1

CANADIAN THESES ON MICROFICHE

I.S.B.N.

THESES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

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 a poor 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.

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

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 mauvaise qualité.

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.

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

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Division Division des thèses canadiennes

Ottawa, Canada
K1A 0N4

59922

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

• Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

MUSTAFA KEMAL MEHMET ALI

Date of Birth — Date de naissance

16/11/52

Country of Birth — Lieu de naissance

CYPRUS

Permanent Address — Résidence fixe

324 CAMBRIDGE ST. Apt. 509
OTTAWA K1R 7B5

Title of Thesis — Titre de la thèse

PACKET VOICE PERFORMANCE MODELS, FOR LONG-HAUL AND
LOCAL NETWORKS, AND DELAYS WITH A
CSMA/CD PROTOCOL

University — Université

CARLETON UNIVERSITY

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

Ph. D.

Year this degree conferred — Année d'obtention de ce grade

1983

Name of Supervisor — Nom du directeur de thèse

Prof. C. M. WOODSIDE

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 otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE 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

7/4/83

Signature

M. K. M. Ali

PACKET VOICE PERFORMANCE MODELS FOR LONG-HAUL AND LOCAL
NETWORKS, AND DELAYS WITH A CSMA/CD PROTOCOL

by

© Mustafa Kemal Mehmet Ali, B.Sc., M.Sc.

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

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

December 1982

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

"PACKET VOICE PERFORMANCE MODELS FOR LONG-HAUL AND LOCAL
NETWORKS, AND DELAYS WITH A CSMA/CD PROTOCOL"

submitted by Mustafa Kemal Mehmet Ali, B.Sc., M.Sc.

in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

Thesis Supervisor

Chairman
Department of Systems and
Computer Engineering

Carleton University

April 1983

ABSTRACT

This thesis concerns packetized voice and CSMA/CD local area networks.

Three speech reconstruction protocols are studied, which buffer voice packets before delivery to compensate for variable network delays. They employ a fixed delay, a fixed delay partly determined by the network, and variable delay due to waiting for all packets. Distributions of buffer size, receiver delay, and packet loss rate are presented for all three protocols under certain stochastic assumptions. The optimal packet size which minimizes the total delay with a constraint on speech quality is found. Curves showing the trade-offs among packet loss rate, delay, packet size, network utilization, and mean buffer length are derived to guide the designer of packet voice systems.

For voice analysis in CSMA/CD networks the probability distribution of delay is determined approximately. Then the capacity for voice channels is shown to be about 100 (1 Mbits/s network) to 1000 (10 Mbits/s network).

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my supervisor, Dr. C. Murray Woodside, for his constant guidance, encouragement and help during the course of this work. The variety of topics in this thesis is due in great measure to his advice.

I would also like to thank Dr. D.A. Dawson of the Department of Mathematics and Statistics for several helpful discussions.

My very sincere thanks go to the residents of Room 328: Morteza Niktash, Rafik Goubbran, Mohsen Rashwan, Dr. D.A. Stanford, Dr. M. Spyrtos for the friendly and productive atmosphere which greatly helped the completion of this work.

My thanks also go to Diane Dodds for her friendship and frequent help, and to Gail Badour for her care and efficiency in typing this thesis.

I very much appreciate the support and understanding which my family has shown and I am especially indebted to my brother Vecdet who shared with me many difficult moments during the course of this work.

TABLE OF CONTENTS

Chapter 1	INTRODUCTION	1
1.1	Object of the Thesis	1
1.2	Packetized Voice	3
1.2.1	Protocols for Reconstitution of Voice	6
1.2.2	The Interaction Between Delay, Packet Size, and Packet Loss Rate	7
1.2.3	Previous Work	8a
1.2.4	Survey of the Thesis	9
1.3	Local Area Networks	10
1.4	Summary of the Results	10
Chapter 2	THE STATEMENT OF THE MODEL	12
2.1	Model	12
2.1.1	Speaker Model	12
2.1.2	Communications Network	14
2.1.3	Receiver Protocol	14
2.2	Analysis and Notation	15
Chapter 3	DISTRIBUTION OF THE NUMBER OF PACKETS IN THE BUFFER	16
3.1	PGF of the Arrivals to the Buffer	16
3.2	The Buffer Distribution for the Protocol FD	18
3.2.1	PGF of the Number of Packets in the Buffer	19
3.2.2	Expected Number of Packets in the Buffer	21
3.3	The Buffer Distribution for the Protocol CD	21
3.4	The Buffer Distribution for the Protocol TD	23
3.4.1	Distribution of the Number of Packets Delivered to the DA Converter	24
3.4.2	PGF of the Number of the Packets in the Buffer	27
3.4.3	Expected Number of Packets in the Buffer	30
3.5	The Expected Number of Packets in the Buffer for Protocol TD with PVC Network	30
3.6	Results and Discussion	34

Chapter 4	DELAY AND LOST PACKET ANALYSIS	42
4.1	Protocol FD	42
4.2	Protocol CD	43
4.3	Protocol TD	44
4.4	Numerical Results and Discussion	45
4.5	A Buffer Design Calculation	46
Chapter 5	FINITE BUFFER ANALYSIS	49
5.1	Markov Chain Analysis	49
5.2	Buffer Distribution and Expected Number of Packets Discarded due to Overflow	55
5.3	Numerical Results and Discussion	56
Chapter 6	THE STEADY STATE ANALYSIS	61
6.1	Model	61
6.2	The Steady-State Buffer Distribution for Protocol FD	63
6.3	Numerical Results and Discussion	65
Chapter 7	OPTIMAL PACKET SIZE	67
7.1	Statement of the Model	67
7.2	Optimal Packet Size for the FD Protocol	70
7.3	Optimal Packet Size for the CD Protocol	72
7.4	Optimal Packet Size for the TD Protocol	74
7.5	Numerical Results and Discussion	78
7.6	A Packet Voice System Design	79
Chapter 8	DELAYS IN A CSMA/CD PROTOCOL FOR LOCAL AREA NETWORKS	84
8.1	The CSMA/CD Protocol	85
8.2	Analysis of the CSMA/CD Protocol	88
8.2.1	Queue Size Distribution	90
8.2.2	The Waiting Time Distribution	90
8.2.3	Asymptotic Waiting Time Distribution in Heavy Traffic	100
8.3	Numerical Results	102
8.4	Applications to Packetized Voice	103
Chapter 9	CONCLUSIONS AND FURTHER WORK	110
9.1	Conclusions	110
9.2	Suggestions for Further Work	112

Appendix 1	Unconditional Expected Number of Packets in the Buffer for FD Protocol	114
Appendix 2	Unconditional Expected Number of Packets in the Buffer for CD Protocol	116
Appendix 3	Probability Density Function of the Receiver Delay for TD Protocol	119
References		121

LIST OF FIGURES

Figure 1.1	A Packet Voice System	4
Figure 2.1	Speaker Model	13
Figure 3.1	Delivery Pattern for Protocol FD	20
Figure 3.2	Delivery Pattern for Protocol TD	25
Figure 3.3a	Packet Voice Receiver for a Communications Network with Dynamic Routing	32
Figure 3.3b	Packet Voice Receiver for a Communications Network with Fixed Routing	32
Figure 3.4	Mean Buffer Lengths for FD and CD Protocols	37
Figure 3.5	Mean Buffer Lengths for TD Protocol	38
Figure 3.6	Mean Buffer Length for TD Protocol with PVC Approach	39
Figure 3.7	Mean No. of Packets Delivered to the DA Converter for TD Protocol	40
Figure 3.8	Probability Distribution of k packets in the Receiver Buffer for FD, CD and TD Protocols	41
Figure 4.1	Probability Distribution of the Receiver Delay τ_R	47
Figure 4.2	Percentage of Packets Discarded, and Mean Receiver Buffer Delay	48
Figure 5.1	State Transition Diagrams of the Packets for FD Protocol with Finite Receiver Buffer	51
Figure 5.2	Mean Buffer Length for FD Protocol with a Finite Buffer	58
Figure 5.3	Probability Distribution of k Packets in the Receiver Buffer for FD Protocol with a Finite Buffer	59
Figure 6.1	States of the Transmitter and Receiver at Steady-State for FD Protocol	62
Figure 6.2	The Steady-State Probability Distribution of k Packets in the Receiver Buffer for FD Protocol	66

Figure 7.1	Optimal Packet Size and Minimum Total Delay for FD Protocol	81
Figure 7.2	Total Delay for FD Protocol	82
Figure 7.3	Mean and Second Moment of the Total Delay for TD Protocol	83
Figures 8.1a,b	Busy and Idle Periods in the CSMA/CD Protocol	91
Figures 8.2a,b,c	Waiting Time of the Tagged User	93
Figures 8.3a,b	Simplified Waiting Time of the Tagged User	96
Figure 8.4	Mean Delay for the CSMA/CD Protocol	106
Figure 8.5	Variance of the Delay for CSMA/CD Protocol	107
Figure 8.6	Distribution Function of the Delay for CSMA/CD Protocol	108

LIST OF TABLES

Table 5.1	Packet Loss Rate for FD Protocol with a Finite Buffer	60
Table 8.1.	Packet Loss Rate and Average Number of Voice Circuits a CSMA/CD Network may support	109

LIST OF FREQUENTLY USED SYMBOLS

Chapters 1-7

a	Arrival rate to the network (packets/sec)
B	Packet size (in no. of information bits)
$B(z, q, r_{\min}, r_{\max})$	Modified Binomial probability generating function
C	Channel capacity (bits/sec)
CDF	Cumulative distribution function
d	Total number of packets discarded
d^*	Packet loss rate
$E\{X\} = \bar{X}$	Expected value of the random variable x
H	Header size (bits)
j	Sequence number of a packet in a talkspurt
$k(n)$	Number of packets in the receiver buffer at time n
$K(z, n)$	Probability generating function of the distribution of $k(n)$
$k_j(n)$	Random variable showing the status of the packet j at time n .
$K_j(z, n)$	Probability generating function of the distribution of the random variable $k_j(n)$
n	Time in number of slots (packet duration)
N	Number of voice circuits
pdf	Probability density function
PGF	Probability generating function
$p_i(x_i)$	Probability density function of the random variable x_i
$P_i(x_i)$	Cumulative distribution function of the random variable x_i

q	Discrete network delay distribution parameter
R	Voice AD rate (bits/sec)
ss	Silent state
t_D	Total delay of a packet (sec)
t_{Dj}	Total delay of the j 'th packet (sec)
t_h	Header duration (sec)
t_N	Network delay of a packet (sec)
t_p	Packet (slot) duration (sec)
t_R	Receiver delay of a packet (sec)
ts	Talkspurt state
t_S	System delay (network + receiver) of a packet (sec)
t_{Sj}	System delay (network + receiver) of j 'th packet
$T_x(\theta)$	Laplace transform of the pdf of x
v	Channel capacity (packets/sec)
z	Probability generating function variable
α	(Talkspurt length - 1) in packets
ϵ	Transmission efficiency
θ	Laplace transform variable
λ	Parameter of continuous network delay distribution
v	Channel utilization
v_0	Useful channel utilization
ρ	Parameter of the distribution of talkspurt length
γ	Parameter of the distribution of silent interval length

μ	Delay of the first packet in the receiver buffer for CD, TD protocols (in slots)
$\tau_{d,j}$	Beginning of the delivery time of a packet to the DA converter (slots)
τ_D	Total delay of a packet (slots)
τ_N	Network delay of j'th packet (slots)
τ_R	Receiver delay of a packet (slots)
τ_S	System (network and receiver) delay of a packet (slots)

Chapter 8

a	Successful transmission time of a packet ($a=r+\Delta$)
a^*	Successful transmission time of the tagged user
b	Service time of a user ($b=a+c$)
b^*	Service time of the tagged user ($b^*=a^*+c^*$)
c	Contention interval
c^*	Contention interval after which the tagged user is served
d	Total delay of a packet (waiting time + successful transmission time)
N	Number of voice circuits
p_s	Transmission probability of a ready user in a slot
q	Success probability in a slot
r	Transmission time of a packet
$T_X(\theta)$	Laplace transform of the probability density function of X
u	Residual waiting time
v	Any waiting time following residual waiting time
w	Waiting time of the tagged user
$\overline{w^k}$	k 'th moment of the waiting time of the tagged user
$\overline{w_h^k}$	k 'th moment of the waiting time of the tagged user under heavy traffic
Δ	Slot duration
θ	Laplace transform variable
λ	Arrival rate of the ready users to the network (packets/sec)
ρ	Useful server (channel) utilization

σ Server (channel) utilization
 μ Ratio of slot duration to packet transmission time
 τ Tagged user

Chapter 1

INTRODUCTION

1.1 Object of the Thesis

This thesis is clearly divided into two parts: In the first part several packet-voice receiver protocols for synchronous delivery of the packetized voice to the listener have been studied. The question of optimal packet voice receiver design has been addressed; buffer size, packet delay, and packet size are determined with a constraint on speech quality. In the second part of the thesis performance of the local area networks has been considered which resulted in packet delay distribution for a CSMA/CD protocol. The delay distribution is very significant for transmitting packetized voice on Ethernet.

