ± 0.0001	0.074 ± 0.0100	0.052 ± 0.0026	0.103
30	0.091 ± 0.0010	0.095 ± 0.0041	0.099 ± 0.0044	0.138 ± 0.0003	0.102 ± 0.0060	0.128
40	0.123 ± 0.0036	0.129 ± 0.0014	0.141 ± 0.0062	0.214 ± 0.0137	0.132 ± 0.0054	0.154
50	0.173 ± 0.0028	0.140 ± 0.0011	0.167 ± 0.0035	0.256 ± 0.0017	0.173 ± 0.0070	0.181
60	0.195 ± 0.0015	0.178 ± 0.0026	0.197 ± 0.0012	0.309 ± 0.0094	0.215 ± 0.0086	0.207
70	0.231 ± 0.0031	0.225 ± 0.0290	0.222 ± 0.0090		0.243 ± 0.0096	0.233
80	0.264 ± 0.0006	0.237 ± 0.0034	0.252 ± 0.0166	0.376 ± 0.0058	0.279 ± 0.0009	0.258
90	0.269 ± 0.0027	0.271 ± 0.0012	0.278 ± 0.0047	0.407 ± 0.0035	0.303 ± 0.0040	0.282
100	0.299 ± 0.0082	0.300 ± 0.0084	0.301 ± 0.0041	0.455 ± 0.0026	0.341 ± 0.0115	0.305

Fig.11.3 The overall grades of service for the various routing strategies with the 95% confidence interval.

11.2.1 Performance comparisons

The conclusions we came to regarding the relative performances of PERR1, PERR2 and MTC in the four-exchange network have, in general, been reinforced in this six-exchange network. The traffic interactions are now much more complex, however, so that it is no longer easy to give clear cut preferences between the different strategies for a large overload range.

11.2.2 Overall grades of service.

PERR1 gives a better overall grade of service than MTC over the whole overload range considered. PERR2 is, as before, ^{definitely} inferior to PERR1 for the small overloads of up to about 25 per cent. Above that, within the confidence level of 95 per cent, PERR1 is sometimes still superior to PERR2.

The sequential alternate routing strategy performs just as well as PERR1 at the rated load (0 per cent overload), but as usual, deteriorates much more rapidly at overloads above ten per cent. PERR1 is superior to the sequential alternate routing strategy with one trunk reserved.

On the strength of the results of the results of Fig.11.3 giving the overall grades of service, I would choose as my routing strategy PERR1 which at between ten and twenty per cent overload carries approximately six per cent more traffic than the direct strategy.

11.2.3 Point-to-point grades of service

As in the four-exchange network, we have chosen the link AB which has ^{the} most vulnerable point-to-point grade of service in order to compare the protection given by the different routing strategies to the small parcels of traffic. The results, with the 90 per cent confidence interval, are tabulated in Fig.11.4.

OVER- LOAD %	PERR2	PERR1	MTC	SEQ. ALT. ROUT.	1 TRNK. RMS.	DIRECT
0	0.029 ±0.019	0.010 ±0.0065	0.010 ±0.0067	0.010 ±0.0064		0.0154
10	0.038 ±0.0245	0.032 ±0.0024	0.027 ±0.0054	0.024 ±0.0024	0.013 ±0.0085	0.0204
20	0.050 ±0.0234	0.082 ±0.0035	0.077 ±0.0047	0.078 ±0.0049	0.042 ±0.0040	0.0262
30	0.073 ±0.0448	0.149 ±0.0051	0.130 ±0.0054	0.155 ±0.0223	0.033 ±0.0025	0.0328
40	0.122 ±0.0146	0.181 ±0.0164	0.205 ±0.0197	0.204 ±0.0005	0.073 ±0.0151	0.0400
50	0.139 ±0.0337	0.231 ±0.0190	0.288 ±0.0074	0.313 ±0.0010	0.116 ±0.0070	0.0480
60	0.158 ±0.0405	0.292 ±0.0218	0.316 ±0.0266	0.364 ±0.0393	0.110 ±0.0106	0.0565
70	0.165 ±0.0042	0.414 ±0.0258	0.376 ±0.0045	0.379 ±0.0155	0.116 ±0.0033	0.0655
80	0.214 ±0.0317	0.355 ±0.0222	0.475 ±0.0303		0.138 ±0.0230	0.0750
90	0.229 ±0.0137	0.457 ±0.0016	0.453 ±0.0304	0.448 ±0.0319	0.171 ±0.0300	0.0850
100	0.212 ±0.0076	0.390 ±0.0483	0.498 ±0.0094	0.471 ±0.0885	0.198 ±0.0696	0.0952

Fig. 11.4 The A-B point-to-point grades of service with the 90% confidence level.

But before proceeding with the comparison, let us make two points.

Firstly, the reader may have noticed from Fig. 11.4 that the confidence intervals for the small parcels of traffic like AB are quite often too big even for the 90 per cent confidence level. This is not altogether suprising.

If the smallest parcel contributes only 1/50 of the total number of simulated calls, and the grade service of this parcel is say 0.01, then out of a total of 100,000 calls processed only (1/50) (1/100) (100,000) or twenty calls will be blocked. This number is not big enough to give

a reliable estimate of the grade of service. A further reason for the big confidence intervals, discussed in detail by Karlsson (52), is that the small trunk groups take a very long time to attain statistical equilibrium. This confidence problem did not arise in the four-exchange network because the simulator used was much faster than the one used in the six-exchange network, and therefore we could process orders of magnitude more calls.

