1

National Library
of Canada

Library of Parliament
Ottawa

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

National Library
of Canada

Bibliothèque nationale
du Canada

CANADIAN THESES
ON MICROFICHE

THÈSES CANADIENNES
SUR MICROFICHE

0-315-22189-5

NAME OF AUTHOR/NOM DE L'AUTEUR MARVIN KRYM
TITLE OF THESIS/TITRE DE LA THÈSE PERFORMANCE ANALYSIS OF A LOCAL AREA NETWORK
WITH APPLICATION TO FIBRE OPTIC
INTEGRATED VOICE/DATA
UNIVERSITY/UNIVERSITÉ CARLETON UNIVERSITY
DEGREE FOR WHICH THESIS WAS PRESENTED/
GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE MASTER OF ENGINEERING
YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE GRADE 1984
NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE PROFESSOR A. HAFEZ

Permission is hereby granted to the NATIONAL LIBRARY OF
CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and neither the
thesis nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈ-
QUE NATIONALE DU CANADA de microfilmer cette thèse et
de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la
thèse ni de longs extraits de celle-ci ne doivent être imprimés
ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATE/DATÉ

Aug 15/84

SIGNED/SIGNÉ

Marvin Krym

PERMANENT ADDRESS/RÉSIDENTIE FIXE

1108 - 1316 CARLING AVE

OTTAWA ONT

K1Z 7L1

CANADA

PERFORMANCE ANALYSIS OF A LOCAL AREA NETWORK
WITH APPLICATION TO
FIBRE OPTIC INTEGRATED VOICE/DATA

by

© Marvin Krym

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements
for the degree of
Master of Engineering

Department of Systems and Computer Engineering
Faculty of Engineering
Carleton University
Ottawa, Ontario
March 1984

LICENCE TO CARLETON UNIVERSITY

In the interests of facilitating research by others at this institution and elsewhere, I hereby grant a licence to:

CARLETON UNIVERSITY

to make copies of my thesis:

PERFORMANCE ANALYSIS OF A LOCAL AREA NETWORK
WITH APPLICATION TO FIBRE OPTIC INTEGRATED VOICE/DATA

or substantial parts thereof, the copyright which is invested in me, provided that the licence is subject to the following conditions:

Only singled copies shall be made or authorized to be made at any one time, and only in response to a written request from the library of any university or similar institution on its own behalf or on behalf of one of its users.

This licence shall continue for the full term of the copyright, or for so long as may be legally permitted.

The Universal Copyright Notice shall appear on the title page of all copies made under the authority of this licence.

This licence does not permit the sale of authorized copies at a profit, but does permit the collection by the institution or institutions concerned of charges covering actual costs.

All copies under the authority of this licence shall bear a statement to the effect that the copy in question "is being made available in this form by the authority of the copyright owner solely for the purpose of private study and research and may not be copied or reproduced except as permitted by the copyright laws without written authority from the copyright owner."

The foregoing shall in no way preclude the granting by the author of a licence to the National Library of Canada to reproduce the thesis and to lend or to sell copies of the same. For this purpose it shall also be permissible for Carleton University to submit the abovementioned thesis to the National Library of Canada.

Mandy Steinberg
Signature of Witness

Aug 15/84 M. ENG
Date Degree

[Signature]
Signature of Student

SYSTEMS ENGINEERING
Department of Student

The undersigned recommend to the Faculty of Graduate Studies and
Research the acceptance of the thesis

**PERFORMANCE ANALYSIS OF A LOCAL AREA NETWORK
WITH APPLICATION TO
FIBRE OPTIC INTEGRATED VOICE/DATA**

by Marvin Krym, submitted in partial fulfillment of the requirements
of the degree of Master of Engineering.

Thesis Supervisor

Chairman,
Department of Systems and
Computer Engineering

Carleton University
August 1984

ABSTRACT

An integrated voice/data fibre optic local area network is investigated. A single frame, subdivided into one data field and a number of voice channels wraps around a loop topology. Communicating devices are connected to the loop by way of a loop interface. The frame rotates around the loop in a fixed interval of time servicing the stations. The data transmission protocol is a synchronous token passing network. Messages queuing up at the stations are serviced nonexhaustively and in a FIFO order. The mean scan time moments, mean message access delay and mean queue lengths are determined. For heterogeneous traffic, stability regions are shown to exist where combinations of stations are stable and unstable simultaneously. It is also shown that a homogeneous network is more stable than a heterogeneous network carrying equal total traffic. Furthermore, a homogeneous network is capable of carrying larger total traffic than its heterogeneous counterpart.

The nonexhaustively serviced network is then compared to the exhaustively serviced network. The results are that exhaustively serviced networks have lower access delays but are less stable than the nonexhaustively serviced network. This led to the mixed service network which has high traffic stations exhaustively serviced. Queuing results for the mixed service case are determined. It is shown to be a compromise between good stability and low access delay.

Some of the results were verified by simulations. The simulations were written in GPSS.

ACKNOWLEDGEMENTS

I wish to acknowledge, gratefully those who have assisted me in the research and writing of this thesis.

Firstly, I wish to express my appreciation to Professor H. Hafez for his continuous guidance and assistance throughout this work.

I also wish to thank Dr. Jan Conradi and Dr. Romain Maciejko both of Bell-Northern Research. Dr. Conradi permitted me the use of the facilities at Bell-Northern Research. Dr. Maciejko suggested the topic and provided valuable advice and criticism.

Finally, as both etiquette and my heart compel me to do, I must thank my wife, Naomi, for her encouragement and support throughout this work.

TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER 1 INTRODUCTION	1
1.1 Introduction	1
1.2 Thesis Objectives	8
1.3 Thesis Outline	9
CHAPTER 2 NETWORK MODEL	11
2.1 Introduction	11
2.2 Transmission Protocol	12
2.2.1 Sync Field	12
2.2.2 Token Field	15
2.2.3 Data Field	16
2.2.4 Voice Channels Field	16
2.2.5 Gap Field	18
2.3 Hardware Considerations	18
2.3.1 Loop Interface	18
2.3.2 Host Devices	21
2.4 Lan System Description	22
2.5 System Assumptions	27

	<u>Page</u>
CHAPTER 3 PERFORMANCE ANALYSIS	29
3.1 Introduction	29
3.2 Homogeneous Traffic Network	30
3.2.1 Mean Scan Time	30
3.2.2 Mean Square Scan Time	34
3.2.3 Mean Access Delay	40
3.2.4 Mean Queue Length	40
3.3 Heterogeneous Traffic Network	45
3.3.1 Mean Scan Time	45
3.3.2 Mean Square Scan Time	45
3.3.3 Stability Conditions	47
3.3.4 Mean Access Delay	52a
3.3.5 Mean Queue Length	53
3.4 Heterogeneous Traffic with Batch Arrivals	53
3.4.1 Mean Scan Time	54
3.4.2 Mean Square Scan Time	55
3.4.3 Mean Access Delay	55
3.4.4 Mean Queue Length	56
3.5 Comparison of Nonexhaustive to Exhaustive Service	56
3.5.1 Statistics for Exhaustive Service	58
3.5.2 Stability Comparison	59
3.5.3 Access Delay Comparison	61
3.6 Mixed Service Network	64
3.6.1 Mean Scan Time	65
3.6.2 Mean Square Scan Time	66
3.6.3 Mean Access Delay	68
3.7 Voice Analysis	70

	<u>Page</u>
CHAPTER 4 SIMULATION	72
4.1 Introduction	72
4.2 Model Description	72
4.3 Simulation Results	77
4.3.1 Homogeneous Network	77
4.3.2 Heterogeneous Network	77
4.3.3 Heterogeneous Network with Batch Arrivals	84
CHAPTER 5 SUMMARY AND CONCLUSIONS	93
5.1 Introduction	93
5.2 Mean Scan Time	94
5.3 Mean Square Scan Time	95
5.4 Stability Conditions	97
5.5 Mean Access Delay	98
5.6 Comparison of Nonexhaustive to Exhaustive Service	100
5.7 Mixed Service Network	101
5.8 Simulation	102
5.9 Conclusions	103
5.10 Scope for Further Work	103
REFERENCES	105
APPENDIX I	107

LIST OF TABLES

	<u>Page</u>
4.1 Homogeneous Network Results for a 5 Node Network	78
4.2 Heterogeneous Network Results. Theoretical Results for a 5 Node Network	85
4.3 Heterogeneous Network Results. Simulation Results for a 5 Node Network	86
4.4 Heterogeneous Traffic Results with Batch Arrivals	90

LIST OF FIGURES

	<u>Page</u>
1.1 Examples of Topology	5
2.1 Configuration of the Token Pass Loop and its Stations	13
2.2 Frame Format	14
2.3 Functional Block Diagram of the Loop Interface	19
2.4 Example of a Frame	24
2.5a A 4 Station Loop	25
2.5b Timing Diagram for the 4 Station Loop	25
3.1 Probability a Station Has a Message to Transmit when a Free Token Arrives versus Traffic for a Homogeneous Network with N Stations	35
3.2 Mean Scan Time versus Traffic for a Homogeneous Network with N Stations	36
3.3 Mean Square Scan Time versus Traffic for a Homogeneous Network with N Stations	39
3.4 Mean Access Delay versus Traffic for a Homogeneous Network with N Stations	43
3.5 Queue Length versus Traffic for a Homogeneous Network with N Stations	44
3.6 Plot of Station Stability Regions	49
3.7 Mean Scan Time versus Traffic for an Exhaustively and Nonexhaustively Serviced Homogeneous Network	60
3.8 Effects of a High Traffic Station on a Nonexhaustively Serviced Network	62
3.9 Mean Access Delay for Exhaustively and Nonexhaustively Serviced Networks with N Stations	63
3.10 Mean Access Delay versus Traffic for a Mixed Service Network with 50 Stations	69

	<u>Page</u>
4.1 Simulation Model Flowchart	73
4.2a p_i versus λ for a 5 Station Homogeneous Network	79
4.2b $\bar{t}_c$ versus λ for a 5 Station Homogeneous Network	80
4.2c $\bar{t}_c^2$ versus λ for a 5 Station Homogeneous Network	81
4.2d $\bar{d}_a$ versus λ for a 5 Station Homogeneous Network	82
4.2e T versus λ for a 5 Station Homogeneous Network	83
4.3a $\bar{t}_c$ versus λ_5 for a 5 Station Heterogeneous Network	87
4.3b $\bar{t}_c^2$ versus λ_5 for a 5 Station Heterogeneous Network	88
4.3c $\bar{d}_a$ versus λ_5 for a 5 Station Heterogeneous Network	89
4.4a $\bar{t}_c$ versus λ_{B5} for a 5 Station Heterogeneous Network with Batch Arrivals	91
4.4b $\bar{t}_c^2$ versus λ_{B5} for a 5 Station Heterogeneous Network with Batch Arrivals	92

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

In recent years there has been an evolution in computing philosophy from the large centralized computer to smaller computers, interconnected to permit communication and distributed over a geographical area. If each computer is an autonomous unit then the system is called a computer network ([1], page 2). Furthermore, if all the computers are in a close geographical proximity so that no two computers are more than a few kilometers apart, then the system is known as a local area network.

The development of low-cost very large-scale integrated circuit computers encouraged the distribution of computing power. A manufacturing plant, for example, using computerized process control could have a small specialized computer for each process. Each computer would be located near the process under control. All the computers would have a communication link between them to a master computer responsible for the control, scheduling and maintenance of each process control computer. This network of process control computers is a local area network and has clear advantages over a single centralized computer. Firstly, small dedicated computers in the local area network would collectively require simpler software. A single computer would require a time sharing operating system in

order to distribute the computing power, in addition to the process control software. Secondly, the local area network can grow incrementally resulting in lower startup costs. The computing power can grow with the demand. Use of a single computer on the other hand may mean projections of the total computing power required in order to obtain a system large enough to meet future needs. If all this computing power is not required immediately, then this under utilization of resources may mean wasted dollars. Thirdly, there is the issue of reliability. In the local area network configuration, all computers are autonomous units and are capable of independent operation. Even if the master computer should fail, the remaining computers in the network may continue performing a subset of the process control functions. Failure of the computer in the single computer case would result in a complete failure of process control computing power.

Applications of computer networks are not restricted to manufacturing and process control. Another application involves the interactive user. Small computer based data concentrators may be distributed over an area in order to service computer terminals. Examples are the airline reservation systems, teleconferencing and most recently interactive computing by way of cable TV. The principal concept is to bring the computing power to the user rather than bringing the user to the computer.