Historically, circuit switching (CS) has been the standard communication system for voice transmission. The characteristics of CS are access with blocking, fixed throughput, low delay, and assignment of the transmission bandwidth to one user during an entire call. Voice consists of alternating active periods (talkspurts) and silences. Statistics have shown that CS systems are under-utilized because of the silent intervals [1], except for TASI which assigns the transmission bandwidth to a user on a talkspurt basis. However, conventional TASI requires large number of voice channels to be multiplexed in order to obtain an acceptable level of speech quality.

Data communications began on circuit-switched lines leased from telephone networks. However the burstiness of data traffic, its high peak to average ratio, resulted in the development of packet-switched networks (PS) for data communications. The characteristics of PS are access with buffering, variable packet delay and throughput seen by a user. PS achieves higher utilization by dynamic sharing of resources.

In recent years a considerable amount of research has gone into integration of voice and data in a digital network. It is both economically and technically advantageous to have one instead of two communication systems [2]. Low utilization of CS systems for voice as well as the success of PS for data have also encouraged this research. The objectives of the integration are: higher system utilization, reduced costs due to sharing of transmission and switching facilities, and greater flexibility in adjusting to changing service demand.

Several approaches have been proposed for integration:

1) CS for voice and PS for data.

Voice is circuit switched, but silent intervals of voice are used for the transmission of data packets.

2) PS for both voice and data.

Voice is digitized, coded, packetized, and then transmitted over a packet switched network; voice packets are transmitted only during talkspurts.

3) Packet Virtual Circuit (PVC) for voice and PS for data.

As in the previous scheme voice and data are transmitted in packets but voice packets follow a fixed route (the virtual circuit).

Each of the above schemes has advantages and disadvantages widely discussed in the literature [2,3,4,5]. Briefly, the first approach has good delay properties while it falls behind in dynamic sharing of network resources. On the other hand because they give randomly varying delays the other two approaches present problems in transmission of voice but have excellent resource sharing. Voice is a real-time phenomenon with stringent delay requirements. Though a fixed delay does not affect the intelligibility of the speech, when one-way delay of a voice packet is more than 0.5 sec., conversations become difficult. (The speaking parties may interrupt each other in mid-sentence, for instance). Variability of the delay is quite undesired due to synchronous voice requirement at the receiver output. In the first half of the thesis we address the delay related problems of packetized voice in either PS or PVC schemes.

1.2 Packetized Voice

A packet voice system consists of several parts as shown in Figure 1.1. At the transmitting end voice is detected, digitized, coded, and following the packetization transmitted to the network. Different methods of coding exist with data rates varying from 1-96K bits/sec. In general speech quality increases with bit rate. It takes a time t_p called the packetization delay to form a packet, given by:

$$t_p = \frac{B}{R}$$

B = packet size (no. of information bits)

R = voice AD rate in bits/sec.

Packetization delay (t_p) will also be referred as packet duration or an slot.

FIGURE 1.1
A Packet Voice System

For $B = 1000$ bits and $R = 64K$ bits/sec, the packetization delay is 16 msec.

Also a header of H bits is attached to each packet, with duration:

$$t_h = \frac{H}{R}$$

In the communications network packets follow a fixed or adaptive route depending on PVC or PS schemes. Network delay of a packet is made up of two parts: propagation delay and nodal delay. Propagation delay is the time required to transmit a pulse along a transmission line, and is usually negligible. Nodal delay is the queueing delay plus transmission time of a packet at various switching centers in the network. This is random and results in asynchronous arrival of the packets to the receiver.

Network delay of a packet will be denoted by t_N seconds or in terms of t_p by $\tau_N = t_N/t_p$ packet-durations, or 'slots'.

As has been pointed out, voice requires synchronous delivery to the listener. To meet this we need to compensate for the random network delay of a packet before it is delivered to the user, and this is done by packet voice receiver (Figure 1.1). The packet voice receiver performs the following duties:

- 1) Reordering Packets

Voice packets are numbered at the transmitter, and the packet voice receiver makes sure that they are delivered to the listener in the correct order. A packet which arrives to the receiver out of sequence

is delayed, until all the packets generated earlier are delivered to the listener. This will be called reordering delay.

2) Delaying Talkspurt

Reordering delay introduces undesired gaps in the talkspurt. In order to limit occurrences of these gaps the first packet of the talkspurt is delayed by an extra amount of time. This will be called intentional delay.

Receiver delay of a packet is the sum of the above two components, and it will be denoted by t_R or in terms of t_p by τ_R .

Total delay of a packet (end-to-end) will be denoted by t_D (or in terms of t_p by τ_D) and is given by:

$$\begin{aligned} t_D &= t_p + t_N + t_R \\ \tau_D &= 1 + \tau_N + \tau_R \end{aligned} \quad (1.1)$$

For convenience the system delay of a packet (t_S) will be defined as its network delay plus its receiver delay:

$$t_S = t_N + t_R \quad (1.2)$$

or in terms of slots, as τ_S :

$$\tau_S = \tau_N + \tau_R$$

1.2.1 Protocols for Reconstitution of Voice

Several protocols have been suggested for implementation in packet voice receivers [6,7]. All of these protocols require synchronous clocks at the transmitter and receiver. Packets are time-stamped at the transmitter, thus allowing the network delay of a packet to be determined by the receiver when the packet arrives there. Next we shall describe three such protocols.

i) Protocol FD:

In this protocol reconstructed speech follows the original by a fixed delay (the name FD), with gaps for the packets which arrive too late. Total delay of a packet is fixed, and usually based on the mean network delay of the previous talkspurt.

ii) Protocol CD:

This protocol is similar to the FD, except that the total delay of a packet is based on the network delay of initial packet of a talkspurt (total delay = network delay of the initial packet + constant). This protocol follows the changes in the channel delay (the name CD) more closely than the protocol FD.

iii) Protocol TD:

This protocol waits for any late arrivals, thus no packets are discarded. If a packet is late the system waits until it arrives and the receiver delay is increased by that amount for all subsequent packets. This protocol keeps tracking the channel delay (the name TD) until all the packets of a talkspurt are delivered.

1.2.2 The Interaction Between Delay, Packet Size, and Packet Loss Rate

The various components of the total delay depend on each other:

Packetization delay is decreased by having shorter packets. However, because of the fixed header attached to each packet, shorter packets mean increased load on the network which results in increased network delay, and in lower transmission efficiency. When network delay is

increased its variation also increases. To compensate, a larger intentional receiver delay is needed, but this is not desirable because of the increased total delay as well as excessive buffering. Voice is a rather redundant information source, and it has been found that infrequent loss of short segments of voice and occasional errors may be tolerated [8]. To limit the total delay as well as its variation retransmission and error checking of voice packets may be neglected and loss of certain percentage of packets is accepted. Experimental studies on speech quality have shown that a loss of 1% of packets causes negligible degradation while a loss of 5% is noticeable but acceptable [5].

The trade-off between the packet size, delay and loss rate leads us to the following design problems:

- 1) From the above discussion it is clear that there is a packet size which minimizes the total delay. Thus the first problem is the determination of the optimal packet size which minimizes some function of the mean square total delay and its variance under a constraint on speech quality.

- 2) Determination of the required buffer size in terms of packet size, and the constraint on speech quality.

1.2.3 Previous Work

Some work has been done on both the above problems, and we summarize it next.

D. Minoli [9] has modelled the packet-switched network by a single server queue, and determined the mean network delay of a packet. Then he has minimized the packetization delay plus mean network delay to determine the optimal packet size. This minimization did not include receiver delay or any constraint on speech quality.

Barberis and Pazzaglia [10] have studied the delay distribution of the protocols which wait for late packets. Two protocols have been studied:

- 1) The TD protocol as described above (where the first packet of the talkspurt is delayed by μt_p seconds in the receiver before it is delivered to the DA converter.)
- 2) A second version of the TD protocol in which, if the network delay of the first packet of the talkspurt is less than μt_p it is delayed in the receiver until μt_p . Otherwise its delivery begins to the DA converter as soon as this packet arrives to the receiver.

For both of these protocols they determined the probability density function (pdf) of system delay of a packet assuming independent, identically distributed (iid) network delays with arbitrary distribution, and a geometrically distributed number of packets in a talkspurt. They

have also proposed to determine the value of μ by minimizing a function of the mean and variance of the system delay wrt μ , but the minimum was not found. They have also described a synchronization algorithm which reduces the gap between transmitter and receiver clocks, which is needed for their second protocol.

In a later work Barberis [11] has derived the pdf of receiver delay of a packet under the same assumptions. He has also done simulation with a finite buffer, to determine the percentage of packets discarded due to overflow. These simulations show that finite buffer capacity in number of packets should be greater than the parameter μ in order to obtain an acceptable overflow probability.

Suzuki et al [13] have done some work on the FD protocol. According to the FD protocol, the system delay of a packet is fixed, and they based it on the mean network delay of the packets belonging to the previous talkspurt. Assuming that the network delay is exponentially distributed they have found a method to estimate the mean network delay in the absence of synchronized clocks at the transmitter and receiver. The method uses the fact that the mean and variance of the exponential distribution are identical.

Naylor and Kleinrock [28] have analysed the gaps occurring in a talkspurt for the CD and TD protocols. They assumed a fixed number of packets in the talkspurt, and iid network delays. They proposed a method to determine a good value for μ : First determine the smallest of the network delays of the most recent μ packets from the previous

talkspurt. Then find the delay difference between the network delay of each of the u packets and the smallest delay, and choose u as the u_k th largest delay difference. As a rule of thumb choose $u > 40$ and u_k approximately 5 percent of u for CD and 7 percent for TD. They found the gap frequency (the average number of packets arriving to an empty receiver buffer) for both protocols, TD giving lower values but at the expense of longer gap durations.

In a recent work Gopal et al [12] have analysed the gap frequency under the same assumptions except with fixed routing in the network and iid packet interarrival times to the receiver. They have found that only those interarrival time distributions with a small coefficient of variation gives satisfactory gap frequency. Further for the CD protocol, the gap frequency is affected by the number of packets in the talkspurt because a late packet delays all the packets coming after it. On the other hand the gap frequency for the TD protocol is not affected because of the resynchronization of the delivery after each new arrival to the receiver.

1.2.4 Survey of the Thesis

This thesis discusses a number of fundamental performance results, and solves the design problems outlined above for protocols FD, CD and TD. Thus we determine the optimal packet size which minimizes total delay, and corresponding nonstationary buffer distribution of packets for each of these protocols. All the results to be presented are new except the delay analysis of the TD protocol. In the following we describe the contents of the chapters; unless otherwise stated the results apply only to PS schemes. Design problems are discussed in opposite order; chapters 3-6 discuss the second problem, while chapter 7, the first one.