Secondly, the reader may also have noticed that the values of the point-to-point grades of service in PERR1, PERR2 and MTC are not necessarily monotone increasing with the offered traffic. See, for example, MTC between 80 and 90 per cent overload, and PERR1 between 70 and 80 per cent.

As far as the comparisons of the point-to-point grades of service are concerned, we come to the same conclusions as we did in the four-exchange network. We note, however, that the sequential alternate routing strategy with one trunk reserved for direct traffic provides, as is to be expected, the strongest protection for the small parcels of traffic.

Finally, we recognize an additional factor in the six-exchange network that did not play an important part in the four-exchange network. The small trunk groups like AB and AD (4 trunks each), and AF and BC (5 trunks each) are less protected in the six-exchange network than they were in the four-exchange network. This is because we can now establish many two- or three-link routes that are entirely made up of small trunk groups. In the four-exchange network, on the other hand, we quite often had to use a big trunk group in tandem with a small one. Thus the big trunk group by trying to protect itself also indirectly gave protection to the small group.

11.3 Networks in which some exchanges have no direct links to other exchanges.

In all the networks we have considered so far, we have assumed the existence of direct junctions or trunk groups

between all the exchanges. In practice for a network of several exchanges, this arrangement is economically unsuitable. We therefore propose two methods, exchange partitioning and exchange lumping, for dealing with networks that are not fully connected.

11.3.1 Exchange and route partitioning.

Consider the network of Fig. 11.5a where n_{ij} is the number of circuits between exchanges i and j .

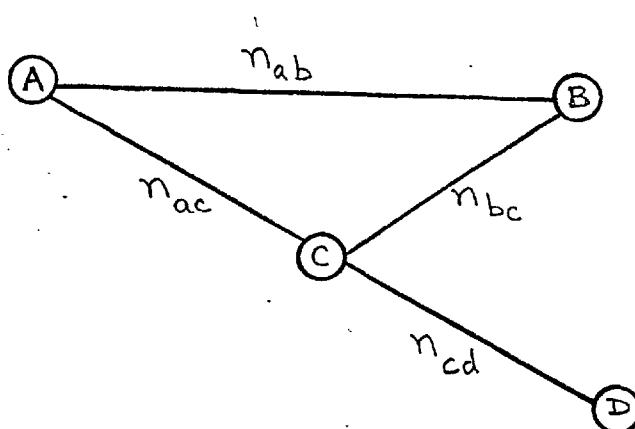


Fig. 11.5a A four-exchange network that is not fully connected

In this case, calls from exchange D to A or B have no direct routes. We can proceed by partitioning the routes AC, BC and CD, and exchange C as in Fig. 11.5b.

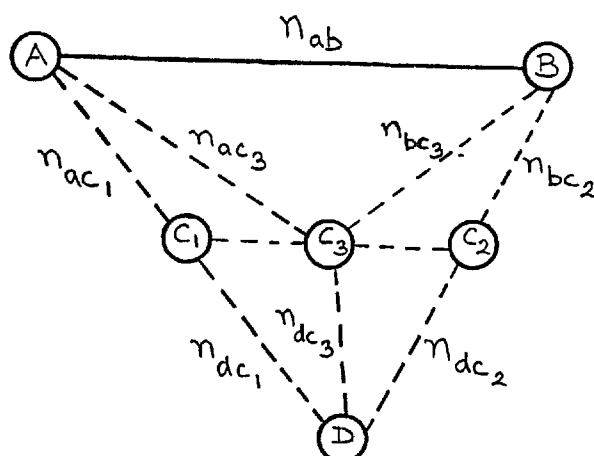


Fig. 11.5b The network of Fig. 11.5a partitioned

Because we have partitioned exchange C into C_1 , C_2 and C_3 , we assume that the circuits connecting C_1 and C_3 , and C_2 and C_3 are infinitely many. It should be understood that

$$n_{ac_1} + n_{ac_3} = n_{ac}$$

$$n_{bc_2} + n_{bc_3} = n_{bc}$$

$$\text{and } n_{dc_1} + n_{dc_3} + n_{dc_2} = n_{cd}$$

We can now use the same methods that have been used before. The partitioning of the routes may be done in the proportion of the traffic offered between the various exchanges. This method of partitioning routes has been used to safeguard small parcels of traffic in the Danish network, Asgersen(102).

11.3.2 Exchange Lumping

= A method that is complementary to partitioning is exchange lumping. (In network flow theory this process is better known as "node condensation"). Consider the six-exchange network of Fig.11.6.

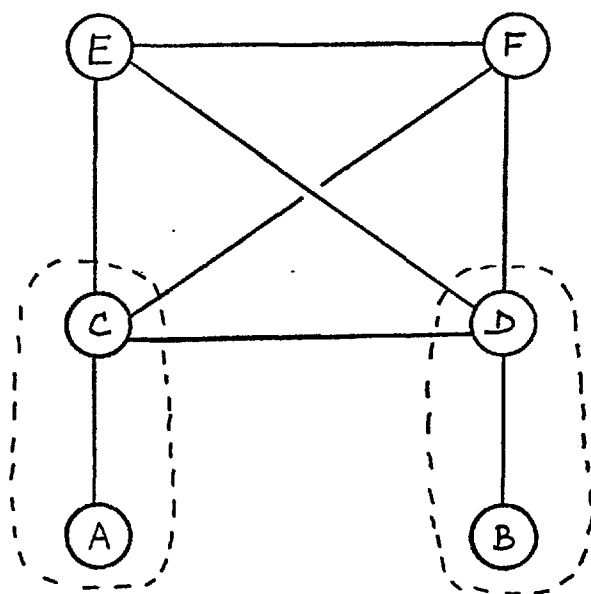


Fig.11.6 A six-exchange network

We can effectively reduce ~~this~~ six-exchange network to a four-exchange one by lumping exchanges A and C , and B and D . This method of lumping is important because it makes it possible to use state-dependent routing rules at the highest exchange levels in a hierarchical network. We can for example use state-dependent rules in the network of Main Switching Centres in the B.P.O. network, or Regional Centres in the U.S. network. Even at the present time, it may be economical to use state dependent routing procedures at these levels.