There are other technical and cost reasons which make computer networks useful. Traditionally remote terminals would be connected to a computer by a single line and a modem. The cost for configuring

a large number of terminals in this manner would be very large. Furthermore, computer traffic and especially terminal traffic is characterized by transmission bursts with comparatively long gaps between the bursts. If the burst length is very short in time compared to the gap length then the transmission facility will be underutilized. This represents a waste of both transmission resource and money. Computers in the network could be connected to a single transmission medium (i.e., coaxial cable) in order to share the bandwidth. Transmission protocols have been developed to permit computers access to the shared transmission medium on a demand basis and in such a way as to obtain a fair sharing of the transmission medium.

The discussion to this point has indicated the viability of computer networks and more specifically local area networks on which this thesis is focused. Many types of local area networks have been developed and may be categorized by consideration of certain aspects of their design.

1. Topological Design and Transmission Protocol:

The topology refers to the interconnection arrangement of the computers while the transmission protocol is the method by which a computer seizes the transmission facility to transmit its data. The topology and transmission protocol are related in the sense that selecting the topology constrains the protocol design. Some examples of these networks will illustrate these concepts. Four general topologies with suitable transmission protocols are described below.

i) STAR TOPOLOGY - The star topology shown in Fig. 1.1a requires all computers to send their messages to a central computer where the messages are relayed to their destination. Each peripheral computer has dedicated access to a transmission medium. If the central computer has a dedicated port for each input and is capable of handling multiple simultaneous inputs then the peripheral computers may transmit messages when required. If the central computer is only capable of handling single messages at a time, then a peripheral computer will only transmit after being polled for messages by the central computer.

ii) MESH TOPOLOGY - The mesh topology network shown in Fig. 1.1b commonly uses the store and forward protocol. In this method a message travelling from a source computer to its destination computer may be required to pass through intermediate computer stations. At each intermediate station, the message is stored, queued and then retransmitted onto the appropriate link to bring the message one hop closer to its destination. This method usually finds application in long distance networks and is generally not used for local area networks.

iii) BUS TOPOLOGY - The bus topology of Fig. 1.1c often uses contention or polling as the transmission protocol. Using the contention method, a computer transmits its message onto the bus in a broadcast fashion (i.e., capable of being received by all computers

a. Star Topology

b. Mesh Topology

Fig. 1.1 Examples of Topologies

c. Bus Topology

d. Loop Topology

connected to the bus). If another computer transmits simultaneously, a collision results destroying both messages. In this case both computers must wait a random period of time before reattempting transmission. The polling method requires one computer on the bus to be a master which queries each computer for messages. A positive acknowledgement causes the master computer to assign the transmission facility to the queried computer.

iv) LOOP TOPOLOGY - Many transmission protocols have been developed for the loop topology shown in Fig. 1.1d. Many of the protocols are described in a paper by Penny [2]. One very interesting method is the token passing method and has been used by IBM [3]. In this method a permission permit called a token is transmitted around the loop from computer to computer. Any computer receiving the token has exclusive access to the loop since only one token can exist in the network at any one time. After a computer has completed using the loop for transmission, the token is transmitted to the neighbouring computer.

2. Transmission Medium:

All the methods described above are for use with physical transmission mediums such as copper wire, coaxial cable or fibre optic cable. Radio is another transmission medium in use for computer networks. Due to the physical separation of the computers contention protocols have become popular with this medium.

Fibre optic cable has made a very recent debut for use in local area networks. Interest in it lies in the fact that fibre optic cable offers some very clear advantages over copper wire. These advantages are a very large bandwidth distance product (>400 MHz·km) [6], immunity to electromagnetic interference, no RF leakage and no inductive crosstalk [4]. Most importantly fibre optics is becoming cost competitive with other transmission media. Furthermore, the large bandwidth can be exploited to permit voice traffic simultaneously with the data traffic. The large bandwidth allows higher traffic rates and therefore a larger number of computers may also be connected to the network.

The computer network has clearly established itself in the telecommunications industry. They are a reality and they work. The study of their precise performance is receiving a large proportion of research attention. Many applications of computer networks have resulted in many types of networks some which have already been described and each with its particular performance.

1.2 THESIS OBJECTIVES

Presently, Bell-Northern Research (BNR) is engaged in the study and development of an integrated voice/data fibre optic local area network which is also described in a paper by Maciejko [13]. The objective of this thesis is to develop an analysis tool which may be used for the BNR local area network. A general form of the

network will be described which will form the basis of a mathematical model also developed in this thesis. The mathematical model should permit a reasonably accurate assessment of the network even before it is built.

The loop will be analyzed under various input traffic assumptions in order to derive mathematical expressions for useful design parameters such as the mean access delay. The traffic bounds to assure network stability will be determined as well as resulting performance bounds. The emphasis will be on the performance of the data aspects of the loop.

1.3 THESIS OUTLINE

In chapter 2 the access or transmission protocol is described. It begins with a description of how data is put on the loop and transmitted to the destination computer. This is followed by a block diagram level description of the hardware used to interface to the loop. The traffic assumptions are then detailed. The chapter concludes with an example to help tie all the concepts together.

In chapter 3 the mathematical model of the network is developed. The input traffic is assumed Poisson and both single arrivals and batch arrivals are considered. The analysis is made for nonexhaustive, FIFO service of the computer message queues. The results for exhaustive service are quoted from the literature and a comparison is made to the nonexhaustive case. A mixed service case

is presented and analyzed in which high traffic computer stations are serviced exhaustively while the remaining computer stations are serviced nonexhaustively. This case is analyzed and shown to be a compromise between low delay and good stability.

In chapter 4, a simulation model for the network is described. The results of a simulation using GPSS are presented to support and verify the results of the previous chapter.

In chapter 5, the thesis results are summarized and concluded.

CHAPTER 2

THE NETWORK MODEL

2.1 INTRODUCTION

The communications network modelled in this thesis is a fibre optic based local area network designed to support simultaneous voice and data traffic. Loop interfaces are connected to a fibre optic cable configured in a loop topology. Communication devices (herein called the host) such as computers and disk drives are connected to the loop interface. The combination of the host and its loop interface will be called a station.

Messages are circulated around the loop in one direction. Interstation communication may be voice, data or both voice and data simultaneously. The loop requires one special station called a supervisor which monitors the loop in order to ascertain and maintain sanity among the stations by inspecting the data passing through it.

This chapter begins by describing the loop protocol, that is how information is passed around the loop. This is followed by a system level hardware description of the station connected to the loop and by a discussion of the traffic assumptions. In the next section all the concepts are brought together to provide a system description of the loop network by way of an example. The chapter concludes with the assumptions that are necessary in order to proceed with the mathematical treatment in the following chapter.

2.2 THE TRANSMISSION PROTOCOL

It is useful to consider the loop and its stations as a single large circular shift register as shown in Fig. 2.1. Each loop interface forms a section of the shift register used to transmit and receive information.

At power up, the supervisor transmits the frame shown in Fig. 2.2 onto the loop. The frame then continues to circulate around the loop indefinitely while the supervisor maintains its correct structure. The frame wraps completely around the loop and shifts around the loop once in a fixed interval of time called the frame time denoted by t_f . For example, if a station is reading the beginning of the SYNC field, then t_f seconds later it must again be reading the beginning of the SYNC field. Due to this property, the loop is called synchronous.

2.2.1 SYNC Field

The sync field is a unique bit pattern or code violation such as a bipolar violation used by the loop interface to maintain frame and clock synchronization. Frame synchronization is necessary for the loop interface to determine the field positions and to determine when a particular field is passing through the interface. Clock synchronization is necessary to ensure that each interface clocks data through its portion of the shift register at the same rate as

Fig 2.1 Configuration of the token pass loop and its stations.

Fig 2.2 Format of the frame wrapped around the loop.

the remainder of the network.

2.2.2 Token Field

The token is a one bit field which takes on one of two states, busy or free. A busy token indicates that the data field contains valid data while a free token indicates that the data field is vacant. Hence the token and data field form an embedded token pass network whose protocol is here described.

A station numbered i , $1 \leq i < N$ in an N station loop receives a free token. Assuming it has a message to transmit, the loop interface changes the token to busy and transmits its message in the data field area after making the token busy. The data field has a header which contains at least the TO and FROM addresses as well as a trailing checkword. The message in the data field travels around the loop and since the token has been marked busy, none of the stations are permitted to alter the token or deposit their message in the data field. All of the stations on the loop will compare the TO address with their own, and will copy the data if an address match occurs. The token will return to the transmitting station t_f seconds later at which time the station has two options. It can set the token free, allowing station $(i+1) \text{ MOD } N$ to use the token or if station i has further messages to transmit, it leaves the token set busy and deposits its next queued message into the data field. If this later mechanism continues until all the buffered messages have been

transmitted before the station sets the token to free, then the mechanism is called exhaustive service. The former mechanism of setting the token to free is called nonexhaustive service. Any station using the token is said to be holding the token.

2.2.3 Data Field

The data field is used by the station to exchange messages. The field will have a header containing at least the TO and FROM address and possibly a trailer containing frame check information.

Access to the data field is obtained only after a station sees a free token and sets the token to busy. The station then deposits its message in the data field. The frame circulates around the loop to allow the destination to copy the data field. Once the frame circulates around the loop, the token field returns to the station which sent the message. This station may leave the token busy in order to exhaustively service its queue or it may set the token field to free thereby permitting its neighbouring station to use the token and data field.

2.2.4 Voice Channels Field

The voice channels field is composed of a fixed number of contiguous channels. Each channel carries an 8 bit PCM codec sample and possibly 2 extra bits for signalling if 10 bit channels are used.

The telephone industry standard has a codec sampling rate of 8 KHz and therefore the channel must be available every 125 μ sec. This forces the frame time t_f to correspond to this period if such a codec is used.

When a station requires a voice channel to set up a voice connection, it must request a voice channel from the supervisor. This may be done by signalling the request via the data field. Once the voice channel has been assigned to a station, the station will continue to deposit its voice sample into its assigned channel every t_f seconds until the station signals the supervisor that it is relinquishing the channel. There may be in the order of a hundred such voice channels in a frame which are assigned to stations on a demand basis to permit these stations digital voice communication.

The use of a voice channel is completely independent of the use of the data field. In other words, the token and data field form an independent token passing network while the voice channels form an independent digital voice transmission system. Access to the data field is obtained by obtaining a free token while independent access to a free voice channel is obtained by making a request to the supervisor. Since the voice and data transmissions are independent, they can and will be studied independently.

2.2.5 Gap Field

The gap field carries no information and is used as a waiting time for interstation fibre propagation delay. The sum of all fibre propagation delays around the loop must not exceed the gap length. Failure to meet this condition will result in a frame requiring longer than t_f to circulate the loop and therefore causing a loss of synchronization.

2.3 HARDWARE CONSIDERATIONS

Each station consists of two components, the host and the loop interface. The host is a device which generates traffic such as a terminal or computer. Host units are categorized as to the type of traffic they generate. The loop interface is an interface between the host and the fibre transmission line.

2.3.1 Loop Interface

The loop interface is a hardware device as shown in block diagram form in Fig. 2.3. The optical signal upon entering the interface is converted to an electrical signal by the O/E (optical to electrical) convertor. With SWA in position 'a' the signal passes through a fixed length shift register, and an E/O convertor before being transmitted back onto the fibre line. This mode of operation

Fig 2.3 Functional block diagram of the loop interface.

is called the repeater state. In this state SWB remains open and only an address match or arrival of a free token causes a change of state.

The controller circuits are responsible for monitoring the incoming data checking for a free token or the TO address for a match. If the controller detects an address match and the token is busy, then the message is copied into a local receive buffer from where it will later be passed onto the host. Copying is done by closing SWB which puts the loop interface into the receiver state.

If the controller circuit detects a free token and a message is queued for transmission then SWA is switched to position 'b' at the correct field boundary in order to transmit a busy token and the message in its correct field. This state is called the transmit state and concludes with SWA being restored to position 'a' at the conclusion of transmitting the data field. This returns the loop interface to the repeater state. When the token field arrives back, it is either set to free or reused.

During power down, power fail, or loop interface failure, it may be necessary to bypass this station. These issues deal with network reliability and while being important, lie outside the scope of this thesis.

2.3.2 Host Devices

Host devices will be categorized into two types depending on the type of traffic they generate. The first type of host considered generates short messages. Short means that the message length does not exceed the data field length, thereby permitting the entire message to be transmitted in a single frame. The traffic is assumed to be generated according to a Poisson process ([5], page 26) given by

$$P[M=k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad (2.1)$$

where $P[M=k]$ is the probability k messages are generated in t seconds and λ is the average number of messages generated per second. Since the data field length is fixed, any messages shorter than this length will be padded.