In the next chapter, the main assumptions of the thesis are outlined, and notation is discussed. In chapter 3, the probability generating function (PGF) of the buffer distribution of the packets is determined for all three protocols assuming a single talkspurt and infinite buffer. For protocol TD, also the distribution of the number of packets delivered

to the listener is derived as a function of time. At the end of the chapter an approximate analysis of the PVC approach with TD protocol is given.

In chapter 4, the delay analysis of all three protocols is given, and the distribution of the discarded packets is derived for protocols FD and CD. Also a design example is given on determining the buffer size requirement.

In chapter 5, the distribution of the packets in the buffer is determined for the protocol FD and a finite buffer, using a Markov chain approach.

In chapter 6, using the approach of chapter 3, the buffer distribution is determined for a speaker which alternates between the talkspurt and silence. This gives the steady state analysis of the FD protocol.

In chapter 7, the first design problem is discussed. For protocols FD and CD packet size which minimizes the expected total delay for a fixed packet loss rate is determined. For protocol TD, packet size which gives minimum mean square delay has been determined. Buffer size requirements have also been related to these results.

1.3 Local Area Networks

Local area networks (LAN's) are designed to serve to a large number of users within a short distance that are characterized by bursty traffic. LAN's provide a high speed communications channel (0.1-10 Mbps) over a distance of approximately 0.1-1 Km. Typical

applications of LAN's include office information exchange (e.g. word processing, electronic mail, file transfer, and terminal access to central computer), process control (assembly line automation) and packetized voice. Two major implementations of LAN's are Ethernet and ring networks; in this work we only deal with the former. Ethernet uses a shared channel with distributed control using carrier sense multiple access with collision detection (CSMA/CD) for access to the channel [14]. So far delay results on CSMA/CD protocols have been limited to the mean value [15,16,17]. For many applications this is sufficient but real time applications such as voice depend on the delay distribution as shown in chapters 2-7. In a very recent work, Tobagi [18] has determined the delay distribution for a CSMA/CD protocol, but without computationally easy results. In chapter 8, for a slightly different CSMA/CD protocol we determine the distribution of the delay and its moments. A very simple asymptotic result is given under heavy traffic, which match the mean value of delay of the LAN for all traffic levels and the variance of delay for intermediate levels of traffic as well.

1.4 Summary of the Results

The most significant results of this thesis are:

- i) The distribution of the number of packets in the buffer (Chapter 3),
- ii) The optimal packet size which minimizes the total delay (Chapter 7),
- iii) The distribution of the delay for a CSMA/CD protocol (Chapter 8).

Chapter 2

THE STATEMENT OF THE MODEL

In this chapter we state the model used in the thesis. We present the main assumptions of this work and establish the notation used in Chapters 2-7.

2.1 Model

The model is based on Figure 1.1, and its main characteristics are discussed below.

2.1.1 Speaker Model

In a typical conversation each speaker alternates between talkspurt and silent intervals, and call durations are on the order of hundreds of seconds [19]. Following Brady [1], lengths of talkspurt and silent intervals will be modelled as random variables with exponential distributions, with respective mean values 1.32 and 1.82 sec. Experimental studies in [1] have shown that while this model fits well to the talkspurts, it is somewhat unrealistic for silent intervals. It is used because it is simple.

$$p_t(\alpha) = P\{(\alpha+1) \text{ packets in a talkspurt}\} = (1-\rho)\rho^\alpha \quad \alpha = 0,1,2,\dots \quad (2.1)$$

$$p_s(\beta) = P\{(\beta+1) \text{ packets in a silent interval}\} = (1-\gamma)\gamma^\beta \quad \beta = 0,1,2,\dots \quad (2.2)$$

As a result of this a speech source may be modelled by a two state Markov chain:

Figure 2.1. The Speaker Model

where above ts and ss stand for talkspurt and silent states respectively.

Steady-state probabilities are given by:

$$P\{X=ts\} = \frac{1-\gamma}{2-\gamma-p}, \quad P\{X=ss\} = \frac{1-p}{2-\gamma-p} \quad (2.3)$$

Except for the analysis of chapter 6, we shall be dealing with a single talkspurt which is geometrically distributed. During a talkspurt starting at time $t = 0$, the speech signal is encoded into packets of length t_p sec., which are transmitted into the communications network at equal intervals. These transmission instants will be referred to by the discrete time index $n = 0, 1, 2, \dots$, and the corresponding time unit as a slot. Thus j 'th packet of the talkspurt enters to the network at the discrete time $n = j - 1$.

2.1.2 Communications Network

The network delays of all packets belonging to a talkspurt are assumed to be independent, identically distributed (i.i.d.). As a result packets may overtake each other in the network and they may arrive out of sequence to the receiver. This assumption was used in [10, 20]; it is most appropriate when a highly dynamic routing policy is used in the network. Network delay of a packet is assumed to be either exponentially distributed as given in [5, 20]:

$$p_N(t_N) = \lambda e^{-\lambda t_N} \quad (2.4)$$

or geometrically distributed in discrete time:

$$p_N(\tau_N) = (1-q)q^{\tau_N-1}, \quad \tau_N = 1, 2, \dots \quad (2.5)$$

For $q = e^{-\lambda p}$, the two distributions are equivalent.

2.1.3 Receiver Protocol

The receiver implements one of the protocols stated in chapter 1. Except for the analysis of the chapter 5, an infinite buffer is assumed. Delivery time of a packet refers to the delivery time of its first bit to the DA converter, and a packet is considered to be in the buffer until its last bit is delivered.

2.2 Analysis and Notation

In the next two chapters a probability generating function approach is used in determining the various distributions. In chapter 5 a more complicated Markov chain analysis is applied.

For any packet we define the following Bernoulli random variable:

$$k_j(n) = \begin{cases} 1 & \text{if } j\text{'th packet has arrived} \\ & \text{buffer before or at } n. \\ 0 & \text{otherwise} \end{cases} \quad (2.6)$$

In the different parts of this work, the above definition will be extended according to the need.

The symbols $t_{Nj}(\tau_{Nj})$, $t_{Sj}(\tau_{Sj})$, $t_{Dj}(\tau_{Dj})$ will refer to the network, system and total delay of the j 'th packet in seconds (in slots) respectively.

Chapter 3

DISTRIBUTION OF THE NUMBER OF PACKETS IN THE BUFFER

In this chapter we determine the buffer distributions for protocols FD, CD, and TD introduced in chapter 1 assuming a single talkspurt, and infinite buffer.

In the first section of this chapter the distribution of the number of arrivals to the buffer is determined. Then in the following sections the distribution and expected number of packets in the buffer are derived for each of the protocols. In the last section, an approximate analysis of a PVC network with protocol TD is given.

3.1 PGF of the Arrivals to the Buffer

The arrival process to the buffer is independent of the protocol applied at the receiver. Assuming a talkspurt of $(\alpha+1)$ packets, the j 'th packet of the talkspurt arrives to the network at time $n = j - 1$, for $j = 1, 2, \dots, \alpha+1$, and spends τ_{Nj} slots in the network before arriving at the receiver. Network delays are i.i.d. according to the geometric distribution (2.5). At time n the status of packet j is described by the Bernoulli random variable, $k_j(n)$, defined in the equation (2.6). Since j 'th packet arrives to the network at time $(j-1)$, the probability distribution of $k_j(n)$ is given by:

$$P\{k_j(n)=1\} = P\{\tau_{Nj} \leq n-j+1\} = \sum_{\tau_{Nj}=1}^{n-j+1} (1-q)q^{\tau_{Nj}-1} = 1 - q^{n-j+1} \quad n \geq j - 1 \quad (3.1)$$

$$P\{k_j(n)=0\} = P\{\tau_{Nj} > n-j+1\} = q^{n-j+1}, \quad n \geq j - 1 \quad (3.2)$$

Let us define the PFG of $k_j(n)$ as $K_j(z, n)$:

$$K_j(z, n) = P\{k_j(n) = 1\}z + P\{k_j(n) = 0\}$$

$$K_j(z, n) = (1 - q^{n-j+1})z + q^{n-j+1} \quad (3.3)$$

Defining the number of packets satisfying $j_1 \leq j \leq j_2$, and arriving to the receiver by time n as $k(n, j_1, j_2)$, we see that:

$$k(n, j_1, j_2) = \sum_{j=j_1}^{j_2} k_j(n) \quad (3.4)$$

Clearly from the independence of $k_j(n)$ for each j , the PGF of the distribution of the sum is the product of the individual PGF's, $K_j(z, n)$ in (3.3). Denoting the PGF of the distribution of the sum as $K(z, n, j_1, j_2)$, it is therefore given by:

$$K(z, n, j_1, j_2) = \prod_{j=j_1}^{j_2} [(1 - q^{n-j+1})z + q^{n-j+1}] \quad (3.5)$$

To simplify the latter notation, we transform j to the index $r = n - j + 1$ in the above equation, and use the modified Binomial PFG defined as:

$$B(z, q, r_{\min}, r_{\max}) = \prod_{r=r_{\min}}^{r_{\max}} [(1 - q^r)z + q^r] \quad (3.6)$$

$$K(z, n, j_1, j_2) = B(z, q, n-j_2+1, n-j_1+1) . \quad (3.7)$$

The above equation gives the PGF of the distribution of the number of arrivals to the buffer from packets in the range $j_1 \leq j \leq j_2$ at time n .

Depending on the receiver protocol some of the packets which have arrived by time n will have been delivered to the listener and others

will be held. Time n is the arrival time of $(n+1)$ st packet to the network, given that such a packet exists ($n+1 \leq \alpha+1$ = number of packets in the talkspurt). But the $(n+1)$ st packet cannot be in the buffer at time n because of nonzero network delay (see equation 2.5). Thus the packet with the highest sequence number which may be in the buffer is $j_2 = \min(n, \alpha+1)$. Those which are held at time n are delimited by $j_1 \leq j \leq j_2$ in equation (3.7). For each protocol, j_1 can be determined as functions of n and talkspurt length, giving the conditional distribution for the buffer contents as $K(z, n | \alpha)$:

$$K(z, n | \alpha) = K_j(z, n, j_1(\alpha), j_2(\alpha)) . \quad (3.8)$$

3.2 The Buffer Distribution for the Protocol FD

First we briefly describe the protocol FD. At the beginning of each talkspurt, FD protocol chooses a fixed system time (τ_S = network delay + receiver delay) for all packets of that talkspurt. One possible way of choosing τ_S is as the mean network delay of the previous talkspurt plus a constant; another is simply to take a single constant for all talkspurts. The life time of each packet is τ_S ; τ_S slots after its arrival to the network a packet is either delivered to the DA converter or from the receiver's point of view vanishes from the network. Let us define the delivery time of the j 'th packet to the DA converter as $\tau_{d,j}$ (or $t_{d,j}$ in sec.), then we have:

$$\tau_{d,j} = j + \tau_S - 1 . \quad (3.9)$$

Solving for j in the above equation with $\tau_{d,j} = n$ gives the sequence number of the packet to be delivered to DA converter at time n as:

$$j = n - \tau_S + 1 \quad (3.10)$$

Thus the packet with smallest sequence number which may be in the buffer at time n is $j_1 = \max(n - \tau_S + 1, 1)$.

Since the life-time of each packet is τ_S slots, only the most recent τ_S packets may be in the buffer. This protocol prevents occurrence of the large gaps in the talkspurt due to late arrivals by choosing τ_S adequately large to absorb the random 'jitter' in reception times (see Figure 3.1).