11.3.3 Exchange lumping : an example.

Let us work through an example using exchange lumping on Fig.11.6. The table of Fig.11.7 gives the number of circuits connecting each pair of exchange the directly offered traffic, and the "direct" route between each pair of exchanges.

Link	Number of circuits	Offered Traffic (erlangs)	Direct route
AB	0	1	A-C-D-B
AC	16	2	A-C
AD	0	1	A-C-D
AE	0	3	A-C-E
AF	0	2	A-C-F
BC	0	2	B-D-C
BD	20	2	B-D
BE	0	3	B-D-E
BF	0	4	B-D-F
CD	15	4	C-D
CE	11	2	C-E
CF	13	4	C-F
DE	15	5	D-E
DF	14	3	D-F
EF	10	4	E-F

Fig.11.7 Number of circuits, exchange-to-exchange offered traffic, and the "direct" routes.

Since some pairs of exchanges have no direct connection, we can no longer compute the performance for the direct strategy. A simulation program was therefore written to determine this performance.

The use of PERR2 in this network also needs some explanation. The procedure is this:

1. For each link we determine the total amount of "directly" offered traffic. Take the link AC as an example. This link carries $1+2+1+3+2=9e$ made up of the traffic parcels AB, AC, AD, AE and AF respectively.
2. Compute the probabilities of routing errors for each link using the total amount of directly offered traffic, and the number of circuits in the link.
3. Set up an alternate routing table that allows up to two links in tandem for each link that forms part of the direct route. For example the routes A-C-D-B and A-C-E-D-B are allowed for an AB call. The routes C-D, C-E-D, C-F-D are allowed, but the route C-E-F-D is not allowed for a CD call.

During the simulation, a call from A is treated as if it was a call from exchange C. Similarly for B and D. But in searching for alternate routes, the conditions for the PERR2 routing strategy must be satisfied.

The simulation results for both the direct and the PERR2 routing strategies are shown in the table of Fig. 11.8.

OVERLOAD %	PERR2 STRATEGY		DIRECT STRATEGY	
	BLOCKING	95% CONFIDENCE INTERVAL	BLOCKING	95% CONFIDENCE INTERVAL
0	0.005	± 0.0002	0.011	± 0.0007
10	0.010	± 0.0018	0.020	± 0.0013
20	0.017	± 0.0001	0.034	± 0.0009
30	0.034	± 0.0004	0.051	± 0.0033
40	0.052	± 0.0008	0.066	± 0.0001
50	0.078	± 0.0018	0.092	± 0.0094
60	0.111	± 0.0016	0.123	± 0.0069
70	0.125	± 0.0011	0.144	± 0.006
80	0.146	± 0.0045	0.157	± 0.008
90	0.170	± 0.0045	0.173	± 0.0055
100	0.201	± 0.0009	0.200	± 0.0014

Fig. 11.8 Simulation results for the lumped six-exchange network. (overall grade of service).

Yet again we see the excellence of the state-dependent routing rules (PERR2 this time). The PERR2 overall grade of service is superior to that given by the direct strategy up to over 90 per cent overload. The point-to-point grades of service for the relatively big trunk groups of this network have been good.

Finally, we emphasize that this exchange lumping example has been designed to show how the routing strategies that we have developed in this thesis can be applied to networks in which some exchanges have no direct links to other exchange.

CHAPTER 12 CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

12.1 Conclusions

12.2 Some aspects of the routing problem needing further research.

12.2.1 Data processing requirements.

12.2.2 Optimal routing rules.

12.2.3 Analytical evaluation of network performance.

12.2.4 State-dependent rules in a hierarchical network.

12.2.5 Routing with imperfect information.

12.2.6 Economic considerations.

12.2.7 Reduction in network status data.

12.2.8 Data and message traffic.

CHAPTER 12. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

12.1 Conclusions

This thesis has been a preliminary study of state-dependent routing rules in a telephone network of exchanges.

By working through some examples of routing in a switch or a network of switches, we have come to the conclusion that the methods used in these examples are inadequate for attacking our problem. We have then drawn analogies between our problem and other problems in network flow, Markov processes with rewards, and inventory systems only to find that those problems are still unsolved in their fields.

It is, therefore, with a sigh of relief that we have come up with the concept of "time to congestion". We have then used this concept in devising routing strategies. These strategies, we maintain, are good strategies. They provide a good overall (network) grade of service, and point-to-point grades of service that are acceptable. In particular, we would recommend PERR1 and PERR2 for further study and development.

A matter for urgent consideration is the establishment of a better (perhaps optimal) "threshold" for use in these strategies. We have used the value of 0.5; we could not have used a higher value. But somewhere between 0.0 and 0.5 there must be a better threshold. If we hit on this threshold, we could produce results that are significantly better, in both overall and point-to-point grades of service, than those we have obtained.

All the same, we believe that there must always be a trade-off between the overall and the point-to-point grades of service. We can, up to a limit, improve the overall grade of service at the expense of some point-to-point grades of service. This compromise between overall and point-to-point grades of service remains the responsibility of the telephone administration. We, on our part, have shown how this balance can be regulated.