The second type of host considered is one which generates messages with an average length which is larger than the data field length. Therefore messages are broken into smaller units called blocks so that each block fits into the data field. Messages are then considered to be generated as a batch of blocks and each block is treated as a separate message. The number of blocks in a batch (which constitutes a single message) is called the batch size and is in general a random variable with a mean value of $\bar{g}$ and second moment $\bar{g}^2$.

Messages are generated according to a Poisson process with

rate λ_B messages (or batches) per second but blocks are generated according to a Compound Poisson process given by ([5], page 256)

$$P[M=k] = \sum_{i=0}^k e^{-\lambda_B t} \frac{(\lambda_B t)^i}{i!} C_g(i) \quad (2.2)$$

where $P[M=k]$ is the probability that k blocks are generated in an interval of t seconds, $C_g^{(i)}$ is the i -fold convolution of C_g with itself and C_g is the distribution function of the batch size.

In the remainder of the thesis, Poisson will always refer to the first type of message while compound Poisson will always refer to the second type of message.

2.4 LAN SYSTEM DESCRIPTION

In order to obtain a clearer understanding of the network a practical network will be discussed.¹ Consider a network that will consist of a maximum of 200 stations. The frame time t_f will be 125 μ sec in order to accommodate voice codecs. Considering that the fibre optic cable has a propagation delay of about 5 ns/m [6], and if the loop is restricted to one kilometer, then the total propagation delay will be 5 μ sec and hence the frame gap will be 5 μ sec long. This network will have 200 voice channels capable of supporting 100 simultaneous conversations, a sync and token field which together

¹ This is an example only in order to explain the concepts and does not necessarily represent the precise LAN in [13].

total 10 bytes and a data field 990 bytes long. If the voice channels are 8 bits long, then the frame will be 9600 bits long and will be 120 μ sec in time. This results in a bit rate of 80 Mbps. The resulting frame is shown in Fig. 2.4. The high bit rate required to realize the network makes fibre optic cable a good choice.

The gap is distributed along the fibre cable. The remaining 9600 bits must be distributed among the 200 stations which is accomplished by making the shift register within each loop interface 48 bits wide. Now, the frame fits into a maximum size loop but the network will not always be of maximum size. Hence, it is the job of the supervisor station to increase its internal delay to make up for stations and propagation delay short of the maximum network.

Consider a loop with only four stations labeled S1 to S4 with S1 also acting as the supervisor station as shown in Fig. 2.5a. If the loop is made, say 400 meters long, the total propagation delay around the loop will be 2 μ sec. Assume that all stations have a message to transmit and use nonexhaustive service.

The operation of the network will be considered as a sequence of events. Begin the sequence with S1 beginning the frame transmission and holding the token shown at point 'a' on the timing diagram of Fig. 2.5b. Because of the smaller than maximum loop size, the head of the frame circles the loop and arrives back to S1 at point 'b'. The supervisor must store the frame as it arrives to prevent the frame from wrapping around itself on the loop. This is accomplished by delaying transmission of the frame until 125 μ sec of time

Fig 2.4 Example of a frame.

Fig 2.5a A 4 station loop.

Fig 2.5b Timing diagram for the 4 station loop.
 The frame times with hashed lines is the time used by the respective station to transmit its message.

has elapsed from point 'a' and is shown at point 'c'. The frame requires the interval ac but the loop is only ab in length. Therefore $S1$ delays the frame by bc which is the time required to shift the frame through 196 stations and three μ seconds of propagation delay. The token has returned to $S1$ where it is set to free, and shifted through $S1$ into $S2$ arriving at point 'd'. The time interval cd is the time representing the propagation delay between $S1$ and $S2$ plus the time to shift the token out of $S1$ and into $S2$ which is 48 bits and is called the walk time.

Station $S2$ transmits the frame with its message at point 'd' and completes transmitting the frame at point 'g'. The area shown by the single hashed lines indicates where the station is transmitting its message. After $S2$ begins transmission at 'd', $S1$ receives the frame at 'e'. Again $S1$ must store and delay the frame until 125 μ sec after 'c' shown as point 'f'. At f , $S1$ transmits the frame so that it arrives back to $S2$ at point 'g' giving the appearance to $S2$ that the loop delay is 125 μ sec and that the frame wraps completely around it. After $S2$ sets the token to free, it is passed to $S3$. Then $S3$ uses the token and passes it to $S4$. Station $S4$ has completed using the token and is ready to pass it back to $S1$ at 'h'. Station $S1$ receives the frame and it becomes available for transmission at point 'i' which is not a 125 μ sec boundary. Therefore the frame is delayed until 'j' where the cycle repeats.

The time k_i represents the sum of the walk times around the loop and is equal to ab or ce . The supervisor delays the transmis-

sion until 'j' thereby increasing the total walk time to 125 μ sec. In general, the sum of the walk times is always made equal to t_f in order to preserve synchronism.

Another important characteristic of the loop is the scan time defined as the time interval beginning when a station receives a free token until it receives a free token again during which time all the stations have used the token if it was required for transmission of a message. A station will receive a token, either free or busy, every t_f seconds but will only receive a free token every scan time interval. The scan time denoted by t_c is shown in Fig. 2.5b by the interval a_j and in general is given by

$$t_c = t_f + mt_f \quad (2.3)$$

where the first term t_f represents the total walk time and m is the number of stations which have used the token during a given scan interval.

2.5 SYSTEM ASSUMPTIONS

In order to achieve any tractable mathematical results, certain assumptions must be made. The major assumptions required are outlined below.

A station receiving a free token will have a message to transmit with probability p independently of any other station. This assumption is justified for loop networks by Leibowitz [7].

The transmit buffers in the loop interface will have an infinite capacity. A realized interface with finite buffers will behave very closely to the infinite buffer case if the probability of buffer overflow is very small ([8] page 648).

All traffic generated at each station will be either Poisson or Compound Poisson as described in section 2.3.2.

All mathematical results apply only to the network during steady state and will be mean value results.

CHAPTER 3

PERFORMANCE ANALYSIS

3.1 INTRODUCTION

In this chapter a mathematical model of the network is developed. Consideration is first given to the most simple case. This is a homogeneous¹ (also called balanced) network with Poisson generated traffic at each host and FIFO nonexhaustive queue service. The mean scan time moments, mean queue lengths and mean access delays are determined for this network. The same statistics are then determined for the network with heterogeneous² (also called unbalanced) Poisson traffic. A stability criterion is developed showing the interdependence of the message generation rates and queue stability. The network is further generalized by modelling it for heterogeneous Compound Poisson traffic.

All the derivations have been for nonexhaustive service. The next section in this chapter compares the derived nonexhaustive service queuing results to published exhaustive service queuing results. The results show that a delay stability tradeoff exists between the two service schemes and leads to the mixed service network.

1 Homogeneous traffic means that the same traffic rate is generated at all stations.

2 Heterogeneous traffic means that different traffic rates are generated at different stations.

In the mixed service network, high traffic rate stations are exhaustively serviced while low traffic rate stations are nonexhaustively serviced. The result is a compromise in delay and stability. The mixed service network is analyzed and queuing results are determined for homogeneous Poisson traffic.

The chapter concludes with performance criterion for the voice traffic.

3.2 HOMOGENEOUS TRAFFIC NETWORK

The network model developed in this section assumes homogeneous Poisson traffic with nonexhaustive service. The results obtained, then form the starting point for the more general cases.

3.2.1 Mean Scan Time

Consider a single station i in an N station network. If at least one message is queued in this station waiting for transmission, then when a free token arrives, it will be used. The token will be set busy, and t_f seconds later after circling the loop the token will be set free and sent to the neighbouring station. The process is repeated by the neighbouring station and its neighbouring station and so on until a free token again arrives to station i . The time interval between arrivals of the free token is a discrete random variable called the scan time. The first and second moments will be

determined using a method outlined by Yeun [9] as the starting point. Also, let l_i be the number of messages queued at station i waiting for transmission.

Begin by considering an event A_i defined by

$$A_i \ (i = 1 \dots N) = \begin{cases} 1, & \text{for } l_i \geq 1 \\ 0, & \text{for } l_i = 0 \end{cases}$$

The scan time is then given by (c.f. (2.3))

$$t_c = t_f + \sum_{i=1}^N A_i t_f \quad (3.1)$$

and taking the expectation yields

$$\overline{t_c} = t_f + \sum_{i=1}^N \overline{A_i} t_f$$

The value of $\overline{A_i}$ is determined by

$$\begin{aligned} \overline{A_i} &= 1 \cdot P(l_i \geq 1) + 0 \cdot P(l_i = 0) \\ &= p_i \end{aligned}$$

and making the substitution yields

$$\overline{t_c} = t_f + \sum_{i=1}^N p_i t_f \quad (3.2)$$

The variable p_i is the probability that station i has at least one message to transmit and is determined by appeal to M/G/1 queuing theory. Within a scan period, the transmission queue evolves according to

$$l_{ik} = l_{i(k-1)} - U_{(k-1)} + a_{ik} \quad (3.3)$$

where l_{ik} is the transmission queue length at the end of the k th scan, a_{ik} is the number of new message arrivals during the k th scan and

$$U_{(k-1)} = \begin{cases} 1 & \text{if } l_{i(k-1)} \geq 1 \\ 0 & \text{if } l_{i(k-1)} = 0 \end{cases}$$

In words (3.3) says that the queue length at the end of the k th scan is equal to the queue length at the end of the $(k-1)$ th scan less one message which is transmitted during the k scan (i.e. $U_{(k-1)}=1$) plus the number of arrivals during the k th scan. If the $(k-1)$ th scan left the queue empty, then no transmission could take place during the k th scan (i.e. $U_{(k-1)}=0$). Taking the expectation of (3.3) yields

$$\overline{l_{ik}} = \overline{l_{i(k-1)}} - \overline{U_{(k-1)}} + \overline{a_{ik}} \quad (3.4)$$

But at steady state $\overline{l_{ik}} = \overline{l_{i(k-1)}}$ (assuming ergodicity) and

$$\overline{U_{(k-1)}} = 1 \cdot P(l_{i(k-1)} \geq 1) + 0 \cdot P(l_{i(k-1)} = 0) = p_i$$

yielding the desired result

$$p_i = \overline{a_{ik}} \quad (3.5)$$

The variable $\overline{a_{ik}}$ is the expected number of arrivals during a scan time t_c which is a random variable. Since the arrival rate is Poisson with rate λ_i , the expected number of arrivals during a scan conditioned on a given scan time is given by

$$E(a_{ik}|t_c) = \lambda_i t_c$$

The condition is removed by multiplying by the scan time pdf (probability density function) and summing over all scan times yielding $\overline{a_{ik}} = \lambda_i \overline{t_c}$, and when substituted into (3.5) gives the final result

$$p_i = \lambda_i \overline{t_c} \quad (3.6)$$

Substitution of (3.6) into (3.2) yields the scan time for the heterogeneous case given by

$$\overline{t_c} = \frac{t_f}{N - t_f \sum_{i=1} \lambda_i} \quad (3.7)$$

From (3.7) the scan time for the homogeneous case is obtained but further consideration of (3.6) is first required. Since (3.6) describes a probability, a boundary condition results given by

$$\frac{1}{\lambda_i} > \overline{t_c} \quad (3.8)$$

which shows that the mean time between message arrivals must be larger than the mean scan time. In other words, on the average, messages should not arrive more frequently than free tokens. As will be shown, failure to meet this condition results in an unstable queue. A queue is defined as being stable if it has a finite steady state mean queue length.

The mean scan time for the homogeneous network is obtained from (3.7)

$$\bar{t}_c = \frac{t_f}{1 - t_f N \lambda}; \lambda_1 = \lambda_2 = \dots = \lambda_N = \lambda \quad (3.9)$$

subject to the restriction of (3.8).

Substituting (3.8) into (3.9) and solving for the scan time yields

$$\bar{t}_c < t_f + N t_f \quad (3.10a)$$

which shows that the scan time is bounded by an upper limit given by equality. This limit is the sum of the total walk time and the time to service every node in the network.

Substituting (3.8) into (3.9) and solving for the traffic rate gives the traffic rate boundary to achieve stability and is given by

$$\lambda < \frac{1}{t_f + N t_f} \quad (3.10b)$$

A plot of p_i and $\bar{t}_c$ versus the traffic rate λ is given in Fig. 3.1 and Fig. 3.2 respectively for various numbers of stations. In these and all subsequent plots, the value of t_f is set to 125 μ sec unless otherwise stated.