3.2.1 PGF of the Number of Packets in the Buffer

For a talkspurt length of $(\alpha+1)$ packets, at any time between $1 \leq n \leq \alpha + \tau_S$, the only packets which may be in the buffer are between $j_1 \leq j \leq j_2$, with j_1 and j_2 given by:

$$\begin{aligned} j_1 &= \max(n - \tau_S + 1, 1) \\ j_2 &= \min(n, \alpha + 1) \end{aligned} \quad (3.11)$$

For $n > \alpha + \tau_S$ there are no packets in the buffer from this talkspurt. Thus the PGF of the distribution of the number of packets in the buffer, conditional on the talkspurt length (α) is given from (3.7):

$$K(z, n | \alpha) = \begin{cases} B(z, q, \max(1, n - \alpha), \min(n, \tau_S)), & 1 \leq n \leq \alpha + \tau_S \\ 1, & \text{otherwise.} \end{cases} \quad (3.12)$$

The PGF in the above equation is in the form of a polynomial therefore its inversion is straightforward. The unconditional PGF can be determined by unconditioning (3.12) wrt α using (2.1):

$$K(z, n) = \begin{cases} \sum_{\alpha=0}^{\infty} (1-\rho)\rho^{\alpha} B(z, q, \max(1, n - \alpha), \min(n, \tau_S)) & 1 \leq n \leq \alpha + \tau_S \\ 1 & \text{otherwise} \end{cases} \quad (3.13)$$

which must be inverted numerically.

Figure 3.1 Delivery Pattern of the FD Protocol

3.2.2 Expected Number of Packets in the Buffer

Next the conditional and unconditional expected number of packets in the buffer are determined:

$$E\{k(n)|\alpha\} = \left. \frac{dK(z,n|\alpha)}{dz} \right|_{z=1} \quad (3.14)$$

Substituting from (3.12) into the above equation results in:

$$E\{k(n)|\alpha\} = 1 + \min(n, \tau_S) + \frac{q^{\min(n, \tau_S) + 1}}{1 - q} - \max(1, n - \alpha) - \frac{q^{\max(1, n - \alpha)}}{1 - q}, \quad n \leq \alpha + \tau_S \quad (3.15)$$

Unconditional expected number of packets in the buffer is given by:

$$E\{k(n)\} = \sum_{\alpha=0}^{\infty} (1-\rho)\rho^{\alpha} E\{k(n)|\alpha\} \quad (3.16)$$

The above unconditioning has been carried out in the appendix I, and the result is given below from (A1.5):

$$E\{k(n)\} = \frac{\rho^{n+1}(1-q)}{(q-\rho)(1-\rho)} + \rho^{\max(0, n-\tau_S)} \left[\frac{1}{1-\rho} - \frac{q^{\min(n, \tau_S)+1}}{q-\rho} \right] \quad (3.17)$$

3.3 The Buffer Distribution for the Protocol CD

This protocol functions as protocol FD, except CD chooses τ_S for each talkspurt as the delay of the initial packet, plus a constant μ . τ_S is thus a random variable constructed according to

$$\tau_S = \tau_N + \mu \quad (3.18)$$

The receiver then delivers all packets in the talkspurt with the same delay τ_S . Thus τ_S adapts to the network delay, and from (2.5) it has the distribution:

$$p_S(\tau_S) = (1-q)q^{\tau_S-\mu-1}, \quad \tau_S = \mu+1, \dots, \infty \quad (3.19)$$

Results conditioned on both τ_S and α are the same as for the FD protocol. Results conditioned only on α are given by:

$$K(z, n | \alpha) = \sum_{\tau_S=\mu+1}^{\infty} (1-q)q^{\tau_S-\mu-1} K(z, n | \alpha, \tau_S) \quad (3.20)$$

$$E\{k(n) | \alpha\} = \sum_{\tau_S=\mu+1}^{\infty} (1-q)q^{\tau_S-\mu-1} E\{k(n) | \alpha, \tau_S\} \quad (3.21)$$

where $K(z, n | \alpha, \tau_S)$, $E\{k(n) | \alpha, \tau_S\}$ are in (3.13) and (3.15). The first unconditioning above requires numerical inversion; the second one has been carried out in appendix 2 and is given below from (A2.3):

$$E\{k(n) | \alpha\} = \begin{cases} \psi_A & n \leq \mu \\ \psi_A + \left[\frac{\max(\mu+1, n-\alpha)}{q} + \frac{1}{1-q} \right] q^{\max(\mu+1, n-\alpha) - \mu} - \left(n + \frac{1}{1-q} \right) q^{n-\mu} - \frac{q^{2n-\mu+2}}{1-q^2} & n \geq \mu + 1 \end{cases} \quad (3.22)$$

where

$$\psi_A = \left[1 - \max(1, n-\alpha) - \frac{q^{\max(1, n-\alpha)}}{1-q} \right] q^{\max(\mu+1, n-\alpha) - \mu-1} \\ + \left[n + \frac{q^{n+1}}{1-q} \right] q^{\max(0, n-\mu)} .$$

The unconditioning of (3.22) wrt α , has also been carried out in the appendix 2, and is given below from (A2.5):

$$E\{k(n)\} = \begin{cases} \psi_B & n \leq \mu+1 \\ \psi_B - \frac{(1-q)}{(1-\rho)(q-\rho)} (\rho^{n-\mu} - \rho q^{n-\mu-1}) & \\ - \frac{(1-q)}{(q-\rho)(\rho-q^2)} (q^{\mu+2} \rho^{n-\mu} - \rho q^{2n-\mu}) , & n > \mu+1 \end{cases} \quad (3.23)$$

where

$$\psi_B = \frac{\rho^{n+1}(1-q)}{(q-\rho)(1-\rho)} + \left[\frac{1}{1-\rho} - \frac{q^{n+1}}{q-\rho} \right] q^{\max(\mu+1, n) - \mu-1} .$$

3.4 The Buffer Distribution for the Protocol TD

This protocol waits for late arriving packets and re-synchronizes its output stream after each wait. The first packet is delayed by the fixed amount of time μ at the receiver to reduce the gaps occurring during the talkspurt. The delivery time of the j 'th packet is given by:

$$\tau_{d,j} = \begin{cases} \tau_{N1} + \mu & j = 1 \\ \max(j-1 + \tau_{Nj}, \tau_{d,j-1} + 1) & j > 1 \end{cases} \quad (3.24)$$

This delivery pattern is shown in Figure 3.2.

Unlike the situation in the previous two protocols the delivery time of a packet is not known at the beginning of a talkspurt. The distribution of the number of packets delivered is rather important for TD.

3.4.1 Distribution of the Number of Packets Delivered to the DA Converter

Define $x(n)$ as the cumulative number packets delivered from the receiver to the DA converter at time n and $p(x(n))$ as its distribution. Let us condition on talkspurt length and divide the time axis into two intervals:

Interval (i) : $n \leq \mu$.

In this interval there are no deliveries to the DA converter:

$$p(x(n) = i | \alpha) = \begin{cases} 1 & i = 0 \\ 0 & i > 0 \end{cases} \quad (3.25)$$

Interval (ii) : $n \geq \mu+1$

In this interval there may be deliveries to the DA converter.

Depending on $x(n)$ two cases are identified:

Figure 3.2 . Delivery Pattern of the TD Protocol

$$1) \quad x(n) = 0$$

$$p(x(n) = 0 | \alpha) = q^{n-\mu-1} \quad (3.26)$$

$$2) \quad 0 < x(n) \leq \min(n-\mu, \alpha+1)$$

Given that there has been $x(n) = i$ deliveries, delivery time of the j 'th packet satisfies the following:

$$\tau_{d,j} | x(n)=i \leq n + j - 1 - i, \quad j \leq i. \quad (3.27)$$

Since the arrival time of the j 'th packet to the network is $(j-1)$:

$$\tau_{Nj} | x(n)=i \leq \begin{cases} n - i - \mu & j = 1 \\ n - i & 1 < j \leq i \end{cases} \quad (3.28)$$

From the above equation:

$$p(x(n) \geq i | \alpha) = P\{\tau_{N1} \leq n-i-\mu, \bigcap_{j=2}^i (\tau_{Nj} \leq n-i)\} \quad (3.29)$$

Substituting from (2.5) gives:

$$p(x(n) \geq i | \alpha) = (1-q)^{n-i-\mu} (1-q)^{n-i} i^{-1} \quad (3.30)$$

The last equation gives:

$$p(x(n)=i | \alpha) = \begin{cases} (1-q)^{n-i-\mu} (1-q)^{n-i} i^{-1} - (1-q)^{n-i-\mu-1} (1-q)^{n-i-1} i^{-1}, & 0 < i \leq \min(n-\mu-1, \alpha) \\ (1-q)^{n-i-\mu} (1-q)^{n-i} i^{-1}, & i = \min(n-\mu, \alpha+1) \end{cases} \quad (3.31)$$

From (3.25-3.26) and (3.31), the distribution of the number of packets delivered to DA is given by:

$$p(x(n)=i|\alpha) = \begin{cases} q^{\max(0, n-\mu-1)} & i = 0 \quad n \geq 0 \\ (1-q)^{n-i-\mu}(1-q)^{n-i}i-1 - (1-q)^{n-i-\mu-1}(1-q)^{n-i-1}i & 0 < i \leq \min(n-\mu-1, \alpha) \quad n \geq \mu+1 \\ (1-q)^{n-i-\mu}(1-q)^{n-i}i-1 & i = \min(n-\mu, \alpha+1) \quad n \geq \mu+1 \end{cases} \quad (3.32)$$

Define α_R as the time it takes to deliver all the packets of a talkspurt beginning with the delivery time of the first packet. From (3.30) cumulative distribution function of (CDF) of α_R is given by:

$$P\{\alpha_R \leq u\} = (1-q)^{u-\alpha+1} (1-q)^{u+\mu-\alpha+1} \alpha \quad u = \alpha+1, \alpha+2, \dots, +\infty \quad (3.33)$$

3.4.2 PGF of the Number of the Packets in the Buffer

First this PGF is to be derived conditioned on the talkspurt length and the number of packets delivered to the DA, giving $(K(z, n|\alpha, x(n)))$. Clearly $K(z, n|\alpha, x(n) = \alpha+1) = 1$, for the empty buffer after the last packet has been delivered. Because the first packet is treated differently, two cases are identified:

Case (1) $x(n) = 0$ Packets Delivered

With $x(n) = 0$ the first packet may still be in the network ($k_1(n) = 0$) or it may have arrived and be waiting in the buffer for its μt_p - sec delay ($k_1(n) = 1$). Introducing the condition that $k_1(n) = 0$ or 1 gives the PGF in this form:

$$K(z, n | \alpha, x(n) = 0) = K(z, n | \alpha, x(n) = 0, k_1(n) = 1)P\{k_1(n) = 1 | x(n) = 0\} \\ + K(z, n | \alpha, x(n) = 0, k_1(n) = 0)P\{k_1(n) = 0 | x(n) = 0\}. \quad (3.34)$$

Bayes rule is applied to determine the conditional probabilities:

$$P\{k_1(n)=0|x(n)=0\} = \frac{P\{k_1(n)=0, x(n)=0\}}{P\{x(n)=0\}} = \frac{q^n}{q^{\max(0, n-\mu-1)}}$$

$$P\{k_1(n)=0|x(n)=0\} = q^{\min(n, \mu+1)} \quad (3.35)$$

$$P\{k_1(n)=1|x(n)=0\} = 1 - q^{\min(n, \mu+1)} \quad (3.36)$$

The conditional PGF's in (3.34) are found from first principles.

The packets which could be in the buffer (aside from number 1 which is there only if $k_1(n) = 1$) have number j in the range $2 \leq j \leq \min(n, \alpha+1)$.