12.2 Some aspects of the routing problem needing further research

12.2.1 Data processing requirements

The routing techniques described in this study are all dependent on

continuously monitoring the network, collecting traffic flow information and sending data to some central processor which in turn issues routing instructions to all the exchanges in the network.

This situation makes great demands on the data processing capability of the exchanges. It is assumed that not only does each exchange have some sort of special purpose computer, but that it can also access and process information held in a common data bank. To say that this data problem is very complex is perhaps an understatement. It is, therefore, important that research should be undertaken to determine if data processing of this magnitude can be carried out efficiently, reliably and economically.

12.2.2 Optimal routing rules

Research effort should also be channeled into finding methods of determining optimal routing rules. In this thesis, we have used an heuristic approach in devising routing rules. We have then evaluated the network performance under these rules, and chosen the rule that leads to the best performance amongst those considered. We have at no stage proved a routing rule to be optimal in any way. Therefore, depending on the criteria for optimality, it may be possible to find routing rules that are not only better than those we have used, but are, in fact, optimal. We may be able to use some form of mathematical programming to obtain the most efficient rules. It may turn out that, like the transportation problem, our problem has a special structure that can be easily exploited.

12.2.3 Analytical evaluation of network performance

The method for evaluating the performance of a network under a given routing rule used in this thesis has been computer simulation. Katz(98, 99) has devised an algorithm for analytically evaluating network performance for certain alternative routing networks with progressive route control. This method has been tested for networks having more than fifty exchanges, and the results have compared very favourably with those obtained by computer simulation. It is worth pointing out that the computer run time required in an analytical evaluation is usually a small fraction of that required for simulation. The availability of analytical methods for evaluating a network's performance under more complex routing rules would, therefore, be a great asset to routing studies.

12.2.4 State-dependent rules in a hierarchical network

The routing rules studied in this thesis have been devised on the implicit assumption that all the exchanges have equal capabilities. It is not conceivable that a practical telephone network could be completely free of any form of hierarchy. The present work should therefore be extended to include the hierarchical situation. Since the hierarchical network is, in fact, the equal-rank exchange network with certain routing constraints, it should not be too difficult to determine its performance under any of the routing rules discussed. Of course, severe routing constraints will reduce the networks traffic carrying capacity.

12.2.5 Routing with imperfect information

Perfect information about the state of the network has been assumed in all the routing rules used. This is the ideal situation. In practice, errors will pass through the data processing system undetected, and the routing control should be able to recover quickly from such errors. This in fact raises an interesting question. Can we devise a routing scheme that performs almost as well as those we have been using, but which needs considerably less information? The reduction of the data processing requirements could well turn the scales of the economic feasibility of using complex routing schemes. Dejean and Grandjean (70) have made an approach in this data reduction field, see Section 12.2.7.

12.2.6 Economic considerations

We must not be so engrossed in devising routing schemes that we lose sight of the economic factors. No attempt has been made to cost the routing schemes that have been considered in this thesis. This is not to say that the economics is not important. It means rather that we have chosen to concentrate our initial effort in trying to determine just by how much we can expect to increase the traffic carrying capacity of the network. We have not even tried to determine or quantify the enhancement of the reliability and survivability of the network. But after an extensive study has been made of the routing techniques, the logical and natural step will be to look at the economics of the class of proposed systems and determine if they make economic sense.

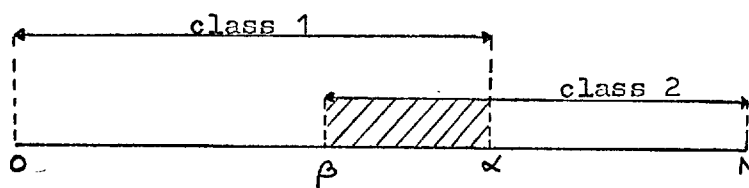
If it should turn out that all the propositions are outrageously inappropriate from an economics point of view, it will not mean that the time and effort spent on the more general routing plans has been wasted.

Quite apart from the discussion and critical review of the more classical routing plans it will have initiated, it will have served a most useful purpose. Has it not been said that the so-called 'negative' results are sometimes more useful than their more popular colleagues the "positive" results?

12.2.7 Reduction in network status data

If every change in the exact number of busy trunks is notified to the routing control centre, the information traffic would be prohibitively large.

An approximate knowledge of the occupation level may be sufficient for an efficient routing process. This approximate knowledge may be, for instance, that the number of busy trunks in a group of N trunks is smaller than α or at least equal to β , for β less than α . We then have two overlapping classes.



In this case we know that the number of busy trunks at time t is between β and α . Dejean and Grandjean (70) have studied the reduction in the information traffic that can be achieved by having these overlapping classes. Their results show that the information traffic may be reduced approximately in the proportion of the overlapping number of trunks.

12.2.8 Data and message traffic

This thesis has addressed itself to problems of telephone traffic routing. But other "on demand" services like telex and data transmission have similar problems, but perhaps with different emphases. For example, in the data problem, holding times tend to be approximately equal to or even less than the call setting up time.

This is not the case for telephone traffic. So while we were justified, at least as a first approximation, in placing the emphasis on the holding times, this emphasis would be misguided for data traffic. If it is true that in data most of, or a large proportion of the time is spent in

setting up the call, then it is the switching centre speed that should receive first attention. These problems should receive some research attention whether we are interested in an integrated data and telephone network, or separate networks.