3.2.2 Mean Square Scan Time

The mean square scan time is found by squaring (3.1) and taking the expected value. Squaring (3.1) yields

Fig 3.1 Probability a station has a message to transmit when a free token arrives versus traffic for a homogeneous network with N stations.

Fig 3.2 Mean scan time versus traffic for a homogeneous network with N stations.

$$t_c^2 = \left(\sum_{i=1}^N A_i t_f \right)^2 + 2t_f \sum_{i=1}^N A_i t_f + t_f^2 \quad (3.11)$$

The first term is the square of a sum of random variables and can be written as

$$\left(\sum_{i=1}^N A_i t_f \right)^2 = t_f^2 \sum_{i=1}^N A_i^2 + t_f^2 \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N A_i A_j \quad (3.11a)$$

Substituting (3.11a) into (3.11) and taking the expected value yields

$$\overline{t_c^2} = t_f^2 \sum_{i=1}^N \overline{A_i^2} + t_f^2 \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \overline{A_i A_j} + 2t_f^2 \sum_{i=1}^N \overline{A_i} + t_f^2 \quad (3.12)$$

It has been shown that $\overline{A_i} = p_i$ and similarly it can be shown that $\overline{A_i^2} = p_i$ but the factor $\overline{A_i A_j}$ requires closer attention. Begin by noting that

$$\begin{aligned} \overline{A_i A_j} &= 1 \cdot 1 \cdot P(A_i=1, A_j=1) + 1 \cdot 0 \cdot P(A_i=1, A_j=0) \\ &\quad + 0 \cdot 1 \cdot P(A_i=0, A_j=1) + 0 \cdot 0 \cdot P(A_i=0, A_j=0) \end{aligned}$$

Using conditional probability this expression can be reduced to

$$\overline{A_i A_j} = P(A_i=1|A_j=1) \cdot P(A_j=1) \quad (3.13)$$

where the term $P(A_i=1|A_j=1)$ is the probability that an arbitrary station i has at least one message queued for transmission given that station j transmits a message during the same scan. This is an important quantity in the development of the model and will be denoted by P_{ij}' . From (3.6) $P_{ij}' = \lambda_i \overline{t_c'}$ where $\overline{t_c'}$ is the mean scan

time given that at least one station j transmits a message during the scan and for the homogeneous case is given by

$$\overline{t_c'} = 2t_f + (N-1)t_f p_{ij}' \quad (3.14)$$

Combining $p_{ij}' = \lambda_i \overline{t_c'}$, (3.14) and solving for p_{ij}' yields the desired result

$$p_{ij}' = \lambda_i \left(\frac{2t_f}{1 - (N-1)\lambda_i t_f} \right) \quad (3.15)$$

Since $P(A_j=1) = p_j$, (3.13) can be written as

$$\overline{A_i A_j} = p_{ij}' p_j \quad (3.16)$$

Making all the substitutions into (3.12), the mean square scan time can be given by

$$\overline{t_c^2} = t_f^2 \sum_{i=1}^N p_i + t_f^2 \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N p_{ij}' p_j + 2t_f^2 \sum_{i=1}^N p_i + t_f^2 \quad (3.17)$$

which for the homogeneous case reduces to

$$\overline{t_c^2} = t_f^2 (N(N-1)pp' + 3Np + 1) \quad (3.18)$$

where $p = \lambda \overline{t_c}$ and $p' = \lambda \left(\frac{2t_f}{1 - (N-1)\lambda t_f} \right)$

A plot of $\overline{t_c^2}$ versus for various values of N is given in

Fig. 3.3.

Fig 3.3 Mean square scan time versus traffic for a homogeneous network with N stations.

3.2.3 Mean Access Delay

The access delay is defined as the time interval beginning when a message arrives at a transmit queue until the message just begins transmission. The time interval is also known as the queuing delay. The mean access delay denoted by $\bar{d}_a$ is found by summing all the components of delay and taking their expected value.

There are two components of delay which must be considered and are the time required for a free token to arrive at the station and the time required to transmit all the messages which have arrived previous to the new arrival. After a new message arrives at a queue, the time taken for a free token to arrive is determined from renewal theory. This interval which is the first component of delay is called the residual scan time and is a random variable denoted by r . The value of the mean residual scan time is given by ([10], page 173)

$$\bar{r} = \frac{\overline{t_c^2}}{2t_c} \quad (3.19)$$

Now consider the second element of delay. After a message has begun transmission, the next message in the queue will not be transmitted until one scan interval later. Since at least one message is being transmitted during the scan, the scan time is given by the conditional scan time random variable t_c' . The queue therefore, appears

as an M/G/1 queue¹ with a service time t_c' whose mean is given by (3.14). The second element of delay is therefore the sum of all the scan times during which the messages are transmitted.

Summing all the delay components and given a newly arrived message finds l messages queued ahead of it, the access delay is given by

$$(d_a|l) = r + \sum_{k=1}^l t_{ck}'$$

where t_{ck}' is the length of the scan time during the transmission of the k th ($0 \leq k \leq l$) message. Taking the expected value yields

$$E(d_a|l) = \bar{r} + \sum_{k=1}^l \overline{t_{ck}'} \quad (3.20)$$

Since the scan times are independent samples of the same pdf, they will have equal mean values, permitting (3.20) to be simplified to $E(d_a|l) = \bar{r} + l\bar{t}_c'$. The conditioning on l is removed by multiplying by the queue length pdf and summing over all possible queue lengths which yields

$$\bar{d}_a = \bar{r} + \bar{l}\bar{t}_c' \quad (3.21)$$

1 The result of M/G/1 theory for queuing delay is not applicable in this case. According to M/G/1 theory, if customers stop arriving, the server actually becomes idle and a new customer would receive immediate service. For the token pass network the token never stops circulating the loop and therefore a newly arriving message to an empty queue may not receive immediate service but must wait for the arrival of a free token.

The mean queue length is determined from Little's result which is given by ([10], page 17)

$$\bar{T} = \lambda \bar{d}_a \quad (3.22)$$

Substituting (3.22) into (3.21) and solving for $\bar{d}_a$ yields the desired result

$$\bar{d}_a = \frac{\bar{r}}{1-p'} \quad (3.23)$$

where p' is given by (3.18). A plot of the access delay versus λ for various numbers of stations is found in Fig. 3.4.

As $p' \rightarrow 1$, the mean queue length increases without bound. It is also easily shown that as $p \rightarrow 1$, $p' \rightarrow 1$ and therefore as previously noted $p = \lambda \bar{t}_c < 1$ is a queue stability bound.

3.2.4 Mean Queue Length

The mean queue length is found by application of Little's result. Substituting (3.23) into (3.22) yields the desired result

$$\bar{T} = \frac{\lambda \bar{r}}{1-p'} \quad (3.24)$$

A plot of the mean queue length is found in Fig. 3.5 for various numbers of stations.

Fig 3.4 Mean access delay versus traffic for a homogeneous network with N stations.

Fig 3.5 Queue length versus traffic for a homogeneous network with N stations.

3.3 HETEROGENEOUS TRAFFIC NETWORK

The network model is extended in this section to include heterogeneous traffic. The traffic generated by each host will be Poisson with rate λ_i for the i th station. There are otherwise, no other changes to the model of section 3.2.

3.3.1 Mean Scan Time

The mean scan time for the heterogeneous network has been calculated and is given by (3.7) which is here repeated.

$$\bar{t}_c = \frac{t_f}{N} \frac{1}{1 - t_f \sum_{i=1}^N \lambda_i}$$

By defining a mean traffic rate $\bar{\lambda} = (1/N) \sum \lambda_i$, (3.7) can be simplified to

$$\bar{t}_c = \frac{t_f}{1 - t_f N \bar{\lambda}} \quad (3.25)$$

Clearly if $\lambda_1 = \lambda_2 = \dots = \lambda_N$ then (3.25) reduces to the homogeneous case.

3.3.2 Mean Square Scan Time

The mean square scan time for the heterogeneous case follows identically the homogeneous case up to (3.12) which is repeated below.

$$\overline{t_c^2} = t_f^2 \sum_{i=1}^N \overline{A_i^2} + t_f^2 \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \overline{A_i A_j} + 2t_f^2 \sum_{i=1}^N \overline{A_i} + t_f^2 \quad (3.12)$$

Also for the heterogeneous case $\overline{A_i} = \overline{A_j} = \lambda_i \overline{t_c}$ and $\overline{A_i A_j} = p_{ij}' p_j = p_{ji}' p_i$ but a more general form of p_{ji}' is required. The expression for p_{ji}' for the heterogeneous case is determined similarly to (3.15). The scan time conditioned on station i having a message to transmit is given by

$$\overline{t_c'} = 2t_f + \sum_{\substack{k=1 \\ k \neq i}}^N t_f p_{ki}' \quad (3.26)$$

Also, as in the homogeneous case $p_{ki}' = \lambda_k \overline{t_c'}$ and substituting for p_{ki}' into (3.26) and solving for $\overline{t_c'}$ yields the mean scan time conditioned on station i transmitting a message during the scan. Now, if $\overline{t_c'}$ is substituted into $p_{ji}' = \lambda_j \overline{t_c'}$, thus the desired result is obtained by

$$p_{ji}' = \lambda_j \left(\frac{2t_f}{1 - t_f \sum_{\substack{k=1 \\ k \neq i}}^N \lambda_k} \right) \quad (3.27)$$

Applying the definition of mean traffic rate, (3.27) reduces to

$$p_{ji}' = \lambda_j \left(\frac{2t_f}{1 - t_f(N\overline{\lambda} - \lambda_i)} \right) \quad (3.28)$$

Substituting the results into (3.12) yields

$$\overline{t_c^2} = t_f^2 \sum_{i=1}^N \lambda_i \overline{t_c} + t_f^2 \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \lambda_j \left(\frac{2t_f}{1 - t_f(N\overline{\lambda} - \lambda_i)} \right) \lambda_i \overline{t_c} + 2t_f^2 \sum_{i=1}^N \lambda_i \overline{t_c} + t_f^2 \quad (3.29)$$

1. The term $p_{ji}' p_i$ obtained by Application of Bayes' Theorem is used as it results in a simpler final expression for $\overline{t_c^2}$.

and with a little algebra reduces to

$$\overline{t_c^2} = t_f^2 (2t_f \overline{t_c} \sum_{i=1}^N \frac{\lambda_i (N\overline{\lambda} - \lambda_i)}{1 - t_f (N\overline{\lambda} - \lambda_i)} + 3N\overline{\lambda} \overline{t_c} + 1) \quad (3.30)$$

If $\lambda_1 = \lambda_2 = \dots = \lambda_N = \lambda$ then (3.30) reduces to the homogeneous case given by (3.18).

3.3.3 Stability Conditions

It has been shown that the condition for stability of the i th station's transmit queue is given by $\lambda_i \overline{t_c} < 1$. Substituting this condition into (3.7) for $\overline{t_c}$ yields

$$\lambda_i \left(\frac{t_f}{1 - t_f \sum_{j=1}^N \lambda_j} \right) < 1 \quad (3.31)$$

Solving (3.31) for λ_i gives the traffic rate bound to ensure stability of the i th station and is given by

$$\lambda_i < \frac{1}{2} \left(\frac{1}{t_f} - \sum_{\substack{j=1 \\ j \neq i}}^N \lambda_j \right) \quad (3.32)$$

Summing the input traffic to all the stations gives an upper bound total traffic rate to maintain network stability. The result is found by summing both sides of (3.32) to give

$$\sum_{i=1}^N \lambda_i < \frac{1}{2} \sum_{i=1}^N \left(\frac{1}{t_f} - \sum_{\substack{j=1 \\ j \neq i}}^N \lambda_j \right) \quad (3.33)$$

which reduces to (c.f. 3.10b)

$$\lambda < \frac{.1}{t_f + Nt_f} \quad (3.34)$$

Satisfying (3.34) does not guarantee that every station in the network will be stable since the distribution of traffic among the stations has not been considered. Satisfying (3.32) for every node in the network will guarantee network stability; however, failure to satisfy (3.34) means that at least one station in the network will be unstable (i.e., have an unstable transmit queue) and therefore (3.32) cannot be satisfied for all stations.