From (3.7):

$$K(z, n | \alpha, x(n) = 0, k_1(n) = 0) = B(z, q, \max(1, n-\alpha), n-1) \quad (3.37)$$

$$K(z, n | \alpha, x(n) = 0, k_1(n) = 1) = zB(z, q, \max(1, n-\alpha), n-1)$$

Substituting (3.35-3.37) into (3.34) and simplifying gives:

$$K(z, n | \alpha, x(n) = 0) = [q^{\min(n, \mu+1)} + (1-q^{\min(n, \mu+1)})z]B(z, q, \max(1, n-\alpha), n-1) \quad (3.38)$$

Case (2) $x(n) > 0$ Packets Delivered

The packets which may be in the buffer satisfy $x(n)+1 \leq j \leq \min(n, \alpha+1)$, so using (3.7) again:

$$K(z, n | \alpha, x(n) > 0) = B(z, q, \max(1, n - \alpha), n - x(n)) \quad (3.39)$$

This completes the distribution conditional on $x(n)$ and α .
The condition on $x(n)$ is removed by using (3.32) as follows:

$$K(z, n | \alpha) = \sum_{x(n)=0}^{\alpha+1} K(z, n | \alpha, x(n)) p(x(n) | \alpha) \quad (3.40)$$

$$K(z, n | \alpha) = \begin{cases} K_a & 0 \leq n \leq \mu \\ K_a + K_b & \mu+1 \leq n \leq \alpha+\mu \\ K_a + K_b + K_c & \alpha+\mu+1 \leq n \end{cases} \quad (3.41)$$

where

$$K_a = q^{\max(0, n-\mu-1)} [q^{\min(n, \mu+1)} + (1-q)^{\min(n, \mu+1)}] z B(z, q, \max(1, n-\alpha), n-1)$$

$$K_b = \sum_{i=1}^{\min(n-\mu-1, \alpha)} [(1-q)^{n-i-\mu} (1-q)^{n-i} i^{-1} - (1-q)^{n-i-\mu-1} (1-q)^{n-i-1} i] B(z, q, \max(1, n-\alpha), n-i)$$

$$K_c = (1-q)^{n-\alpha-1-\mu} (1-q)^{n-\alpha-1} \alpha$$

Equation (3.41) gives the PGF of the distribution of the number of packets in the buffer for the TD protocol conditioned on the talkspurt length (α). This result is in the form of a polynomial, therefore inversion is straightforward. The condition on the talkspurt length may also be removed by using (2.1):

$$K(z, n) = \sum_{\alpha=0}^{\infty} (1-\rho) \rho^{\alpha} K(z, n | \alpha) \quad (3.42)$$

which requires numerical inversion.

3.4.3 Expected Number of Packets in the Buffer

Differentiating (3.41) gives the conditional expected number of packets in the buffer at time n :

$$E\{k(n)|\alpha\} = \begin{cases} \psi_c & n \leq \mu \\ \psi_c + \sum_{i=1}^{\min(n-\mu-1, \alpha)} \sum_{r=\max(1, n-\alpha)}^{n-i} (1-q^r) [(1-q^{n-i-\mu})(1-q^{n-i})^{i-1} - (1-q^{n-i-\mu-1})(1-q^{n-i-1})^i] & n \geq \mu+1 \end{cases} \quad (3.43)$$

where

$$\psi_c = q^{\max(0, n-\mu-1)} [1 - q^{\min(n, \mu+1)}] + \sum_{r=\max(1, n-\alpha)}^{n-1} (1-q^r)$$

The unconditional result is found numerically from:

$$E\{k(n)\} = \sum_{\alpha=0}^{\infty} (1-\rho)\rho^\alpha E\{k(n)|\alpha\} \quad (3.44)$$

3.5 The Expected Number of Packets in the Buffer for Protocol TD with PVC Network

For packet-switched networks using virtual circuits for voice, independence of network delays is an inappropriate model because packets do not pass each other in the network, and thus cannot get out of order. In this section, an approximate solution to this problem is given for the TD protocol. In a packet-switched network using dynamic routing the receiver buffer may in imagination be divided into two parts:

i). Reordering Buffer:

This part reorders the packets arriving from the network, and gives them to the delivery buffer. This is called the reordering delay of a packet. The same problem has also been analyzed in the context of distributed database management in [20] with Poisson arrivals and exponential network delays.

ii) Delivery Buffer:

This part delivers the packets arriving from the reordering buffer to the DA converter. It delays the first packet of the talkspurt by μt_p and keeps the minimum time difference between the deliveries of the two consecutive packet at t_p . A block diagram of the system with this imaginary division of the buffer is shown in Figure 3.3a. $r(n)$, $d(n)$, and $k(n)$ are the number of packets in the reordering, delivery, and receiver buffers respectively, with:

$$k(n) = r(n) + d(n) \quad (3.45a)$$

In a network with virtual circuits, packets are kept in order, with attendant delays in queueing for packets ahead. Thus the receiver buffer needsto carry only the delivery function (see Figure 3.3b). We will assume that the arrival sequence from the network with a virtual circuit is the same as that from the reordering buffer in a network with dynamic routing policy.

Figure 3.3a. Packet Voice Receiver for a Communications Network with Dynamic Routing

Figure 3.3b. Packet Voice Receiver for a Communications Network with Fixed Routing.

For a PVC network, we only need to build the delivery buffer. From equation (3.45a):

$$E\{d(n)\} = E\{k(n)\} - E\{r(n)\} \quad (3.45b)$$

The first term on the right side is given by (3.43), and next we determine the second term. Let $R(z,n)$ be the PFG of the number of packets in the reordering buffer and $x(r,n)$ be the number of deliveries from this buffer until n . The distribution of $x(r,n)$ is given by:

$$p(x(r,n)=i|\alpha) = \begin{cases} q^n & i = 0 \\ \prod_{j=1}^i (1-q^{n-j+1}) q^{n-i} & 0 < i \leq \min(n-1, \alpha) \\ \prod_{j=1}^i (1-q^{n-j+1}) & i = \min(n, \alpha+1) \end{cases} \quad (3.46)$$

Following the derivation of (3.40) in the previous section, first conditional PGF is determined:

$R(z,n|\alpha, x(r,n) = \alpha+1) = 1$ for the empty buffer after the last packet has been delivered. For $x(r,n) \leq \alpha$, the packets which may be in the buffer satisfy $x(r,n) + 1 \leq j \leq \min(n, \alpha+1)$, and using (3.7) gives:

$$R(z,n|\alpha, x(r,n)) = \begin{cases} B(z, q, \max(1, n-\alpha), n-x(r,n)) & x(r,n) \leq \alpha \\ 1 & x(r,n) = \alpha+1 \end{cases} \quad (3.47)$$

Using (3.46) condition on $x(r,n)$ may be removed to give:

$$R(z,n|\alpha) = q^n B(z,q,\max(1,n-\alpha),n) + \sum_{i=1}^{\min(n-1,\alpha)} B(z,q,\max(1,n-\alpha),n-i) \prod_{j=1}^i (1-q^{n-j+1}) q^{n-i} + \prod_{j=1}^{\min(n,\alpha+1)} (1-q^{n-j+1}) \quad (3.48)$$

Finally expected number of packets in the reordering buffer is given by:

$$E\{r(n)|\alpha\} = q^n \sum_{r=\max(1,n-\alpha)}^n (1-q^r) + \sum_{i=1}^{\min(n-1,\alpha)} \sum_{r=\max(1,n-\alpha)}^{n-i} (1-q^r) \prod_{j=1}^i (1-q^{n-j+1}) q^{n-i} \quad (3.49)$$

Substituting (3.43) and (3.49) into (3.45b) give the expected number of packets in the buffer for protocol TD with PVC network.

3.6 Results and Discussion

In this section we present some numerical work on the results of this chapter. Results conditioned on the talkspurt length assume a length of 40 packets; for the unconditional results this has been taken as the mean talkspurt length. Mean network delay of a packet has been chosen to be 2 slots. Though results are plotted in dimensionless slots of time, if the assumed talkspurt length (40 packets) is taken to be equal to the measured mean value (1.32 sec), it corresponds to 33 msec of packet duration, and 66 msec mean network delay. This value

of the network delay is reasonable for voice packets travelling few hops [6].

Figure 3.4 shows the conditional and unconditional expected number of packets in the buffer for the protocols FD (equations 3.15, 3.17) and CD (equations 3.22, 3.23) respectively. Results are plotted against time for a variety of values of τ_S or μ , conditional plots are given by solid lines, while unconditional plots are broken lines. For $n \gg 1$, we see that the mean number of packets is $\tau_S - 1$; this is $\tau_S - E\{\tau_N\} + 1$, where the 1 represents a packet in the process of being delivered. The unconditional curve is shaped primarily by the talkspurt distribution. The conditional curve is constant during the central section. The largest buffer requirement occurs in the central section ($\tau_S \leq n \leq \alpha$) on the conditional curve, and the peak value is $n = E\{\tau_S\}$ on the unconditional curves.

Figure 3.5 gives the same results for the TD protocol (equations 3.43, 3.44) for various values of μ . Clearly the curves all increase monotonically up to a worst case at $n = \alpha + 1$, and they appear to be converging towards each other. Compared to CD with the same μ , or FD with $\tau_S = \mu + E\{\tau_N\}$, the peak demand by the TD protocol is consistently larger. Figure 3.6 shows the conditional expected number of packets in the buffer for TD with PVC approach (equation 3.45b). Because of the ordered arrivals, expected buffer size is reduced.

Figure 3.7 shows the expected number of packets to be delivered by time n for the TD (equation 3.32) μ as a parameter. As μ

increases the slope of the curves approach to one packet per slot, which results in significant reduction of gaps in the talkspurt.

Figure 3.8 shows the conditional buffer distributions at the worst-case time point, for each protocol, and for various values of τ_S or μ (equations 3.12, 3.20, 3.41 for FD, CD, TD respectively). The distribution tails are such that the probability of exceeding twice the mean is well under 1% in all these cases. Thus the time-varying mean can be used to estimate the buffer requirement with some confidence.

FIGURE 3.4

Mean Buffer Lengths for FD and CD Protocols against Time (Network Delay $E\{\tau_N\}=2$).

Fixed Talkspurt $\alpha = 39$ Random $\alpha, E\{\alpha\} = 39$

FD ————— FD — — — — —

CD + + + + + CD + - + - + - +

τ_S = System delay of a packet in slots

μ = Buffer delay of the first packet in slots

FIGURE 3.5
Mean Buffer Lengths for TD Protocol ($E\{\tau_N\} = 2$).

Fixed $\alpha = 39$

Random α ; $E\{\alpha\} = 39$

μ = Buffer delay of the first packet in slots

Figure 3.6 Mean Buffer Length for TD Protocol with
 PVC Approach
 Talkspurt Length $\alpha = 39$
 Network Delay $E\{\tau_N\} = 2$

μ = Buffer delay of the first packet in slots

Figure 3.7. Mean Number of Packets Delivered to the DA Converter for TD Protocol
Talkspurt Length $\alpha = 39$
Network Delay, $E\{\tau_N\} = 2$

μ = Buffer delay of the first packet in slots

FIGURE 3.8
Probability Distribution of k packets in the Receiver Buffer, for FD, CD and TD

Protocols ($E\{\tau_N\} = 2$, various τ_S and μ).

— FD; --- CD; - · - · - TD

τ_S = System delay of a packet in slots

μ = Buffer delay of the first packet in slots

Chapter 4

DELAY AND LOST PACKET ANALYSIS

In this chapter we derive the distribution of receiver delay for the packets which are not discarded and the distribution of the number of packets discarded for each of the protocols for a single talkspurt. In the last section a receiver design example is also outlined.

4.1 Protocol FD

Define the distribution of the receiver delay as $p_R(\tau_R)$. The network delay (τ_N) and receiver delay (τ_R) and system delay (τ_S) of a packet are related by:

$$\tau_R = \tau_S - \tau_N \quad \tau_N \leq \tau_S \quad (4.1)$$

$$\begin{aligned} p_R(\tau_R) &= P\{\tau_N = \tau_S - \tau_R | \tau_N \leq \tau_S\} \\ &= \frac{(1-q)q^{\tau_S - \tau_R - 1}}{1 - q^{\tau_S}} \quad 0 \leq \tau_R \leq \tau_S - 1 \end{aligned} \quad (4.2)$$

Thus expected value of the receiver delay in slots is given by:

$$E\{\tau_R\} = \frac{\tau_S}{1 - q^{\tau_S}} - \frac{1}{1 - q} \quad (4.3)$$

The probability distribution of the number d of packets discarded is determined next. Because the network delays are independent each packet is discarded with probability q^{τ_S} and d has a binomial distribution with the number of trials equal to the talkspurt length in packets.