Otterman (85, 86) has treated the related problem of increasing the trunk group utility, especially of long haul links, by mixing short holding time Store-and-Forward (S/F) message traffic with telephone traffic for various ratios of the holding time of telephone traffic to S/F traffic holding time. He has also investigated the effect on the grade of service of the telephone traffic, and the average message delay of the S/F traffic if the two types are mixed. This work could be usefully extended for the "on demand" telephone and data situation.

REFERENCES

- (1) Grandjean, Ch. - "Call routing strategies in
Telecommunication Networks"
Electrical Communication 42(1967) ; 3,
pp380-391
- (2) Feller, W. - "An introduction to probability theory
and its applications"
Wiley, New York (1957) 2nd Ed. Vol. 1
- (3) Feller, W. - "An introduction to probability theory
and its applications"
Wiley, New York Vol.2
- (4) Hu, T.C - "Integer Programming and Network Flows"
Addison-Wesley (1969) Reading, Mass., U.S.A
- (5) Ford, L.R. - "Flows in Networks"
Fulkerson, D. Princeton University Press (1962)
- (6) Silk, D.J. - "Routing doctrines and their implementation
in message switching networks"
Proc. IEE Vol.116(1969) ; 10, pp1631-1638
- (7) Ling, S.T. - "Optimal allocation of channels in an
et al alternate route communication network"
Trans. IEEE, COM-12(1964); pp185-190
- (8) Wallstrom, B. - "Congestion studies in Telephone systems
with overflow facilities"
Ericsson Technics Vol.22(1966) ;3, pp187-351
- (9) Wilkinson, R.I. "Theories for Toll traffic engineering
in the USA.
BSTJ Vol.35(1956) ; 2, pp421-514
- (10) Rubin, M. - "Communication Switching" Systems
Haller, C.E. Chapman Hall, London (1966) .Reinhold,
New York (1966)
- (11) Wallstrom, B. - "Methods for optimising Alternate routing
networks"
Ericsson Technics Vol.25(1969); 1, pp3-28
- (12) Elldin, A. - "Elementary Telephone Traffic Theory"
Lind, G. - L.M.Ericsson, Stockholm (1967)
- (13) Palm, C. - "Intensit" atsschwankungen in
Fernsprechverkehr"
Ericsson Technics Vol. 44(1943); 3-189
- (14) Kosten, L. - "Uber sperrungswahrscheinlichkeiten bei
staffelschaltungen"
Elektr. Nachr. Techn. Vol.14(1937) ;
1, pp5-12

- (15) Gomory, R.E. "Synthesis of a Communication Network"
Hu, T.C. Jour. SIAI Vol. 12(1964); 2, pp348-369
- (16) Onaga, K. - "Optimum Flows in general communication
 Networks"
 Jour. Franklin Intitute Vol. 283(1967);
 4, pp308-327
- (17) Frank, H. "On the optimum Synthesis of statistical
Hakimi, S.L. communication nets-Pseudo-parametric
 Techniques"
 Jour. Franklin Institute Vol.284(1967);
 6; pp407-416
- (18) Frank, H. - "Dynamic communication networks with
El-Bardai, M. capacity constraints"
 Trans. IEEE COM-17(1969) ;4, pp432-437
- (19) Green, T.V. "Route control in Autovon Electronic
Brown, L.C. Switching Centrrers"
 Trans. IEEE COM-17(1969); 4, pp442-446

- (20) Boehm, B.W. - "Adaptive routing techniques for
Mcbley, R.L. distributed communication Systems"
 Trans. IEEE COM-17(1969) ;4, pp340-349
- (21) Kruithof, J. "Theoretical and practical aspects of
 telephone traffic"
 Electrical Communication Vol.38(1963) ;
 2, pp252-263
- (22) Syski, R. - "Introduction to congestion theory in
 telephone systems"
 Oliver and Boyd, Edinburg and London (1960)
- (23) Atkinson, J. - "Telephony. Vol.2"
 Pitman(1950)
- (24) Benes, V.E.- "Mathematical theory^{of} connecting networks
 and telephone traffic"
 Academic Press, New York (1965)
- (25) Benes, V.E. "Programming and control problems arising
 from optimal routing in telephone networks"
 BSTJ Vol. 45(1966) ; pp1373-1438
- (26) Kalaba, R.E. "Optimal design and Utilisation of
Juncosa, H.L. communication networks"
 Management Science Vol. 3(1956); pp33-44
- (27) Kalaba, R.E. "Optimal utilisation and extension of
 interoffice trunking facilities"
 Communication and Electronics Vol.
 40(1959) ; 1, pp998-1003
- (28) Truit, C.J. "Traffic engineering techniques for deter-
 mining trunk requirements in alternate
 routing trunk networks"
 BSTJ Vol.33(1954), p277