Further insight into station stability can be obtained by considering a two station network with station one and station two generating traffic at a mean rate λ_1 and λ_2 respectively. For network stability (3.32) must be satisfied for both stations and are given by

$$\lambda_1 < \frac{1}{2t_f} - \frac{\lambda_2}{2} \quad (3.35a)$$

$$\lambda_2 < \frac{1}{2t_f} - \frac{\lambda_1}{2} \quad (3.35b)$$

Both of these equations are plotted in Fig. 3.6 for $t_f = 125 \mu\text{sec}$. The area under each curve are the operating points which makes the station represented by the curve stable. An operating point chosen in the region common to both curves, shown by the crossed hash lines in Fig. 3.6a, will satisfy stability at both stations. Outside the common area, regions of combinations of stability and instability of

a. Stability regions

Fig 3.6a Plot of stability regions.

Fig 3.6b Plot of stability regions.

both stations can be found.

Consider an example to help determine these regions. The two station network has a t_f of 125 μsec . From (3.34) the mean station traffic must not exceed 2666 msg/sec. The operating point given by $\lambda_1 = \lambda_2 = 2000$ msg/sec, shown as point 'A' in Fig. 3.6 satisfy this condition as well as conditions of (3.35a) and (3.35b). The mean scan time is 250 μsec . Let the traffic to station two increase until $\lambda_2 = 3000$ msg/sec as shown by point 'B'. At this point, the mean scan time has increased to 333.3 μsec ; condition (3.35b) is violated and $\lambda_2 \bar{t}_c = 1$. Therefore a circulating token will find a message to transmit at station two on every scan with probability one. Any increase in traffic rate by station two beyond the line $\lambda_2 = 1/(2t_f) (\lambda_1)/2$ will be lost with no effect on the stability of station one. Therefore station two will be unstable while station one remains stable in this region as shown in Fig. 3.6a.

Beginning at point 'A', let the traffic to station one increase until $\lambda_1 = 3000$ msg/sec shown by point 'C'. At this point $\bar{t}_c = 333.3$ μsec and $\lambda_1 \bar{t}_c = 1$ indicating the arrival of a free token to station one is used with probability one. An increase in traffic rate to station one cannot increase this probability and therefore the excess rate is the traffic loss rate. Therefore an increase in traffic to station one such that the operating point crosses the line $\lambda_1 = 1/(2t_f) (\lambda_2)/2$ results in station one being unstable while station two remains stable. An operating point in the region shown without hash lines in Fig. 3.6a will result in both stations one and two

being unstable.

Now consider an operating point shown as point 'D' in Fig. 3.6b. This point is very close to the line of instability. While 'D' is the mean value of the operating point, the actual instantaneous operating point at some instant in time would be in some region around 'D'. The circle drawn around 'D' represents the distance the operating point may move while maintaining system stability. If the operating point 'D' is moved along a line of constant total traffic (i.e., $\lambda_1 + \lambda_2 = 4000$ in this example), then the distance to any instability boundary can be maximized and in this example is given by point 'A'. Excursions about point 'A' can be larger than about 'D' and therefore point 'A' is said to be more stable than point 'D'. The most stable operating points occur along the line of homogeneous traffic given by $\lambda_1 - \lambda_2 = 0$ for the two station example and in general a homogeneous network is more stable than a heterogeneous network carrying equal total traffic. Also, all points on a line of constant total traffic have equal mean scan times. Therefore the mean scan time at points 'A' and 'D' are the same. Finally, it is interesting to note that the maximum possible total traffic can be put onto the loop when traffic is homogeneous.

3.3.4 Mean Access Delay

The mean access delay for a message generated at station i is given by (3.21) and is repeated below

$$\bar{d}_{a_i} = \bar{r} + \bar{T}_i \bar{t}_c'$$

where $\bar{t}_c'$ is the mean scan time conditioned on station i transmitting. Proceeding as with (3.26), the result is given by

$$\bar{t}_c' = \left(\frac{2t_f}{N} \right) \frac{1 - t_f \sum_{\substack{k=1 \\ k \neq i}} \lambda_k}{1} \quad (3.36a)$$

Substituting Little's result for $\bar{T}$ and (3.36a) for $\bar{t}_c'$ into (3.21) yields

$$\bar{d}_{a_i} = \bar{r} + \bar{d}_{a_i} p_i' \quad (3.36b)$$

where

$$p_i' = \left(\frac{2t_f}{N} \right) \frac{\lambda_i}{1 - t_f \sum_{\substack{k=1 \\ k \neq i}} \lambda_k}$$

Solving for $\bar{d}_{a_i}$ yields the final result

$$\overline{da_i} = \frac{\overline{r}}{1-p_i'} \quad (3.36)$$

where $\overline{r} = (\overline{t_c^2})/2\overline{t_c}$ and $\overline{t_c}$, $\overline{t_c^2}$ and p_i' are given by (3.25), (3.30) and (3.36b) respectively.

3.3.5 Mean Queue Length

Again by application of Little's result the mean queue length can be found from (3.36) and for station i is given by

$$\overline{T_i} = \frac{\lambda_i \overline{r}}{1-p_i'} \quad (3.37)$$

3.4 HETEROGENEOUS TRAFFIC WITH BATCH ARRIVALS

In this section batch messages are considered. A file server such as a disk drive may need to send an entire file as a message. If the file is larger than the data field, then the message must be broken up into blocks so that the length of each block does not exceed the length of the data field. Therefore each message is broken up into smaller units called blocks. A set of blocks corresponding to a single message is called a batch message or simply a batch.

Batches are assumed to arrive to the transmit queue according to a Poisson process with rate λ_b . The number of blocks in each

batch (called the batch size) depends on the size of the DATA field. It is in general a random variable and has a mean value of $\bar{g}_i$. The actual batch size distribution is only required to determine the block access delay.

Using $M[x]/G/1$ theory for batch arrivals, the probability of a station transmit queue holding at least one message for transmission at the instant of arrival of a free token is found similarly to the single block message case derived in section 3.2 and is given by ([5], page 257)

$$p_i = \lambda_i \bar{t}_c = \lambda_{Bi} \bar{g}_i \bar{t}_c \quad (3.38)$$

where $\bar{t}_c$ is the scan time for batch arrivals to be determined and $\bar{g}_i$ is mean batch size generated at station i .

3.4.1 Mean Scan Time

The development of the mean scan time result follows exactly as for the heterogeneous case except that for batch arrivals (3.38) replaces (3.6). The mean scan time is therefore given by

$$\bar{t}_c = \frac{\bar{t}_f}{1 - \bar{t}_f N \bar{\lambda}_B} \quad (3.39)$$

where

$$\bar{\lambda}_B = \frac{1}{N} \sum_{i=1}^N \lambda_{Bi} \bar{g}_i$$

If $\bar{g}_i=1$ then the result for the Poisson heterogeneous case is obtained, given by (3.25).

3.4.2 Mean Square Scan Time

The mean square scan time is given by (3.30) and making the substitution

$$\lambda_i = \lambda_{Bi} \bar{g}_i \quad (3.40)$$

yields the result

$$\overline{t_c^2} = t_f^2 (2t_f \overline{t_c} \sum_{i=1}^N \frac{\lambda_{Bi} \bar{g}_i (N\bar{\lambda}_B - \lambda_{Bi} \bar{g}_i)}{1 - t_f (N\bar{\lambda}_B - \lambda_{Bi} \bar{g}_i)} + 3N\bar{\lambda}_B \overline{t_c} + 1) \quad (3.41)$$

3.4.3 Mean Access Delay

The access delay given by the time interval beginning when a batch arrives to a queue until the first message in the batch begins transmission is given by (3.36) where $\bar{r} = (\overline{t_c^2} / 2\overline{t_c})$ and $\overline{t_c}$, $\overline{t_c^2}$, and p_i' are given by (3.39), (3.41) and

$$p_i' = \lambda_{Bi} \bar{g}_i \left(\frac{2t_f}{N \sum_{\substack{k=1 \\ k \neq i}}^N \lambda_{Bk} \bar{g}_k} \right) \quad (3.42)$$

The access delay for an arbitrary message within the group is a

function of the batch size distribution. The access delay is given by [11]

$$\overline{da_i} = \frac{\overline{r}}{1-p_i} + \frac{\overline{t_c}}{2(1-\lambda_{Bi}\overline{g_i}\overline{t_c})} \left(\frac{\overline{g_i^2}}{\overline{g_i}} - 1 \right) \quad (3.43)$$

where $\overline{g^2}$ is the second moment of the batch size. If the batch size is constant then $\overline{g^2} = (\overline{g})^2$ and (3.43) can be simplified. For a geometrically distributed batch size given by ([12] page 142)

$$P[\text{batch size}=g] = p(1-p)^{g-1} \quad (3.44)$$

where $p = (\overline{g})^{-1}$, $\overline{g}$ is the mean batch size and $\overline{g^2} = 2\overline{g}^2 - \overline{g}$.

3.4.4 Mean Queue Length

The mean queue length is determined from Little's result and is given by

$$T_i = \lambda_{Bi}\overline{g_i}(\overline{da_i})$$

where $\overline{da_i}$ is given by (3.43).

3.5 COMPARISON OF NONEXHAUSTIVE TO EXHAUSTIVE SERVICE

Some insight into nonexhaustive service can be gained by comparing identical loops with exhaustive and nonexhaustive service. The definition of exhaustive service will be one used by Hayes [14].

A station is exhaustively serviced if after the arrival of a free token, all messages which have arrived previous to the arrival of the free token are transmitted on successive frames, that is on intervals of t_f . Messages which arrive during the transmission of these messages (i.e. after the arrival of the free token) must wait until the next scan.

In order to make the exhaustive versus nonexhaustive service comparison, the loops must be identical in all other aspects. Both loops will be synchronous with a frame time t_f . The network will be homogeneous and each station will generate Poisson traffic with rate λ . All buffers will be assumed infinite in length unless stated otherwise.

Consider the problem by a simple analogy. Consider a series of service counters in a loop, perhaps sales counters. Customers queue up to the counters according to the service they require. The counters are serviced by a single clerk who requires a fixed amount of time to walk from one counter to the next. Exhaustive service means that all customers in the queue at the moment the sales clerk arrives will be serviced before the clerk moves to the next counter. Nonexhaustive service implies the clerk services only the first customer in the queue before moving on to the next counter.

An exhaustively serviced loop will have a 'long' scan time since all the customers in a queue must be serviced before the clerk can move to the next counter but once the clerk arrives at a counter all messages queued there are serviced in succession. A nonexhaus-

tively serviced loop will have a shorter scan time since only one customer in each queue is serviced before the clerk moves to the next queue but a customer in a queue must always wait as many scan intervals as there are customers ahead of it before receiving service. The questions to be answered are for which service discipline will a customer be serviced more quickly on the average and which discipline produces a more stable system. The solution begins by first considering the exhaustive service network.

3.5.1 Statistics For Exhaustive Service

A local area network synchronous loop with exhaustive service, homogeneous Poisson traffic with rate λ is characterized as follows [14].

1) MEAN SCAN TIME

$$\overline{t_s} = \frac{t_f}{1 - t_f N} \quad (3.46)$$

2) MEAN SQUARE SCAN TIME

$$\overline{t_s^2} = \frac{t_f}{1 - N(\lambda t_f)^2} \{ N(N-1)(\lambda \overline{t_s})^2 + 3N(\lambda \overline{t_s}) + 1 \} \quad (3.47)$$

3) STABILITY CONDITION

$$N t_f < 1 \quad (3.48)$$

4) MEAN ACCESS DELAY¹

$$\overline{d_s} = \overline{r} (1 + \lambda t_f); \overline{r} = \frac{\overline{t_s^2}}{2\overline{t_s}} \quad (3.49)$$

These results are derivable using the methods described in this chapter.

3.5.2 Stability Comparison

Begin by noting that the exhaustive service network has an unbounded scan time in contrast to the nonexhaustive network as shown in Fig. 3.7. Also as shown in this figure, the exhaustively serviced network will remain stable for higher traffic rates assuming infinite buffers and for homogeneous traffic.

But more realistically, stations do not have infinite length buffers, do not operate only at their mean values, do not have homogeneous traffic and so the picture is quite different. Consider an exhaustively serviced network with about equal average traffic rate to each station. A sharp increase in traffic to any one station will cause the token to be held by that station for a period of time long enough to transmit all of its messages.