Defining $D(z)$ as the unconditional PGF of the distribution of d , we find

$$D(z) = \frac{(1-q)^{\tau_S} + zq^{\tau_S}(1-\rho)}{1-\rho(1-q)^{\tau_S} + zq^{\tau_S}} \quad (4.4)$$

with expected number of discarded packets per talkspurt of:

$$E\{d\} = \frac{q^{\tau_S}}{1-\rho} \quad (4.5a)$$

Defining packet loss rate (d^*) as fraction of packets lost, then

$$d^* = (1-\rho)E\{d\} = q^{\tau_S} \quad (4.5b)$$

4.2 Protocol CD

Results for CD are obtained by unconditioning the results above (for FD) wrt τ_S using $p_S(\tau_S)$ in (3.19) as follows:

$$p_R(\tau_R) = \sum_{\tau_S=\max(\tau_R+1, \mu+1)}^{\infty} \frac{(1-q)q^{\tau_S-\tau_R-1}}{1-q^{\tau_S}} (1-q)q^{\tau_S-\mu-1}; \quad \tau_R \geq 0 \quad (4.6)$$

$$E\{d\} = \frac{q^{\mu+1}}{(1-\rho)(1+q)} \quad (4.7a)$$

$$d^* = \frac{q^{\mu+1}}{1+q} \quad (4.7b)$$

4.3 Protocol TD

Delay analysis of this protocol has been given by Barberis and Pazzaglia [10]. Following them continuous time variable will be used in this section, thus network delay is assumed to be exponentially distributed as in equation (2.4). In [10], probability density function (pdf) of the system delay of a packet (network plus receiver delay) has been derived with arbitrary network delay distribution. Let us define the pdf of the system delay averaged over all packets in a talkspurt as $p_S(t_S)$, then for exponential network delay from [10]:

$$p_S(t_S) = \frac{1 - \rho}{1 - \rho(1 - e^{-\lambda t_S})} \lambda e^{-\lambda(t_S - \mu t_p)} + \frac{\rho(1 - \rho)\lambda e^{-\lambda t_S}}{[1 - \rho(1 - e^{-\lambda t_S})]^2} \quad t_S \geq \mu t_p \quad (4.8)$$

In a later work Barberis [11] has also derived the pdf of the average receiver delay of a packet, given by:

$$p_R(t_R) = (\overline{F(0)} + 1 - \rho)\delta(t_R) + \overline{f(t_R)} \quad (4.9)$$

where $\delta(t_R)$ is an impulse function

$$\overline{f(t_R)} = \rho \int_{-\infty}^{\infty} p_N(t_S - t_R) p_S(t_S) dt_S \quad (4.10)$$

$$\overline{F(0)} = \int_{-\infty}^0 \overline{f(t_R)} dt_R \quad (4.11)$$

This pdf has been specialized* for exponential network delay in appendix 3, a result not found in [11], and from (A3.3):

$$p_R(t_R) = (1-\rho)\left(1 + \frac{1-\rho}{\rho} e^{\lambda \mu t_p}\right) \left\{ \ln\left(1 + \frac{\rho}{1-\rho} e^{-\lambda \mu t_p}\right) \delta(t_R) + \lambda e^{\lambda t_R} \left[\ln\left(1 + \frac{\rho}{1-\rho} e^{-\lambda \max(\mu t_p, t_R)}\right) - \frac{\rho e^{-\lambda \max(\mu t_p, t_R)}}{1 - \rho + \rho e^{-\lambda \max(\mu t_p, t_R)}} \right] \right\} \quad (4.12)$$

4.4 Numerical Results and Discussion

As in section 3.6 we assume a mean talkspurt length of 40 packets, and mean network delay of 2 slots. Results for TD have been calculated using exponential network delay with equal mean.

Figure 4.1 shows the pdf of the receiver delay for various τ_S and μ from 4.2, 4.6, and 4.12 (TD result has been plotted at $t_R = n$). It is easily seen that CD distribution has greater variance than FD and TD has much greater variance again. In Figure 4.2 the mean delays are plotted against τ_S (or μ) and TD gives much larger delays for small μ , converging to the FD and CD delay only when μ and τ_S are quite large.

Figure 4.2 also shows the percentage of packets discarded for the FD and CD protocols against τ_S and μ , with CD doing somewhat worse. Clearly to obtain acceptable loss rates (below 5%), μ or $\tau_S - E\{\tau_N\}$ must be at least as large as the mean network delay.

CD was introduced as an improvement over FD but found to have slightly poorer performance. This is because FD is assumed to have perfect knowledge of mean network delay. In practice, FD will use an estimate of the mean network delay of the previous talkspurt in the present talkspurt. As a result of the errors in estimation as well as changing channel conditions, the estimate will differ from the instantaneous mean network delay. Thus, in practice, FD will probably have poorer performance than CD.

4.5 A Buffer Design Calculation

Buffer design example illustrates the use of the results derived so far. Design involves determining τ_S , μ , and the required buffer size given the mean network delay and the packet loss rate (for FD and CD). In the design, we shall use results conditional on the talkspurt length since resulting expected number of packets in the buffer is larger than the unconditional case (see Figures 3.4 and 3.5). Let us assume that the mean network delay is 2 slots, and packet loss rate under 2% (for FD and CD).

For protocol FD, from Figure 4.2, system packet delay τ_S must be 5 slots to give the desired loss rate, which will give a total delay of $\tau_D = 6$ slots. From Figure 4.2 again we find an expected receiver delay of 4 slots. From Figure 3.4 the mean buffer length will be about 4 packets during talkspurts, and from the distribution in Figure 3.8 a buffer for 7 packets should give negligible packet losses due to overflow.

Design with CD follows the steps of FD and is neglected here. For protocol TD there is not a mathematical measure of the gaps occurring during the talkspurt such as packet loss rate in FD and CD. From Figure 3.7, it is seen that expected time to deliver all packets of a 40 packet length talkspurt is 50, 48 and 44 slots for $\mu = 0, 2, 4$ respectively. We shall assume that $\mu = 4$ gives an acceptable level of speech quality without introducing extra unnecessary delay. Then the mean receiver delay could be about 13 slots (from Figure 4.2), mean buffer length 6 packets (from Figure 3.5), and a buffer length of 12 packets (from Figure 3.8) would make losses negligible.

FIGURE 4.1

Probability Distribution of the Receiver Delay $\tau_R(E\{\tau_N\} = 2$, various τ_S and μ).

————— FD; - - - - - CD; - - - - - TD

τ_S = System delay of a packet in slots

μ = Buffer delay of the first packet for TD in slots

FIGURE 4.2

Percentage of Packets Discarded, and Mean Receiver Buffer Delay, vs τ_S and μ ($E\{\tau_N\} = 2$).

———— FD; - - - - - CD; TD

τ_S = System delay of a packet in slots

μ = Buffer delay of the first packet for CD, TD protocols

Chapter 5

FINITE BUFFER ANALYSIS

In the present chapter, the distribution of the number of packets in a finite buffer is determined for the protocol FD. We also give the expected number of packets in the buffer and the percentage of packets lost due to overflow. The results assume a constant talkspurt length, of $(\alpha+1)$ packets.

5.1 Markov Chain Analysis

Under the finite buffer assumption and the FD protocol, packets lost due to overflow are treated like late packets at their delivery time to the DA converter. In this problem the PGF approach used in chapter 3 is impossible due to the dependence of packets on each other while they pass through the buffer. This dependence may be taken into account by a transient Markov chain analysis.

As before, let us describe the status of the j 'th packet in the system by the random variable $k_j(n)$, with a new state added for the possibility that a packet may be discarded due to overflow:

$$k_j(n) = \begin{cases} 0 & j\text{'th packet is in the network at time } n. \\ 1 & j\text{'th packet is in the buffer at time } n. \\ 2 & j\text{'th packet has been discarded due to overflow before time } n. \end{cases} \quad (5.1)$$

All packets in the system follow an identical state transition diagram as shown in Figure 5.1. It may be seen that, given a packet is in state 0, the probability that it stays in or leaves this state in the next slot is independent of everything else. On the other hand, given that a packet leaves state 0, the probability that it goes to state 1 or 2 also depend on the other packets. As will be shown the former probabilities are sufficient to determine the distribution of the number of packets in the buffer.

Under the FD protocol, the life of each packet is τ_S slots, thus the state of the system at any time is determined by the state of the most recent τ_S arrivals to the system. We define a system Markov chain whose states are vectors of τ_S components, with packet states taken as the components of this vector. Since there are three packet states the total number of system states is $M_S = 3^{\tau_S}$. The first component of the system state is assumed to be the state of the most recent arrival, while the τ_S 'st component is the state of the oldest arrival to the system. Let i and l be the system states at time n and $(n+1)$, then:

$$i = [i_1, i_2, \dots, i_{\tau_S}] \quad (5.2)$$

$$l = [l_1, l_2, \dots, l_{\tau_S}]$$

or in terms of the packet states:

$$i = [k_n(n), k_{n-1}(n), \dots, k_{n-\tau_S+1}(n)] \quad (5.3)$$

$$l = [k_{n+1}(n+1), k_n(n+1), \dots, k_{n-\tau_S+2}(n+1)]$$

Figure 5.1. State Transition Diagrams of the Packets for FD Protocol with Finite Receiver Buffer

With these definitions we may have nonphysical packets (sequence numbers, $j \leq 0$) in the system, these will be assumed to stay in the network during their time in the system, $k_j(n) = 0$, with probability one. At time $n = 0$, only the following state has nonzero probability:

$$P\{i_1=0, i_2=0, \dots, i_{\tau_S}=0\} = 1. \quad (5.4)$$

Next we determine the transition probabilities of the system Markov chain denoted by $p_{\ell/i}$; Let us make the following definitions:

$I_0(i), I_1(i), I_2(i)$: Number of packets in the network, buffer, and discarded at or before n from among the most recent τ_S arrivals to the system respectively (excluding the nonphysical packets).

$L_0(\ell), L_1(\ell), L_2(\ell)$: are similarly defined for time $(n+1)$.

KP: Buffer capacity in number of packets.

EP(n): Number of the empty places in the buffer at time n , just an instant before the current arrivals from the network.

U(n): Number of arrivals at time n .

EP(n+1), U(n+1) depend on $I_r(i), L_r(\ell), r = 0, 2$ as well as the state of the packet to be delivered to the DA converter at time n :

$$EP(n+1) = \begin{cases} KP - I_1(i) & k_{n-\tau_S+1}(n) = 0 \text{ or } 2 \\ KP - I_1(i) + 1 & k_{n-\tau_S+1}(n) = 1 \end{cases} \quad (5.5)$$

$$U(n) = \begin{cases} \sum_{r=1}^2 (L_r(\ell) - I_r(\ell)) & k_{n-\tau_S+1}(n) = 0 \\ \sum_{r=1}^2 (L_r(\ell) - I_r(i)) + 1 & k_{n-\tau_S+1}(n) > 0 \end{cases} \quad (5.6)$$

Clearly certain system transition probabilities are zero, these transitions involve:

- i) Packets which make transitions between states 1 and 2
(see Figure 5.1)
- ii) Nonphysical packets which make transitions to states 1 or 2.
- iii) Packets to be discarded when there are empty places in the buffer.
- iv) Packets to be stored in the buffer more than its capacity.