- (29) Wright, E.P.G. "Flexible routing plans"
Electrical Communication Vol.41(1966);
1, pp48-54
- (30) Weber, J.H. "Some traffic characteristics of communication networks with automatic alternate routing"
BSTJ Vol.41(1962) ; pp769-796
- (31) Weber, J.H. "A simulation study of routing and control in communication networks"
BSTJ Vol.43(1964) ; pp2639-2676
- (32) Fu, Yuen-sun "Optimal routing procedures in distributed switching networks"
IEEE International Convention Record(1964),
Part 5 , pp250-260
- (33) Whybourne, J. "An approach to selecting switching equipment to develop a national telephone switching system"
Telecom. Jour. Vol.35(1968); 11, pp591-606
- (34) Prabhu, N.U. "Stochastic Processes"
Macmillan, New York(1965)
- (35) Kemeny, J.G. "Finite Markov Chains"
Snell, J.L. Van Nostrand, New York and London(1960)
- (36) Chung, K.L. "Markov Chains with stationary transition probabilities"
Springer, Berlin(1960)
- (37) Chung, K.L. "Contributions to the theory of Markov chains Part 1"
Jour. Res. Nat. Bur. Stand. Vol.50(1953) ;
pp203-208
- (38) Chung, K.L. "Contributions to the theory of Markov chains. Part 2"
Trans. Amer. Math. Soc. Vol.76(1954), pp397-419
- (39) Cohen, J.W. "The single server queue"
North-Holland, Amsterdam(1969)
- (40) Mina, R.R. "Modifications of the input process used in classical traffic theory"
International Teletraffic Cong. Munich, Sept '70
- (41) Clarke, A.B. "Probability and random processes for Engineers and Scientists"
Disney, R.L. J.Wiley, New York(1970)
- (42) Gordon, G. "System simulation"
Prentice-Hall, Englewood Cliffs, N.J.(1969)
- (43) Smith, J.R.W. "System simulation techniques"
Smith, J.L. Plessey Comm. Jour. Vol.1(1966); 1, pp34-37
- (44) Mitchell, G.H. "Simulation"
Inst. Maths. and its Application(GB), Vol.6
1969 ; 3, pp59-62

- (45) Bernard, P. "Solution by means of computers of call and trunk routing problems"
L'Echo des Recherches Vol.48(1967); Jan., pp4-13
- (46) Bernard, P. "Solution by means of computers of routing problems in linked telecommunication networks"
Annales des Telecom. Vol.21(1966) ; pp11-30
- (47) Hakimi, S.L. "Simultaneous flows through a communication network"
IRE Trans. Vol. CT-9(1962) ; June, pp169-175
- (48) Tang, D.T. "Communication networks with simultaneous flow requirements"
IRE Trans. Vol. CT-9(1962) ; June, pp176-182
- (49) Hu, T.C. "Multi-commodity network flows"
Jour. Oper. Res. Vol.11(1963); May, pp344-360
- (50) Parr, J.C.
Eisner, A.M. "Improved trunk allocation in a congested network"
West. Union Tech. Review, 1969; No.3, pp96-103
- (51) Kotov, A.V. "The development of electronic or quasi-electronic telephone systems"
Telecom. Jour. Vol.35(1968); 9, pp455-462
- (52) Karlsson, S.A. "Dimensioning telephone routes"
Res. Inst. Helsinki Telephone Co., May 1967
- (53) Magnani, R. "Some theorems and procedures for loop-free routing in directed communication networks"
BSTJ Vol.47(1968) ; 4, pp465-486
- (54) MacGuire, C.H. "Network concepts for worldwide automatic telephone service"
Telecom. Jour. Vol.34(1967) ; 1, pp25-28
- (55) Otterman, J. "Matrix multiplication in search for alternate routes"
Electr. Comm. Vol.38(1963) ; 2, pp246-251
- (56) Ott, K.W.
De Los Rios "Computation of urban trunking networks with alternate routing by computer"
Electr. Comm. Vol.43(1968) ; 2, pp157-162
- (57) Rapp, Y. "Planning of junction networks in a multi-exchange area. Part 1 . General Principles"
Ericsson Technics Vol.20(1964); 1, pp77-130
- (58) Rapp, Y. "Planning of junction networks in a multi-exchange area. Part 2. Extensions of the principles and applications"
Ericsson Technics Vol.21(1965); 2, pp187-240
- (59) Schwartzel, H. "On automatic traffic control in telephone networks"
5th Intern. Teletraffic Cong., New York 1967

- (60) Truitt, C.J. "Traffic engineering techniques for determining trunk requirements in alternate routing trunk networks"
BSTJ Vol.33(1954) ; 2, pp277-302
- (61) Wormald, E.G. "Consecutive overflow telephone traffic graphs"
Elec. Eng. Trans. Inst. Eng., Australia, Vol.44(1968) ; Sept., pp168-172
- (62) Wernander, M. "Systems engineering for communication networks"
IEEE Trans. Comm. and Electronics(1964); Nov., pp603-611
- (63) Wolman, E. "On definitions of congestion in communication networks"
BSTJ Vol.44(1965) ; pp2271-2294
- (64) Wion, F.W. "Electronic exchanges with Stored Program Control"
Telecom. Jour. of Aust.(1968); June, pp141-146
- (65) Yaged, B. "Traffic flow information for minimum cost routing procedures"
Automatica Vol.5(1969); pp167-173
- (66) Chernoff, H. "Rational selection of decision functions"
Econometrica Vol.22(1954); pp422-443
- (67) Luce, R.D.
Raiffa, H. "Games and Decisions"
J.Wiley, New York(1957)
- (68) Milnor, J. "Games against Nature" in H.Shubik(Ed)
'Game theory and related approaches to social behaviour'
London and New York(1964)
- (69) Hall, A.D. "A methodology for Systems Engineering"
Van Nostrand , Princeton N.J.(1962)
- (70) Dejean, J.H.
Grandjean, Ch. "Transitions in the occupation conditions of a group of switches"
Electrical Comm. Vol.42(1967); 2, pp261-269
- (71) Riordan, J. "Stochastic service systems"
Wiley, New York 1962
- (72) Sunde, E.D. "Communication systems engineering theory"
Wiley, New York 1969
- (73) Dejean, J.H,
Cohen, D.
Grandjean, Ch.
Seneque, P. Dispersed telecommunication network structure
Electr. Comm. Vol.42(1967); 3, pp402-414
- (74) Benes, V.E. "Optimal routing in connecting networks over finite time intervals"
BSTJ Vol.46(1967); 10, pp2341-2351
- (75) Kleinrock, L. "Communication nets: Stochastic message flow and delay"
McGraw-Hill, New York(1964)
- (76) Dreyfus, S.E. "Dynamic programming and calculus of variations"
Academic Press, New York(1965)