Simultaneously traffic continues to arrive to the remaining stations. If the token is held too long, other stations may overflow their buffers and therefore become unstable at the expense of

¹ See Appendix I

Fig 3.7 Mean scan time versus traffic for an exhaustively and nonexhaustively serviced homogeneous network. ($N=50$)
 Y- Traffic bound for an exhaustively serviced network to maintain stability
 X- Traffic bound for a nonexhaustively serviced network to maintain stability

servicing one high traffic station. A screaming station¹ could therefore crash the entire network. If the network was nonexhaustively serviced, a high traffic station could not hold the token longer than required to transmit only one message per scan thereby having a smaller effect on the remainder of the network. Even if one station begins screaming, the remainder of the network could remain stable. The graph of Fig. 3.8 illustrates this effect for a 200 station network. All stations are nonexhaustively serviced and generate traffic at a rate of 30 msg/sec. Station 200 increases its traffic up to 1000 msg/sec while the remainder of the network remains constant. The access delay to station 200 increases sharply as shown in Fig. 3.8 while the affect on the remainder of the network is comparatively small. This graph can be obtained by plotting (3.36). The conclusion is that the nonexhaustively serviced network is more stable than the exhaustively serviced network.

3.5.3 Access Delay Comparison

The access delay for a network using exhaustive and nonexhaustive service is graphed in Fig. 3.9. The nonexhaustive network has a larger delay for all values of traffic rates. The reason for this is that after a nonexhaustive service the token is always passed to the next station resulting in a walk time between all message transmis-

1 A station which is jammed always transmitting.

Fig 3.8 Effects of a high traffic station on a nonexhaustively serviced network. Traffic at stations 1-199 remained constant at 30 messages/sec.

Fig 3.9 Mean access delay for an N station network with exhaustive and nonexhaustive service.

sions. The exhaustively serviced station, on the other hand, may transmit many messages before a walk time is required to send the token to the next station.

The solution to the question is now known. The nonexhaustively serviced network has been shown to be more stable but the price paid is a higher mean access delay. The classical delay, stability tradeoff is demonstrated here. The extra walk time is wasted bandwidth and some of it may be utilized while still maintaining good stability. The technique is to service some of the high traffic stations exhaustively while the remaining stations are serviced nonexhaustively. The effect is to offer higher traffic stations a higher priority compared to the remainder of the network. The stability can be further improved by fixing a maximum number of messages which can be transmitted at the exhaustively serviced stations but this further improvement will not be considered. This service mechanism is called mixed service and is discussed in the next section.

3.6 MIXED SERVICE NETWORK

Any performance between exhaustive and nonexhaustive service can be obtained by using mixed service. The concept involves allowing high traffic stations to be serviced exhaustively and thereby giving them high priority service while the remainder of the network is serviced nonexhaustively. In this section results are

derived for a mixed service network.

In general a network will have N_n nonexhaustively serviced stations and N_e exhaustively serviced stations. Traffic generation is Poisson distributed with rate λ_n for nonexhaustively serviced stations and λ_e for exhaustively serviced stations where λ_n does not necessarily equal λ_e . Traffic generation is homogeneous within each service discipline.

3.6.1 Mean Scan Time

The mean scan is given by

$$\bar{t}_m = t_f + t_f \sum_{i=1}^{N_n} \bar{A}_i + t_f \sum_{i=1}^{N_e} \bar{n}_i \quad (3.50)$$

where A_i is defined in section 3.1.1 as an event such that $A_i=1$ if station i has a message to transmit and $A_i=0$ otherwise. Also, n_i is the number of messages waiting for transmission at the i th exhaustively serviced station. The mean scan time has three terms which are the total walk time, the time to service all nonexhaustively serviced stations and the time to service all exhaustively serviced stations.

As in section 3.2.1 it can be shown that $\bar{A}_i = \lambda_n \bar{t}_m$. Since arrivals are Poisson, $\bar{n}_i$, the mean number of message arrivals during a scan to an exhaustively serviced station is given by

$$\bar{n}_i = \lambda_e \bar{t}_m \quad (3.51)$$

Substituting these results into (3.50) yields

$$\bar{t}_m = t_f (1 - t_f \sum_{i=1}^{N_n} \lambda_{ni} - t_f \sum_{i=1}^{N_e} \lambda_{ei})^{-1} \quad (3.52)$$

and for the homogeneous case reduces to

$$\bar{t}_m = \frac{t_f}{1 - t_f (N_n \lambda_n + N_e \lambda_e)} \quad (3.53)$$

The stability conditions have already been determined as given by (3.8) and (3.48) and for the mixed serviced case become

$$\lambda_n \bar{t}_m < 1 \quad (3.54a)$$

$$t_f (N_n \lambda_n + N_e \lambda_e) < 1 \quad (3.54b)$$

3.6.2 Mean Square Scan Time

The mean square scan time is found by squaring the expression for the scan time and then taking the expected value. The result is given by

$$\begin{aligned} \overline{t_m^2} = t_f^2 & \left(1 + 2 \sum_{i=1}^{N_n} \frac{N_n}{A_i} + 2 \sum_{i=1}^{N_e} \frac{N_e}{n_i} + \sum_{i=1}^{N_n} \frac{N_n}{A_i^2} + \sum_{i=1}^{N_n} \sum_{\substack{j=1 \\ j \neq i}}^{N_n} \frac{N_n}{A_i A_j} \right. \\ & \left. + \sum_{i=1}^{N_e} \frac{N_e}{n_i^2} + \sum_{i=1}^{N_e} \sum_{\substack{j=1 \\ j \neq i}}^{N_e} \frac{N_e}{n_i n_j} + 2 \sum_{i=1}^{N_n} \frac{N_n}{A_i n_i} + 2 \sum_{i=1}^{N_n} \sum_{\substack{j=1 \\ j \neq i}}^{N_e} \frac{N_e}{A_i n_j} \right) \quad (3.55) \end{aligned}$$

Recognizing that $\overline{n_i^2}$ is the mean square number of arrivals for a Poisson distribution and is therefore given by [10]

$$\overline{n_i^2} = \lambda_e^2 \overline{t_m^2} + \lambda_e \overline{t_m} \quad (3.56)$$

Also assuming independence between terminals yields

$$\overline{n_i n_j} = \overline{n_i} \cdot \overline{n_j} \quad (3.57a)$$

$$\overline{A_i n_j} = \overline{A_i} \cdot \overline{n_j}$$

and using the result given by (3.16) permits the final result to be obtained, given by

$$\begin{aligned} \overline{t_m^2} = & \frac{t_f^2}{1 - N_e(\lambda_e t_f)^2} (N_n(N_n-1)p_m p_m' + 3N_n p_m + N_e(N_e-1)p_m^2 \\ & + \lambda_e 3N_e \overline{t_m} + 2N_n N_e p_m \lambda_e \overline{t_m} + 1) \end{aligned} \quad (3.58)$$

$$\text{where } p_m = \lambda_n \overline{t_m} \text{ and } p_m' = \lambda_n \left(\frac{2t_f}{1 - (N_n-1)\lambda_n t_f - N_e \lambda_e t_f} \right)$$

Note that if N_e is set equal to zero then (3.58) reduces to the purely nonexhaustive service case given by (3.18) and if N_n is set equal to zero then (3.58) reduces to the purely exhaustive case given by (3.47).

3.6.3 Mean Access Delay

The mean access delay for a nonexhaustively serviced station is given by (3.23) and is repeated below

$$\overline{d_a} = \frac{\overline{r_m}}{1 - p_m} \quad (3.59)$$

where $\overline{r_m} = (\overline{t_m^2})/2\overline{t_m}$. The mean access delay for an exhaustively serviced station is given by (3.49) and is repeated below

$$\overline{d_s} = \overline{r_m} (1 + \lambda_e t_f)$$

It is useful to define an overall mean access delay given by [16]

$$T = \frac{1}{\lambda_T} \left[\sum_{i=1}^{N_e} \lambda_{ei} \overline{d_{ai}} + \sum_{i=1}^{N_n} \lambda_{ni} \overline{d_{si}} \right] \quad (3.60)$$

$$\text{where } \lambda_T = \sum_{i=1}^{N_e} \lambda_{ei} + \sum_{i=1}^{N_n} \lambda_{ni}$$

As an example consider a network with 50 stations. Figure 3.10 is a plot of the access delay showing the delay versus station traffic. The traffic rate generated at both exhaustive and nonexhaustive stations is the same. The delay is plotted for all 50 stations obtaining nonexhaustive service and all 50 stations obtaining exhaustive service. The network is then considered to be operating with mixed service with 25 stations exhaustively serviced and 25 stations nonexhaustively serviced. The delay for the mixed service

Fig 3.10 Mean access delay versus traffic for a mixed service network with 50 stations.

network, and for the exhaustively and nonexhaustively serviced networks are plotted as well as the overall access delay given by (3.60).

It is interesting to note that the mean mixed service delay is not half way between the delays of the exhaustively serviced network and nonexhaustively serviced network even though this example of mixed service has half the stations with each type of service. Rather, the mean mixed delay is closer to the exhaustively serviced network. This is due to the nature of exhaustive service, servicing all queued messages thereby removing all walk times between messages. Therefore in a network with say many low traffic interactive users and a few high traffic file servers, the network can have an improved delay performance which is very close to the exhaustively serviced network by only allowing the file servers exhaustive service. The resulting stability would also approach that of the nonexhaustively serviced stations as discussed in section 3.3.3.

3.7 VOICE ANALYSIS

The frame which circulates around the loop contains a block of voice channels whose number must be determined. Once the SYNC FIELD, TOKEN and GAP FIELD have been determined, the remainder of the frame will be allocated to the data and voice. These two fields must be chosen as a compromise between the probability of blocking a call due to an insufficient number of channels and high access delay due to a

small data field. A small data field means that messages will have to be broken into a larger number of blocks to transmit the message thereby increasing the delay. Results for determining the access delay have been given but a result for the blocking probability must be determined.

Begin by defining a resource as the number of voice channels required to carry a two way conversation. Typically this would be one or two channels. Once a resource is allocated to a call, it is held by the call until it terminates. The required assumptions are that calls arrive to the network according to a Poisson distribution with rate λ , a call holds a resource according to a general distribution with a mean holding time τ and the arrival rate and holding times are independent. Furthermore, blocked calls are cleared which means that if an arriving call does not find a free resource, it is dropped. Given these assumptions, the probability of blocking can be determined by the Erlang B loss formula given by ([15] page 67)

$$B(s, a) = \frac{(a^s/s!)}{\sum_{k=0}^s (a^k/k!)} \quad (3.61)$$

where $a = \lambda\tau$ and s are the number of resources.

The number of resources and therefore voice channels in the frame is determined by (3.61) in order to obtain a specified blocking probability or grade of service. Once the grade of service and number of voice channels have been specified, the data field size can be determined.

CHAPTER 4

SIMULATION

4.1 INTRODUCTION

This chapter describes the model used to simulate the network described in chapter 2 in order to verify the major results of chapter 3. The model was written using the IBM GPSS (General Purpose Simulation System) language. A GPSS flowchart is also provided which indicates the path a message would take as it moves through the network. The simulation results are plotted and compared to their theoretical counterpart. Only the data transmission has been simulated.

4.2 MODEL DESCRIPTION

In Fig. 4.1 a flowchart of the simulation model is given. The diagram indicates the flow of messages through the network. Message statistics are calculated and stored automatically by the GPSS program for output at the conclusion of the simulation.

The flowchart is divided into two main sections called STATION and LOOP. Messages are represented by transactions which are markers that move from block to block determined by the flowchart. The STATION section simulates the host and loop interface together while the LOOP section simulates the token circulating through the loop.

a. Station section model

Fig 4.1 Simulation Model Flowchart.

Although only one STATION section is shown, the actual simulation model would have as many station sections as the actual system. For example a loop with 10 stations would be simulated using 10 STATION sections in the simulation model.

The STATION section has the following operation referring to Fig. 4.1. The generate block injects transactions representing messages into the model according to a Poisson distribution and at a specified rate. For homogeneous traffic the rate would be the same for all stations, heterogeneous traffic would have different rates for the generate block at each station and batch arrivals would have a number of transactions (determined by a pdf) generated at each Poisson arrival instant. The transactions enter the queue block where they queue waiting for entry onto the loop. The transactions are blocked from further movement by the gate block. The gate block is opened allowing a transaction to pass whenever a free token arrives to the station as determined by the loop section. The transaction then passes the logic block which closes the gate. The depart block is passed indicating the message has left the queue and then moves into the terminate block in order to exit the model. The STATION section is mainly responsible for managing message transactions and their queuing statistics. This model section determines, mean queue lengths, mean access delay and maximum queue length during the simulation.