In the case of nonzero transition probabilities there are two possibilities: on the first hand, if the number of arrivals to the buffer is less than the empty places in the buffer ($U(n) \leq EP(n)$), then all arrivals go to the buffer. On the other hand, if the number of arrivals is greater than the empty places ($U(n) > EP(n)$), then some of them have to be discarded. We assume that the probability of

discarding a packet is the same for all these arrivals. There are

$\frac{U(n+1)}{EP(n+1)}$ equally likely ways of placing the arrivals into the buffer and

discarding them. Thus transition probabilities are given by:

$$P_{\ell/i} = [P\{k_j(n+1)=0|k_j(n)=0\}]^{L_0(\ell)} [P\{k_j(n+1)>0|k_j(n)=0\}]^{\frac{U(n+1)-1}{bc}} \quad (5.7)$$

where

$$bc = \begin{cases} \frac{U(n+1)}{EP(n+1)} & U(n+1) > EP(n+1) \\ 1 & U(n+1) \leq EP(n+1) \end{cases}$$

Substituting from Figure 5.1 for the unknown probabilities:

$$P_{\ell/i} = \frac{q^{L_0(\ell)} (1-q)^{U(n+1)-1}}{bc} \quad (5.8)$$

Using the above equation to calculate the nonzero transition probabilities, the system transition matrix at any time, $P(n)$, may be determined. Let us define $\pi^{(n)}$ as the probability state vector of the system at time n :

$$\pi^{(n)} = [\pi_0^{(n)}, \pi_1^{(n)}, \dots, \pi_i^{(n)}, \dots, \pi_{M_S-1}^{(n)}] \quad (5.9)$$

where $\pi_i^{(n)}$ is the probability that the system is in state i defined in (5.2). State probabilities may be determined recursively from:

$$\pi^{(n+1)} = \pi^{(n)} P(n) \quad n \leq \alpha + \tau_S \quad (5.10)$$

where initial state vector $(\pi^{(0)})$ is obtained using (5.4).

5.2 Buffer Distribution and Expected Number of Packets Discarded due to Overflow

Buffer distribution at any time is determined by summing over all those state probabilities which give the same number of packets in the buffer, or mathematically:

$$P\{k \text{ packets in the buffer at time } n\} = \sum_{i=0}^{M_S-1} \pi_i^{(n)} \quad (5.11)$$

and $I_1(i) \forall I_1(i) = k$

where $I_1(i)$ is as defined before.

The probability that a packet is discarded due to overflow is similarly determined. n is the delivery time of $(n - \tau_S + 1)$ 'st packet to the DA converter:

$$P\{(n - \tau_S + 1)\text{'st packet discarded before } n\} = P\{k_{n - \tau_S + 1}(n) = 2\}$$

which is given by,

$$P\{k_{n - \tau_S + 1}(n) = 2\} = \sum_{i=0}^{M_S-1} \pi_i^{(n)} \quad (5.12)$$

and $i_{\tau_S} = 2$

Defining total number of packets discarded due to overflow as d_0 , its expected value is given by:

$$E\{d_0|\alpha\} = \sum_{j=1}^{\alpha+1} P\{k_j(j-1+\tau_S) = 2\} \quad (5.13)$$

A similar expression may be given to the packets discarded due to late arrivals, d_1 :

$$E\{d_1|\alpha\} = \sum_{j=1}^{\alpha+1} P\{k_j(j-1+\tau_S) = 0\} \quad (5.14)$$

Total expected number of packets discarded (d) is sum of the two:

$$E\{d|\alpha\} = E\{d_0|\alpha\} + E\{d_1|\alpha\} \quad (5.15)$$

5.3. Numerical Results and Discussion

In the numerical analysis we have assumed a mean network delay of 2 slots. Compared to the example in section 3.6 smaller talkspurt length (16 packets) has been assumed to keep the required computations manageable.

Figure 5.2 shows the expected number of packets in the buffer for both finite and infinite buffers for different τ_S . The latter results agree with the corresponding results determined by the PGF analysis in chapter 3 (see Figure 3.4). Finite and infinite buffer results follow the same curve except in the central section where they are almost parallel to each other.

Due to asymptotic convergence we have found that the distribution of the packets in the buffer becomes stationary at the end of the central section in Figure 5.2. This distribution has been plotted in Figure 5.3, clearly finite and infinite buffer results are very similar.

Finally in Table 5.1 percentage of the packets lost has been tabulated. Because of the small buffer size, lost due to overflow is quite high.

The main drawback of this analysis is the high computation time due to the size of the transition matrix. As an example for $\tau_S = 10$, it has the dimensions of $3^{10} \times 3^{10}$. Because of this, it cannot be applied to CD and TD protocols.

Figure 5.2 Mean Buffer Length for FD Protocol with
a Finite Buffer (of Length KP packets)
Fixed Talkspurt $\alpha=15$
Network Delay $E\{\tau_N\} = 2$
 τ_S = System delay of a packet in slots

Figure 5.3 Probability Distribution of k Packets in the Receiver Buffer for FD Protocol with a Finite Buffer (of Length KP packets)
 Fixed Talkspurt $\alpha=15$
 Network Delay $E\{\tau_N\} = 2$
 τ_S = System delay of a packet in slots

τ_S	KP	% Packet Lost		
		Late	Overflow	Total
4	3	6.25	13.36	19.61
5	4	3.12	10.78	13.90

Table 5.1 Packet Loss Rate for FD Protocol with a Finite Buffer (of Length KP packets)

Network Delay $E\{\tau_N\} = 2$

Chapter 6

THE STEADY STATE ANALYSIS

In the previous chapters, we have determined the buffer distribution assuming a single talkspurt since a system which is working well should not have much overlap between successive talkspurts. Nevertheless, when talkspurts follow each other with short silent intervals, overlap may occur. Furthermore, previous analysis fails to give information about the long-term buffer occupancy.

In this chapter, the steady state buffer distribution is determined for a speaker alternating between talkspurt and silent intervals under an infinite buffer assumption. The analysis is confined to the FD protocol.

6.1 Model

The speaker model of section (2.1) is used, where the speaker follows a two state Markov chain, with the steady-state probabilities given by equation (2.3). Geometric network delay is assumed with the same mean value for all talkspurts. Since under FD protocol, the system time of a packet, τ_S , depends on the mean network delay, all talkspurts will have the same τ_S . Further we assume that the silent intervals are composed of τ_S packets which need no storage and arrive to the receiver instantaneously. Under these assumptions the receiver state follows the speaker state with a fixed delay of τ_S slots (See Figure 6.1) which allows use of results in Chapter 3 to determine the distribution of the packets in the buffer.

ts = talkspurt state ss = silent state

Figure 6.1. States of the Transmitter and Receiver at Steady-State for FD Protocol.

6.2 The Steady-State Buffer Distribution for Protocol FD

The steady-state distribution is determined by a transient PGF analysis similar to the one in chapter 3.

Suppose time zero ($n=0$) is an instant long after the conversation has started. Let the sequence number of arrivals to the network in the interval $(0, \tau_S]$ be $1 \leq j \leq \tau_S$, including any silent packets. Then the steady-state buffer distribution at τ_S is equal to the distribution of the number of arrivals to the buffer from among these packets until $n = \tau_S$.

Let us describe the transmitter speaker state at time j as m_j , then $m_j = ts$ or ss with ts and ss defined as in Figure 2.1 and also as shown in Figure 6.1. Further defining the state of the transmitter in the interval $(0, \tau_S]$ by the vector m with τ_S components:

$$m = [m_1, m_2, \dots, m_j, \dots, m_{\tau_S}] \quad (6.1)$$

Clearly the number of transmitter states, m , is 2^{τ_S} , and their respective probabilities are given by:

$$P\{m\} = P\{m_1\}P\{m_2|m_1\} \dots P\{m_{\tau_S}|m_{\tau_S-1}\} \quad (6.2)$$

where $P\{m_1\}$ is given by equation (2.3) and $P\{m_j|m_{j-1}\}$ are given in Figure (2.1).

Let us describe the state of the talkspurt packets at time $n = \tau_S$ by the random variable $k_j(\tau_S)$, then:

$$k_j(\tau_S) = \begin{cases} 0 & \text{j'th talkspurt packet in the network at time } \tau_S \\ 1 & \text{j'th talkspurt packet is in the buffer at time } \tau_S \end{cases} \quad (6.3)$$

The probability distribution of $k_j(\tau_S)$ is given by equations (3.1-3.2) with $n = \tau_S$: let the PGF of a packet at $n = \tau_S$ be defined as $K_j(z, \tau_S)$. Then, conditioned on the transmitter state when the packet arrives to the network, we have:

$$K_j(z, \tau_S | m_j) = \begin{cases} (1-q)^{\tau_S-j+1} z + q^{\tau_S-j+1} & m_j = ts \\ 1 & m_j = ss \end{cases} \quad (6.4)$$

where above we have defined the PGF of silent packets as 1.

Then conditioned on the transmitter state vector m , the PGF of the number of arrivals to the buffer at $n = \tau_S$ from among the packets $1 \leq j \leq \tau_S$ is given by:

$$K(z, \tau_S | m) = \prod_{j=1}^{\tau_S} K_j(z, \tau_S | m_j) \quad (6.5)$$

The steady-state buffer distribution is determined by unconditioning (6.5) wrt m :

$$K(z, \tau_S) = \sum_m \prod_{j=1}^{\tau_S} K_j(z, \tau_S | m_j) P\{m\} \quad (6.6)$$

6.3 Numerical Results and Discussion

In numerical calculations a mean network delay of 2 slots, mean talkspurt and silent interval lengths of 40 and 52 packets have been used as in section 3.6. Following equation (6.6), the buffer distribution has been plotted for different τ_S (Figure 6.2). These plots are similar to the conditional single-talkspurt plots given in Figure 3.8, except for the large peaks at a zero number of packets due to the silent intervals. As a result of this high empty buffer probability, sharing of buffer space among users will be advantageous.

The analysis of the CD protocol by the same approach is straightforward but tedious, and is expected to behave much like FD. TD could not be analyzed because the model over a long period of time does not have a stationary limit.

Figure 6.2 The Steady-State Probability Distribution of k Packets in the Receiver Buffer for FD Protocol
Talkspurt Length $E(\alpha)=39$, Silent Interval $E(\beta)=51$
Network Delay $E(\tau_N)=2$

τ_S = System delay of a packet in slots

Chapter 7

OPTIMAL PACKET SIZE

In this chapter we address the first design problem discussed in chapter 1, namely determining the optimal packet size which minimizes the total delay under a constraint on speech quality. The minimization form to be used has been suggested by Barberis and Pazzaglia in [11], but has not been solved. D. Minoli [9] has solved this problem by omitting the constraint on speech quality and excluding the receiver delay from the total delay. In this chapter, for the protocols studied so far, we determine the optimal packet size by solving the constrained minimization problem on the total delay. In the last section, the design of a packet voice system which uses the results derived in the thesis is discussed.

7.1 Statement of the Model

The objective is to minimize some function of the mean total delay and its variance under a constraint on speech quality related to the packet loss rate defined in section 4.1. For the sake of simplicity a linear function will be used:

$$\begin{aligned} \min_{t_p} \quad & \bar{t}_D^2 + b(\bar{t}_D^2 - \bar{t}_D^2) \\ \text{subject to} \quad & \text{packet loss rate } d^* = b^* \end{aligned} \tag{7.1}$$

where both b and b^* are constants.

b is to be chosen on the basis of the relative effects of the mean total delay and variance of this delay. Experimental studies have shown that [4,8] the effect of increased total delay is psychological, and when the one way delay of a packet reaches to 0.5 sec, conversations become difficult. On the other hand the variance of the delay causes gaps in the speech which affect intelligibility. Because these studies are not comprehensive enough to suggest a value for b , for simplicity b will be chosen as unity, which reduces the minimization problem to:

$$\min_{t_p} \overline{t_D^2} \quad (7.2)$$

subject to $d^* \leq b^*$

Next let us make the following definitions:

C channel capacity (bits/sec)

v channel capacity (packets/sec)

a arrival rate to the network (packets/sec)

$\rho = \frac{a}{v}$ channel utilization

$\epsilon = \frac{t_p}{t_p + t_h}$ transmission efficiency

where t_p and t_h are packet information and header duration in sec. respectively.

The following relationship may be written among these variables:

$$v = \frac{C}{R(t_p + t_h)} \quad (7.3)$$

where as defined before R is the AD rate in bits/sec.