- (77) Hambrock, H. "Saturation signalling - towards optimum
Svala, C.G. alternate routing"
Prince, L.J. 9th Nat. Comm. Symp. Oct. 1963, Utica, N.Y.
- (78) Parikh, "Finding the most reliable routes in
Frisch, communication systems"
Trans. IEEE Vol. CS-11(1963); pp402-406
- (79) Frisch "Optimum routes in communication systems
with channel capacities and channel
reliabilities"
Trans. IEEE Vol. CS-11(1963) ; pp241-245
- (80) Ford, L.R. "Suggested computation for maximal multi-
Fulkerson, D. commodity network flows"
Management Science Vol.5(1958); pp97-101
- (81) Tomlin, J.A. "Minimum cost multi-commodity network flows"
Oper. Res. Jour. Vol.14(1966); pp45-51
- (82) Bellman, R "On the concepts of a problem and problem
Brock, P. solving"
Amer. Math. Monthly Vol.67(1960); pp119-134
- (83) Hakimi, S.L. "Analysis and Design of the Communication
networks with memory"
Jour. Franklin Inst. Vol.287(1969);1, pp1-18
- (84) Weber, J.H. "UNISIM - A simulation program for
communication networks"
Fall Joint Computer Conf., San Francisco
Oct. 1964
- (85) Otterman, J. "Grade of service of direct traffic mixed
with Store and Forward traffic"
Trans. IEEE Vol. CS-10(1962);1, pp130-138
- (86) Otterman, J. "Probability distribution of delays for
Store and Forward traffic mixed with
direct traffic."
Trans. IEEE Vol. CS-11(1963);1, pp41-48
- (87) Grinold, R.C. "A multi-commodity max-flow algorithm"
Jour. Oper. Res. Vol.16(1968);6, pp1234-1238
- (88) Grinold, R.C. "A note on multi-commodity max-flow"
Jour. Oper. Res. Vol.17(1969);4, pp755
- (89) Schehrer, R. "Optimal design of alternate routing systems"
5th Intern. Teletraffic Cong., New York 1967
- (90) Pratt, C.W. "The concept of marginal overflow in
alternate routing"
5th Intern. Teletraffic Cong., New York, 1967
- (91) Gupta, J.L. "Alternative routing"
Telecommunications(India), Vol.16;1, pp124-130
- (92) Gupta, J.L. Telecommunications(India), Vol.16;2, pp194-201

- (93) Olsson, K.M. "Simulation on Computers. A method for determining the traffic carrying capacities of telephone systems"
Tele(Eng. Ed.) Vol.22(1970); 1, pp39-58
- (94) Gosztony, G. " 'Telmod' - a simulating equipment for the loss analysis of switching networks"
Endrody, T. Telecom. Review(Hungary), 1968, No.3-4, pp1-21
- (95) Lehmann, E.L. "Testing Statistical hypotheses"
Wiley, New York 1959
- (96) Martin, F.F. "Computer modeling and simulation"
Wiley, New York 1968
- (97) Jefferis, R.P. "The application of dynamic programming to routing problems"
Fegley, K.A. Trans. IEEE Vol.SSC-1(1965); 1, pp21-26
- (98) Katz, S.S. "Statistical performance analysis of a switched communications network"
5th ITC New York, June 1967
- (99) Katz, S.S. "Trunk engineering of non-hierarchical networks"
6th ITC, Munich, Sept.1970
- (100) Arhnmung, J. "Network dimensioning with alternate routing in a multi-exchange area of a city"
6th ITC Munich 1970
- (101) Arhung, J. "Traffic carrying capacity on link routes in a transit exchange of thr type ARM20"
Teletechnik(Eng. Ed.) Vol.11(1967);1, pp30-41
- (102) Asgersen, Chr. "A Traffic Routing Strategy designed for Overload protection downwards the Hierarchy"
6th ITC Munich, Sept. 1970
- (103) Harva, M. "Continual Checking of route loading in telephone traffic"
6th ITC Munich, Sept. 1970
- (104) Schwartzel, H. "Telephone Information Systems and associated traffic theory problems"
6th ITC Munich, Sept. 1970
- (105) Caballero, P.A. "A method for Optimisation of telephone trunking networks with alternate routing"
6th ITC Munich. Sept. 1970
- (106) Schehrer, R. "The design of Alternate Routing System with regard to the Variance Coefficient"
Lotze, A. NTZ-GI (1968) ; 2/3 , pp52-56
- (007) Schehrer, R. "On the exact calculation of overflow systems"
6th ITC Munich, Sept.1970
- (108) Butrimenko, A.V. "Adaptive Routing Technique and Simulation of a Communication Networks"
6th ITC Munich, Sept. 1970