The LOOP model section simulates the frame and token circulating through the loop interfaces and loop. The section begins with

a generate block which generates only one token at the beginning of the simulation. This transaction represents the token. It first enters the block which records the simulation time to be used later. The seize block is entered indicating a scan has begun. The test block is next entered to determine if the queue of the first section contains any messages (transactions). This is equivalent to a free token arriving to that station. If there are any messages, the 'yes' path is taken leading to the logic block which opens the gate at station one, thereby permitting a station one message to leave its queue. The token transaction then moves into the advance block which acts to delay the token for a period of time t_f . This is the time period that the token will remain busy as it travels around the loop once. At the end of this time period the token transaction is permitted to move indicating the token is once again free. Therefore the token moves to the next block which is a test block to determine if any messages are queued at station two. If no messages are waiting then the 'no' path is taken leading to the test block of the next station.

After passing through the final station, the token transaction moves into final advance block used to simulate the sum of all the walk times around the loop. This is followed by the release block indicating the scan has been completed. The assign block is entered to determine the time interval since entering the mark block enabling certain scan time statistics to be determined. This token then reenters the mark block to begin a new scan.

4.3 SIMULATION RESULTS

Presented in this section are the tables and graphs comparing the simulation results to theoretical results. The figures are self explanatory. The simulation results are in good agreement with the theoretical results. The simulation was allowed to continue until the results reached a constant mean value. The mean value obtained was assumed to be the steady state value.

4.3.1 Homogeneous Network

The network was simulated with homogeneous Poisson traffic at different rates. The network contained five stations. Tables comparing the results of simulation versus theoretical results are given in Table 4.1. The graphs of these results are found in Figs. 4.2a-4.2e.

4.3.2 Heterogeneous Network

The heterogeneous network was simulated with five stations. Traffic generated at four of the stations remained constant at $\lambda_1=400$ msg/sec, $\lambda_2=800$ msg/sec, $\lambda_3=345$ msg/sec and $\lambda_4=1070$ msg/sec. These values were chosen to give a wide range of traffic rates. Traffic generated at station five was varied from 0 to 1200 msg/sec. This simulation is designed to demonstrate the independence of station

Table 4.1 Homogeneous Network Results for a 5 Node Network

	Traffic Rate (msg/sec)	Mean Scan Time ($\times 10^{-6}$ sec)	Mean Square Scan Time ($\times 10^{-8}$ sec)	Mean Access Delay ($\times 10^{-6}$ sec)	Mean Queue Length (msg)	Maximum Queue Length (msg)	Load (p)
Simu- tation	400	166.7	3.4	114.5	.045	3	.067
	700	220	6.3	201	.14	5	.154
	1000	334	14.3	444	.444	7	.334
	1200	497	28.5	1209	1.45	15	.596
Theory	400	166.7	3.4	116.1	.046	-	.067
	700	222.2	6.5	200.7	.14	-	.156
	1000	333.3	14.6	437.5	.44	-	.333
	1200	500.0	29.7	1187.5	1.43	-	.6

Fig 4.2a P_i versus λ_i for a 5 station homogeneous network.

Fig 4.2b $\overline{t_c}$ versus λ for a 5 station homogeneous network.

Fig 4.2c t_c^2 versus λ for a 5 station homogeneous network.

Fig 4.2c t_c^2 versus λ for a 5 station homogeneous network.

Fig. 4.2 $\bar{d}_a$ versus λ for a 5 station homogeneous network.

22

OF / DE

Fig 4.2e $\bar{L}$ versus λ for a 5 station homogeneous network.

stability. It can be seen from Fig. 4.3c that the mean access delay to station five messages, increases sharply compared to the remainder of the network which is due to the increase in traffic generated at station five.

Tables 4.2 and 4.3 are tables of the theoretical results and simulation results respectively. The tables are divided into four groups where each group represents a simulation run with a different value of traffic for station five. Fig. 4.3a-4.3c show the results in plotted form.

4.3.3 Heterogeneous Network with Batch Messages

The network with heterogeneous traffic and batch messages was simulated. The mean batch size generated by all stations was held constant at five blocks. The traffic at four stations remained constant at $\lambda_{B1}=80$ msg/sec, $\lambda_{B2}=160$ msg/sec, $\lambda_{B3}=69$ msg/sec, $\lambda_{B4}=214$ msg/sec. These values were chosen to represent a wide range of traffic rates. The traffic rate generated at station five was varied from 0 to 240 msg/sec. The results were tabled and plotted as shown in Table 4.4 and Fig. 4.4a and Fig. 4.4b.

Table 4.2 Heterogeneous Network Results: Theoretical Results for a 5 Node Network (cf. Table 4.3)

	Traffic Rate (msg/sec)	Station Number	Mean Scan Time ($\times 10^{-6}$ sec)	Mean Square Scan Time ($\times 10^{-8}$ sec)	Mean Access Delay ($\times 10^{-6}$ sec)	Mean Queue Length (msg)	Maximum Queue Length (msg)	Load (p)
Group 1	400	1	196.3	4.9	146.2	.058	-	.078
	800	2	196.3	4.9	171.5	.137	-	.157
	345	3	196.3	4.9	143.1	.049	-	.068
	1070	4	196.3	4.9	191.3	.2	-	.21
	290	5	196.3	4.9	140.	.041	-	.057
Group 2	400	1	204.7	5.4	155.4	.062	-	.082
	800	2	204.7	5.4	183.5	.147	-	.164
	345	3	204.7	5.4	151.9	.052	-	.071
	1070	4	204.7	5.4	205.8	.22	-	.219
	500	5	204.7	5.4	161.9	.081	-	.102
Group 3	400	1	218.1	6.2	169.2	.068	-	.087
	800	2	218.1	6.2	202.1	.162	-	.174
	345	3	218.1	6.2	165.2	.057	-	.075
	1070	4	218.1	6.2	228.6	.245	-	.233
	800	5	218.1	5.2	202.1	.162	-	.174
Group 4	400	1	238.9	7.5	189.3	.076	-	.096
	800	2	238.9	7.5	230.2	.184	-	.191
	345	3	238.9	7.5	184.4	.064	-	.082
	1070	4	238.9	7.5	263.7	.282	-	.256
	1200	5	238.9	7.5	281.9	.338	-	.287

Table 4.3 Heterogeneous Network Results: Simulation Results for a 5 Node Network (cf. Table 4.2)

	Traffic Rate (msg/sec)	Station Number	Mean Scan Time ($\times 10^{-6}$ sec)	Mean Square Scan Time ($\times 10^{-8}$ sec)	Mean Access Delay ($\times 10^{-6}$ sec)	Mean Queue Length (msg)	Maximum Queue Length (msg)	Load (p_i)
Group 1	400	1	197.2	4.9	159.8	.069	3	-
	800	2	197.2	4.9	171.5	.14	3	-
	345	3	197.2	4.9	130.	.045	2	-
	1070	4	197.2	4.9	191.	.2	5	-
	290	5	197.2	4.9	125.	.03	2	-
Group 2	400	1	203.8	5.3	145.4	.053	3	-
	800	2	203.8	5.3	188.1	.145	4	-
	345	3	203.8	5.3	156.2	.054	3	-
	1070	4	203.8	5.3	213.7	.244	5	-
	500	5	203.8	5.3	172.5	.078	4	-
Group 3	400	1	217	5.1	168.0	.068	3	-
	800	2	217	6.1	198.7	.153	4	-
	345	3	217	6.1	155.2	.052	3	-
	1070	4	217	6.1	232.8	.261	5	-
	800	5	217	6.1	203.8	.153	5	-
Group 4	400	1	237	7.2	186.	.078	3	-
	800	2	237	7.2	235.9	.188	5	-
	345	3	237	7.2	186.2	.07	3	-
	1070	4	237	7.2	253.4	.262	5	-
	1200	5	237	7.2	283.6	.334	5	-

Fig. 4.3a t_c versus λ_5 for a 5 station heterogeneous network.
 Traffic at stations 1-4 are constant at $\lambda_1=400$,
 $\lambda_2=800$, $\lambda_3=345$, $\lambda_4=1070$ messages/sec.

Fig 4.3b $\overline{t_c^2}$ versus λ_5 for a 5 station heterogeneous network.
 Traffic at stations 1-4 are constant at $\lambda_1 = 400$,
 $\lambda_2 = 800$, $\lambda_3 = 345$, $\lambda_4 = 1070$ messages/sec.

Fig 4.3c $\bar{d}_a$ versus λ_5 for a 5 station heterogeneous network.

Traffic at stations 1-4 are constant at $\lambda_1 = 400$,
 $\lambda_2 = 800$, $\lambda_3 = 345$, $\lambda_4 = 1070$ messages/sec.

Table 4.4 Heterogeneous Traffic Network Results with Batch Arrivals. Traffic Rate at Stations 1 to 4 Remained Constant at B1=80 msg/sec, B2=160, B3=69, B4=214 and Each Had a Mean Batch Size of 5 Blocks

	Station 5 Traffic Rate msg/sec	Mean Batch Size Blocks	Mean Scan Time $\times 10^{-6}$ sec	Mean Square Scan Time $\times 10^{-8}$ sec ²
Simulation	58	5	194.3	4.9
	100	5	204	5.3
	160	5	217.6	6.1
	240	5	238.2	7.4
Theory	58	5	196.3	4.9
	100	5	204.7	5.4
	160	5	218.1	6.2
	240	5	238.9	7.5

Traffic generated at station 5 with a mean batch size of $g=5$ (messages/sec)

Fig 4.4a t_c versus λ for a 5 station heterogeneous network with batch arrivals. Traffic at stations 1-4 are constant at $\lambda_1=80$, $\lambda_2=160$, $\lambda_3=69$, $\lambda_4=214$ messages/sec and each station having a mean batch size of 5 blocks.

Traffic generated at station 5 with a mean batch size of $g=5$ (messages/sec)

Fig. 4.4b t_c^2 versus λ for a 5 station heterogeneous network with batch arrivals. Traffic at stations 1-4 are constant at $\lambda_{s1}=80$, $\lambda_{s2}=160$, $\lambda_{s3}=69$, $\lambda_{s4}=214$, messages/sec and each station having a mean batch size of 5 blocks per message.

CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 INTRODUCTION

The local-area network described in chapter 2 was analyzed and expressions for the steady state mean value statistics were determined. The analysis was made for an N station network with homogeneous¹ Poisson traffic, heterogeneous² Poisson traffic and heterogeneous Poisson traffic with batch arrivals. The service discipline at the loop interface transmit queue was assumed to be nonexhaustive and FIFO.

In this chapter, only the results for the network with heterogeneous Poisson traffic, that is single arrivals rather than batch arrivals are given as this solution provides the most insight into the network. The homogeneous case is found simply by setting the arrival rates equal.

These results are followed by a comparison of exhaustive to nonexhaustive service for otherwise identical homogeneous Poisson traffic networks. The results presented next are of a homogeneous network in which some stations are nonexhaustively serviced while the

-
- 1 Homogeneous traffic, also called balanced traffic, means that equal traffic rates were generated at each station.
 - 2 Heterogeneous traffic, also called unbalanced traffic, means that different traffic rates are generated at each station in general.

remaining stations are exhaustively serviced. This scheme is called mixed service.

5.2 MEAN SCAN TIME

The mean scan time is defined as the average interval of time between successive arrivals of a free token to a station. The mean scan time is dependent upon the probability that a station requires the token to transmit a message when the free token arrives. This probability was determined to be

$$p_i = \lambda_i \bar{t}_c \quad (3.6)$$

where λ_i is the arrival rate to the i th station and t_c is the mean scan time. Since the probability is bounded by $p_i < 1$, the first boundary condition can be determined and is given by

$$\frac{1}{\lambda_i} > \bar{t}_c \quad (3.8)$$

The result indicates that on the average, messages must not be generated at a station more frequently than the arrival of a free token to that station. It was shown in chapter 3 that failure to meet this criterion results in an unstable queue where instability means that the queue does not have a finite steady state queue length.

Summing up the transmission times of all the stations weighted

by the probability a station transmits a message and adding the total walk time permits the derivation of the mean scan time given by

$$\bar{t}_c = \frac{t_f}{1 - t_f N \bar{\lambda}} \quad (3.25)$$

where t_f is the frame time, $\bar{\lambda}$ is the mean arrival rate given by $\bar{\lambda} = (1/N) \sum \lambda_i$ and N is the number of stations in the network. If the condition of (3.8) is substituted into (3.25) a boundary condition on the scan time is found given by

$$\bar{t}_c < t_f + N t_f \quad (3.10a)$$

The result indicates that the mean scan time will not exceed the time required to transmit one message from every station plus the total walk time. The result is somewhat obvious and is a consequence of the service being nonexhaustive.

A plot of p_i versus station traffic and $\bar{t}_c$ versus station traffic are found in Figs. 3.1 and 3.2 respectively.

5.3 MEAN SQUARE SCAN TIME

The mean square scan time is the second moment of the scan time. By itself the mean square scan time offers very little extra insight into the understanding of the network, but it is required to determine the mean queue length and the mean access delay.

A useful result emerges from the development of the expression for the mean square time. It is the probability a station j has a

END

12.06.86

FIN

5.4 STABILITY CONDITIONS

The transmit queue in the loop interface will be considered stable if it has a finite mean queue length. The condition $\lambda_i \bar{t}_c < 1$ must be satisfied for stability. Substituting for $\bar{t}_c$ and solving for λ_i yields the traffic rate bound per queue for maintaining stability and is given by

$$\lambda_i < \frac{1}{2} \left(\frac{1}{t_f} - \sum_{\substack{j=1 \\ j \neq i}}^N \lambda_j \right) \quad (3.32)$$

It is interesting to sum the traffic generated at each station and the result is given by

$$\bar{\lambda} < \frac{1}{t_f + N t_f} \quad (3.34)$$

Using (3.10), the result (3.34) can be rewritten as $\bar{\lambda} < 1/\bar{t}_c$ (cf (3.8)). The interpretation of this result is that average message arrival rate to the network must not exceed the average arrival rate of a free token to a station in order to maintain queue stability. Satisfying (3.34) does not guarantee that every station in the network will have a stable queue since the distribution of traffic among the stations has not been considered. Satisfying (3.32) does guarantee stability for a specific station.

By plotting (3.32) for pairs of stations, stability regions are obtained as shown in Fig. 3.6. There are regions where both stations are stable, both stations are unstable, or only one of two

stations is stable. In an actual network the traffic rates will not be a stationary point in this graph but rather would be represented by a region about a mean traffic rate. Therefore a station with a mean traffic rate close to the boundary given by equality in (3.32) may become periodically unstable as the instantaneous traffic rate crosses the boundary. A station will therefore be considered more stable the further the mean traffic rate is from the stability boundary. It was shown in section 3.2.3 by way of an example that for a fixed total network traffic, that is $\sum \lambda_i = \text{constant}$, the stability of each station is maximized when the traffic is homogeneous. Furthermore, the maximum possible total network traffic while maintaining stability is achieved only when the traffic is homogeneous.

5.5 MEAN ACCESS DELAY

The access delay is defined as the time interval beginning when a message arrives at a queue until the message just begins transmission onto the loop. The delay has two components which are considered individually and summed to obtain the total delay. The components of delay are the time for a free token to arrive to service the station and the time to transmit all messages ahead of the newly arrived message. The first component is called the mean residual scan time and is determined from renewal theory to be

$$\bar{r} = \frac{\overline{t_c^2}}{2\bar{t}_c} \quad (3.19)$$

When a message begins transmission, a free token does not arrive again until one scan time later. Since a message is being transmitted with probability one during the scan, a mean scan time conditioned on the scan containing one transmission with probability one is needed and is denoted by $\bar{t}_c'$. This conditional scan time is the time each message ahead of the new arrival consumes. It was shown in section 3.2.3 that the second component of delay is therefore the average number of messages a newly arrived message finds multiplied by the conditional scan time. Therefore the mean access delay is given by

$$\bar{d}_{ai} = \bar{r} + \bar{l}_i \bar{t}_c' \quad (3.21)$$

for the i th station where $\bar{l}_i$ is the mean queue length. Using Little's result to substitute for $\bar{l}_i$ yields the final result

$$\bar{d}_{ai} = \frac{\bar{r}}{1-p_i'} \quad (3.36)$$

where $\bar{t}_c'$ is given by (3.36a) and p_i' is given by (3.36b). It was stated earlier that $p = \lambda_i \bar{t}_c' < 1$ must be satisfied to maintain stability. Also, if $p \rightarrow 1$ then $p' \rightarrow 1$ and $d_a \rightarrow \infty$ as indicated by (3.36). As the mean delay ($\bar{d}_a$) grows without bound so to does the mean queue length grow without bound showing that $p < 1$ is required for stability.

A plot of the mean access delay versus traffic for a homogeneous network and various numbers of stations is given in Fig. 3.4.

5.6 COMPARISON OF NONEXHAUSTIVE TO EXHAUSTIVE SERVICE

The network statistics as developed for the nonexhaustively serviced network were quoted from the literature for the exhaustively serviced case and are given in section 3.5.1. Using these results, two identical networks with the exception that one is exhaustively serviced while the other is nonexhaustively serviced, are compared. Two main conclusions are made about the comparison and are given below.

Firstly the nonexhaustively serviced network tends to be more stable than the exhaustively serviced network. As an example consider a stable loop network. The stations are serviced nonexhaustively and one station has a sharp increase in traffic until the station becomes unstable. The effect of the unstable station on the remainder of the network is to increase the scan time because the unstable station will require a free token every scan with probability one. Since only one message can be transmitted the effect on the remainder of the network is small. On the other hand, if the stations were exhaustively serviced, then the unstable station may hold the token for a very long time. This would result in the messages overflowing the queues in the remaining stations, making the entire network unstable.

Secondly, the mean access delay for the nonexhaustively serviced station is larger for the exhaustively serviced station with an equal traffic rate as shown in Fig. 3.9. The reason for this is

due to the time required for the token to be transmitted from one station to the next called the walk time. For nonexhaustive service, a walk time occurs between every message transmission. The walk time represents wasted bandwidth since it is not being used for transmission. For exhaustive service many messages may be transmitted from a queue before a walk time occurs. The result is that more messages are transmitted for a given number of walk time intervals; there is less waste bandwidth and lower mean access delays.

The nonexhaustively serviced network has better network stability but higher delays compared to exhaustive service. The exhaustively serviced network has poorer network stability but lower mean access delay. It is possible to obtain a compromise between good network stability and low access delay by using the mixed service network described in the next section.

5.7 MIXED SERVICE NETWORK

Good stability results from nonexhaustive service while a short mean access delay results from exhaustive service. It is possible to achieve a compromise between the two service disciplines by using mixed service. This method involves servicing higher traffic stations with exhaustive service while lower traffic stations are serviced nonexhaustively. This network was analyzed for messages generated according to a Poisson distribution. Assume the network has N_n nonexhaustively serviced stations each having a traffic rate

of λ_n and N_e exhaustively serviced stations each with a traffic rate λ_e . Therefore the traffic is homogeneous for each service discipline but in general $\lambda_e \neq \lambda_n$. The expression for the mean value statistics, such as mean scan time can be found in section 3.6.

An interesting result of the mixed service is that the improvement in access delay is not proportional to the number of stations exhaustively serviced. Consider a network with all stations nonexhaustively serviced and a small number of these stations have a high traffic rate compared to the remainder of the network. If all the stations would be converted to exhaustive service, there would be a large improvement in access delay but a large degradation in stability as previously described. A better solution is to only service the high traffic stations exhaustively. The improvement in this case would still be very large so that the mean network delay as given by (3.60) would approach the delay performance of an exhaustively serviced network while maintaining good stability.

5.8 SIMULATION

All the results except for the mixed service case have been verified by simulation. The simulation was written in IBM GPSS. The simulation results verify the theoretical results with a very good correlation.

5.9 CONCLUSIONS

It has been shown that the integrated voice/data loop described in chapter 2 and modelled in chapter 3 does work. In order to maintain stable operation, though, certain traffic performance bounds must be maintained. It has also been shown that a network with nonexhaustive service has a larger mean access delay compared to a network with exhaustive service, but the nonexhaustively serviced network is intrinsically more stable than the exhaustively serviced network. A compromise between delay and stability can be achieved by using nonexhaustive service on low traffic stations and exhaustive service on high traffic stations. This service scheme is called mixed service.

The voice channels were modelled as blocked calls cleared facilities. The grade-of-service for the voice aspect of the network is given by the probability of a new call being blocked. The blocking probability is given by the Erlang B equation.

It is felt by the author that this thesis offers more than the analysis of a local area network. It offers an approach to solving similar queuing problems for their mean value results.

5.10 SCOPE FOR FURTHER WORK

Mixed service can be more finely tuned by setting a maximum number of messages that may be transmitted from a station each time a

free token arrives. The maximum number of messages transmitted would in general be different for each station and would be a function of station traffic.

The problem in queuing terms is to consider a series of N queues (one per station) serviced cyclically by the token. If up to M_i messages could be serviced at queue i , then expressions for the queue statistics at each queue would be required.

REFERENCES

- [1] A.S. Tanenbaum, Computer Networks, Prentice Hall, 1981.
- [2] B.K. Penny, A.A. Baghadadi, "Survey of computer communications loop networks: part I," Computer Communications, vol. 2, no. 4, Aug. 1979.
- [3] W. Bux et al., "A local-area communication network based on a reliable token-ring system," IBM Research report RZ1117, Dec. 1981.
- [4] C. Tseng, B. Chen, "D-net, a new scheme for high data rate optical local area networks," Globecom '82, p. 949, 1982.
- [5] D. Gross, C.M. Harris, Fundamentals of Queueing Theory, John Wiley and Sons, 1974.
- [6] R. Maciejko, private communication.
- [7] M.A. Leibowitz, "An approximate method for treating a class of multiqueue problems", IBM Journal, pp. 204-209, July 1961.
- [8] W.W. Chu, A.G. Konheim, "On the analysis and modelling of a class of computer communication systems," IEEE Trans. Comm., vol. COM-20, pp. 645-660, June 1972.
- [9] M.L.T. Yuen, et al., "Traffic flow in distributed loop switching systems," Symp. on Computer-Communication Networks and Teletraffic, Polytechnic Institute of Brooklyn, pp. 29-46, April 1972.
- [10] L. Kleinrock, Queueing Systems, Vol. I: Theory. Wiley, 1975.
- [11] P.J. Kuehn, "Multiqueue systems with nonexhaustive cyclic service," BSTJ, vol. 58, no. 3, pp. 671-697, March 1979.
- [12] M. Schwartz, Computer-Communication Network Design and Analysis, Prentice-Hall, 1977.
- [13] R. Maciejko, et al., "A synchronous, folded-loop fibre optic local network," OFC Conference Proceedings, New Orleans, Jan. 23-25, 1984.
- [14] J.F. Hayes, D.N. Sherman, "A study of data multiplexing techniques and delay performance," BSTJ, vol. 51, no. 9, pp. 1983-2011, Nov. 1972.

- [15] R.B. Cooper, Introduction to Queueing Theory, Collier-Macmillan, 1972.
- [16] A. Thomasian, H. Kanakia, "Performance study of loop networks using buffer insertion," Computer Networks 3, pp. 419-425, 1979.

APPENDIX I

The mean scan time moments and access delay given by (3.49) were derived under the assumption of homogeneous traffic. The applicability of the mean access delay equation to the more general heterogeneous case is not evident from the derivation. Therefore, the following simpler derivation is offered.

Consider the loop shown in Fig. A1.1 where only one station's transmit queue is shown. Assume that the instant a new message arrives at this queue, the token is being held by a station shown on the loop as a dot. The access delay for this new arrival is the residual life of the scan time r plus the time required to service all the messages ahead of the new arrival. Since the service is exhaustive and the arrival process is Poisson with rate λ , the time required to service all messages ahead of the new arrival is $\lambda t_f(t_s - r)$ where t_s and r are random variables. The access delay is therefore given by

$$d_s = r + \lambda t_f(t_s - r) \quad (\text{A1.1})$$

It is important to recognize that this is not a work conserving queue due to the time the queue is not being serviced by a free token. Therefore, Little's result is not applicable in this case. The mean access delay can be found by multiplying (A1.1) by the joint pdf of t_s and r given by renewal theory ([10], page 172) and summing over all possible scan times. The mean access delay for the hetero-

geneous, exhaustively serviced case becomes

$$\bar{d}_S = \int_0^\infty \int_0^{t_S} (r + \lambda t_f(t_S - r)) \frac{f(t_S)}{t_S} dr dt_S \quad (A1.2)$$

where $f(t_S)/t_S dr dt_S$ is the joint density of t_S and r and $f(t_S)$ is the pdf of t_S alone. Solving (A1.2) leads directly to the result given by (3.49).

Fig. A1.1 Token pass loop

END

1 2 0 6 . 8 6

FIN