- (109)Kharkevich,A. "Some considerations about the construction of a communication network"
6th ITC Munich, Sept. 1970
- (110)Kosten,L. "On the measurement of congestion quantities by means of fictitious traffic"
PTT Bedriff, Vol.2(1948/49); pp15-25
- (111)Kosten,L. "Simulation in traffic theory"
6th ITC Munich, Sept. 1970
- (112)Rotenberg,A. "A new pseudo-random number generator"
Jour. Assoc. Comp. Mach. Vol.7(1960);pp75-77
- (113)Greenberger,M "Notes on a new pseudo-random number generator".
Ibid Vol.8(1961);pp163-167
- (114)I.B.M. "Random number generation and testing"
Ref. Manual C20-8011, New York 1959
- (115)Ralston,A.(Ed.) "Mathematical methods for digital computers"
Wilf,H.S. Wiley, New York(1960)
- (116)Dixon,W.J.(Ed) "BMD Biomedical computer programs"
Univ. Calif.,Health Sciences Facility
Los Angeles, Sept.1965
- (117)Jacobaeus,C. "Telephone traffic research in Sweden"
Ericsson Review, Vol.47(1970);3,pp89-98
- (118)Kemperman,J. "The passage problem for a stationary Markov chain"
Chicago Univ. Press(1961)
- (119)Hamsher,D.H. "Communication System Engineering Handbook"
MacGraw-Hill , New York(1967)
- (120)Martin,J. "Telecommunications and the computer"
Prentice-Hall, New York(1969)
- (121)Klein,H. "A survey of users' views of discrete simulation languages"
Simulation Vol.14(1970);5,pp225-230
- (122)Le Gall,P. "L'ergodisme et la convergence des simulations de phenomenes aleatoires"
Annales des Telecommunications, Vol.22(1967);
No.7-8, pp211-228
- (123)Deloraine,E.M. "Telecommunication traffic"
Electr. Communication, Vol.44(1969);1,pp4-13
- (124)Wright,E.P.G. "Grade of service considerations"
Electr. Communication Vol.45(1970);1,pp13-17
- (125)Huntley,H.R. "Some ideas regarding Economics of Telecommunication Engineering"
Electr. Communication Vol.42(1967);1,pp6-21

- (126)Markowitz,H.M. "SIMSCRIPT - A simulation programming language"
Hausner,B.
Karr,H. Prentice-Hall , Englewood Cliffs, N.J.
- (127)C.D.C. "SIMSCRIPT Reference Manual"
Control Data Corporation, California.
- (128)Adelaar,H.H. "The influence of Stored Program Control on the development of Communication Networks"
Electr. Communication Vol.45(1970);3,pp183-5
- (129)Wilkinson,R.I. "Theories for toll traffic engineering in the U.S.A."
BSTJ Vol.35(1956); pp421-514
- (130)Cosgrove,YT. "Economic considerations for communication systems"
IEEE Trans. Vol.16(1968);4, pp513-525
- (131)Povey,J.A. "The use of Electronic Digital Computers for telephone traffic engineering"
Cole,A.C. Post Office Electr. Eng. Jour. Vol.58(1965); 3, pp203-209
- (132)Akimaru,H. The derivatives of Erlang's B formula"
Nishimura,T. 4th ITC London 1964
- (133)Schehrer,R. "The design of alternate routing systems with regard to the variance coefficient"
Lotze,A.
- (134)Clos,C. "A study of non-blocking switching networks"
BSTJ Vol.32(1953); pp406-424
- (135)Kibble,W.F. "The determination of telephone traffic blocking in groups of trunk circuits"
G.E.C.-A.E.I. Telecommunications, No.37(1968), pp11-22
- (136)Barton,D.E. "The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal"
Dennis,K.E. Biometrika, Vol.39(1952);pp425-427
- (137)Ledermann,W. "Spectral theory for the differential equations of simple Birth and Death processes"
Reuter,G.E. Phil. Trans. Royal Soc. London, Vol.246, 1953-54; Ser. A , pp321-369
- (138)Bellman, R. "Introduction to Matrix analysis"
McGraw-Hill , London(1960)
- (139)Ewens,W.J. "The mean time to absorption in a process of genetic type"
J.Aust. Math. Soc. Vol.3(1963), pp375-83
- (140)Moran,P.A.P. "Introduction to probability theory"
Oxford Univ. Press(1968)
- (141)Crew,G.L. "The CCITT System No.6 - A common channel signalling scheme"
Telecom. J. of Aust. Vol.18(1968);3,pp251-6

- (142) Rubas, J. "Supervision of Plant Congestion"
Carroll, B.J. Telecom. J. of Aust. Vol.18(1968);1, pp5458
- (143) Edstrom, N.H. "Stored Program Techniques and their impact
on trends in Telecommunications"
- (144) Horsfield, B.R. "Fast signalling in the U.K. telephone
network"
Proc. IEE Vol.117(1970);3, pp521-31
- (145) Welch, S. "The signalling problems associated with
Register-controlled automatic telephone
exchanges"
Proc. IEE Vol.109B; pp465-80
- (146) Welch, S. "Telephone switching and network control"
Paper 24b, ITU Seminar, 'The Telephone
Service', London, Oct. 1966
- (147) Punter, H. "Some economic and financial aspects of
Christiaanse, automatisisation"
C.H. HET PTT-BEDRIJF, Vol.11(1962);3/4, pp185-98
- (148) Wallstrom, B. "A distribution model for telephone traffic
with varying call intensity, including
overflow traffic"
Ericsson Technics, Vol.20(1964);2, pp183-202
- (149) Moran, P.A.P. "Theory of Storage"
Methuen Monographs, London(1959)
- (150) Jacobaeus, C. "Applications of telephone traffic theory
in industry and administration"
Ericsson Review, Vol.48(1971);1; pp2-13
- (151) Asgersen, Chr. "A new Danish traffic routing plan with
single stage alternate routing and
consistent use of service protection finals
for first routed traffic to tandem offices"
Teleteknik Vol.11(1967);1, p11-
- (152) Howard, R.A. "Dynamic programming and Markov processes"
Wiley, London(1960)
- (153) Fry, T.C. "Probability and its engineering uses"
Van Nostrand, 1965