If an average of N voice circuits are active on the network, and further speech is active only about half of the time, the two state Markov chain model of the speech (section 2.1.1) may be approximated by a Poisson process [5] with average arrival rate to the network given by:

$$a = \frac{N}{2t_p} \quad (7.4)$$

Then the channel utilization is given by:

$$v = \frac{v_o}{\epsilon} \quad (7.5)$$

where v_o is defined as the useful channel utilization, and is given by:

$$v_o = \frac{NR}{2C} \quad (7.6a)$$

Following substitution for the transmission efficiency (ϵ) in equation (7.5), the stability requirement of the channel ($v < 1$) gives the following minimum packet size:

$$t_p > \frac{v_o t_h}{1-v_o} \quad (7.6b)$$

The pdf of the network delay given by (2.4) becomes:

$$p_N(t_N) = v(1-v)e^{-v(1-v)t_N} \quad (7.7)$$

and q defined in (2.5) becomes

$$q = e^{-v(1-v)t_p} \quad (7.8)$$

Next we solve the minimization problem for each of the protocols. In this, we shall use a continuous time variable, fortunately this does not introduce any new conditions on the protocols.

7.2 Optimal Packet Size for the FD Protocol

For the FD protocol total packet delay is a constant given by:

$$t_D = t_p + \tau_S t_p \quad (7.9)$$

where t_p is the packet duration and τ_S system delay of a packet.

From (4.5b) the packet loss rate is given by:

$$d^* = q^{\tau_S} \quad (7.10)$$

or substituting for q from (7.8):

$$d^* = e^{-v(1-v)t_p \tau_S} \quad (7.11)$$

Thus minimization problem of (7.2) becomes:

$$\text{Min}_{t_p} t_D = t_p + \tau_S t_p \quad (7.12)$$

where τ_S is a function of t_p^* defined implicitly by:

$$e^{-v(1-v)t_p \tau_S} = b^* \quad (7.13)$$

Substituting from (7.4-7.6) into (7.13), we have:

$$\tau_S = \frac{D_F(t_x+1)v_o}{t_x(1-v_o)-v_o} \quad t_x > \frac{v_o}{1-v_o} \quad (7.14)$$

where

$$D_F = \frac{2 \ln(1/b^*)}{N} \quad (7.15)$$

and

$$t_x = \frac{t_p}{t_h} \quad (7.16)$$

The condition on t_x in (7.14) comes from the equation (7.6b). Substitution of (7.14) into (7.12) gives the total delay as:

$$t_D = t_h t_x \left[1 + \frac{D_F(t_x+1)v_o}{t_x(1-v_o)-v_o} \right] \quad t_x > \frac{v_o}{1-v_o} \quad (7.17)$$

Taking the derivative in the above equation wrt t_p , and setting it equal to zero gives the optimal packet size as:

$$t_p = \frac{v_o t_h}{1-v_o} \left[1 + \left(\frac{D_F}{1-v_o+D_F v_o} \right)^{1/2} \right] \quad (7.18)$$

The plus sign gives the desired solution (t_D is negative under the minus sign. Thus the condition $t_x \geq \frac{v_o}{1-v_o}$ is always satisfied at the minimum, and it is not binding). In the above equation the factor

$\left(\frac{v_0 t_h}{1-v_0} \right)$ is the minimum packet size (see 7.6b), and in all practical cases $D_F \ll 1$.

It may seem surprising that the optimal t_p does not depend on the talkspurt length, but just on the total rate of transmission of speech samples, which is taken as $NR/2$ bits/sec., and enters the solution via v_0 . Any effect of talkspurt length was dropped in constructing the Poisson arrival model for the channel behaviour. This is eminently reasonable if we are looking at a large number of speech circuits with their packets interleaved in time.

7.3 Optimal Packet Size for the CD Protocol

Analysis of this protocol is very similar to the FD protocol.

Total packet delay is given by:

$$t_D = t_p + \tau_S t_p \quad (7.19)$$

where τ_S (system delay of a packet) is a random variable given by equation (3.18). As in the FD, total delay is minimized with a constraint on the packet loss rate except that here the mean of t_D is used:

$$\bar{t}_D = t_p + \bar{\tau}_S t_p \quad (7.20)$$

Substituting from (3.18) and (7.7) gives:

$$\bar{t}_D = t_p + \left[\frac{1}{v(1-v)t_p} + \mu \right] t_p \quad (7.21)$$

From (4.7b) packet loss rate, d^* for this protocol is given by:

$$d^* = \frac{q^{\mu+1}}{1+q} \quad (7.22)$$

or substituting for q from (7.8):

$$d^* = \frac{e^{-v(1-v)t_p(\mu+1)}}{1 + e^{-v(1-v)t_p}} \quad (7.23)$$

Thus the minimization problem is given by:

$$\min_{t_p} \bar{t}_D = t_p + \left[\frac{1}{v(1-v)t_p} + \mu \right] t_p \quad (7.24)$$

subject to

$$\frac{e^{-v(1-v)t_p(\mu+1)}}{1 + e^{-v(1-v)t_p}} = b^* \quad (7.25)$$

Solving for μ in the last equation gives

$$\mu = \frac{1}{v(1-v)t_p} \ln \frac{1}{b^*(1 + e^{-v(1-v)t_p})} \quad (7.26)$$

Substituting the above equation and from (7.4-7.6) into (7.24) gives:

$$\bar{t}_D = t_h t_x \left\{ 1 + \frac{2(t_x+1)v_o}{N[t_x(1-v_o)-v_o]} \left[1 + \ln \frac{1}{b^* \left[1 + e^{-\frac{N}{2} \frac{t_x(1-v_o)-v_o}{(t_x+1)v_o}} \right]} \right] \right\} \quad (7.27)$$

$$t_x > \frac{v_o}{1-v_o}$$

Comparing the above equation with the corresponding result of FD (7.17), it is seen that though two are quite similar, (7.27) is more complicated and its minimization will not be followed any further.

7.4 Optimal Packet Size for the TD Protocol

In the TD protocol, total packet delays are random, and later delays are dependent on the earlier ones. Furthermore, later delays have larger mean and variance than the earlier ones. Therefore we shall consider the j 'th packet in the talkspurt, and minimize its total delay denoted by t_{Dj} wrt t_p . Clearly, in this protocol we do not need the constraint on the packet loss rate. Thus the minimization problem of (7.2) becomes:

$$\min_{t_p} \overline{t_{Dj}^2} \quad (7.28)$$

where t_{Dj} is given by:

$$t_{Dj} = t_p + t_{Sj} \quad (7.29)$$

with t_{Sj} being the system delay of j 'th packet (network + receiver delay). After substituting from the above equation into (7.28), minimization problem becomes:

$$\min_{t_p} \overline{t_{Dj}^2} = t_p^2 + 2t_p \overline{t_{Sj}} + \overline{t_{Sj}^2} \quad (7.30)$$

Next we determine the first and second moment of the system delay of a packet. The pdf of the system delay of a packet has been determined by Barberis and Pazzaglia [10], and from there it is given by the following formula:

2 2
OF / DE

$$P_{Sj}(t_S) = \frac{d}{dt_S} [P_N(t_S) P_{S(j-1)}(t_S)] \quad (7.31)$$

where $P_{Sj}(t_S)$, and $P_{Sj}(t_S)$ are the pdf and CDF of j 'th packet system delay, and $P_N(\cdot)$ is the CDF of network delay. Substituting for the CDF of the network delay from (7.7) in the last equation:

$$P_{Sj}(t_S) = \frac{d}{dt_S} [P_{S(j-1)}(t_S) - e^{-v(1-v)t_S} P_{S(j-1)}(t_S)] \quad (7.32)$$

Let us define the Laplace transform of the pdf of j 'th packet system delay as:

$$T_{Sj}(\theta) = \int_0^{\infty} P_{Sj}(t_S) e^{-\theta t_S} dt_S \quad (7.33)$$

Now, taking the Laplace transform of the both sides of (7.32) give:

$$T_{Sj}(\theta) = T_{S(j-1)}(\theta) - \frac{\theta T_{S(j-1)}(\theta + v(1-v))}{\theta + v(1-v)}, \text{ for } j \geq 2 \quad (7.34)$$

The above recursive relationship may be expressed in terms of the Laplace transform for the first packet as:

$$T_{Sj}(\theta) = \sum_{i=0}^{j-1} \binom{j-1}{i} (-1)^i \frac{\theta [T_{S1}(\theta + iv(1-v))]}{\theta + iv(1-v)}, \text{ for } j \geq 2 \quad (7.35)$$

From the definition of TD protocol, the system delay of first packet is given by its network delay plus an intentional delay μt_p , therefore:

$$P_{S1}(t_S) = P_N(t_S - \mu t_p) \quad t_S \geq \mu t_p \quad (7.36)$$

Substituting from (7.7) in the above equation and taking Laplace transforms give:

$$T_{S1}(\theta) = \frac{e^{-\mu t_p \theta} v(1-v)}{\theta + v(1-v)} \quad (7.37)$$

Substituting the above result into (7.35) gives:

$$T_{Sj}(\theta) = \sum_{i=0}^{j-1} \binom{j-1}{i} (-1)^i \frac{\theta v(1-v) e^{-\mu t_p [\theta + iv(1-v)]}}{[\theta + iv(1-v)] [\theta + (i+1)v(1-v)]}, \text{ for } j \geq 2 \quad (7.38)$$

This gives the Laplace transform of the pdf of the j 'th packet's system delay, which is a result not found in [10]. Successive differentiation of (7.38) gives the first and second moment of the j 'th packet system delay as:

$$\bar{t}_{Sj} = - \left. \frac{dT_{Sj}(\theta)}{d\theta} \right|_{\theta=0} = \mu t_p + \frac{1}{v(1-v)} - \sum_{i=1}^{j-1} \binom{j-1}{i} (-1)^i \frac{e^{-\mu t_p iv(1-v)}}{i(i+1)v(1-v)} \quad (7.39)$$

$$\begin{aligned} \overline{t_{Sj}^2} &= \left. \frac{d^2 T_{Sj}(\theta)}{d\theta^2} \right|_{\theta=0} = 2\mu t_p \overline{t_{Sj}} - \mu^2 t_p^2 + \frac{2}{v^2(1-v)^2} \\ &\quad - \sum_{i=1}^{j-1} \binom{j-1}{i} (-1)^i e^{-\mu t_p i v(1-v)} \frac{2(2i+1)}{i^2(i+1)^2 v^2(1-v)^2} \end{aligned} \quad (7.40)$$

(Substituting from (7.39-7.40) and (7.4-7.6) into (7.30) gives the minimization problem as:

$$\begin{aligned} \min_{t_p} \overline{t_{Dj}^2} &= t_h^2 t_x^2 \left\{ \left[\mu + 1 + \frac{2(1+t_x)v_o}{N(t_x(1-v_o)-v_o)} \right]^2 + \frac{4(1+t_x)^2 v_o^2}{N^2(t_x(1-v_o)-v_o)^2} \right. \\ &\quad - \frac{4(\mu+1)(1+t_x)v_o}{N(t_x(1-v_o)-v_o)} \sum_{i=1}^{j-1} \binom{j-1}{i} (-1)^i e^{-\frac{\mu N i(t_x(1-v_o)-v_o)}{2(1+t_x)v_o}} \frac{2(2i+1)}{i^2(i+1)^2} \\ &\quad \left. - \frac{4(1+t_x)^2 v_o^2}{N^2(t_x(1-v_o)-v_o)^2} \sum_{i=1}^{j-1} \binom{j-1}{i} (-1)^i e^{-\frac{\mu N i(t_x(1-v_o)-v_o)}{2(1+t_x)v_o}} \frac{2(2i+1)}{i^2(i+1)^2} \right\} \\ &\quad t_x > \frac{v_o}{1-v_o} \end{aligned} \quad (7.41)$$

The above problem is quite involved, therefore probably the best way is computer minimization such as Fletcher Powell method [21]. Once the optimal value of t_p has been determined, other quantities such as mean total delay of the j 'th packet may easily be calculated.

7.5 Numerical Results and Discussion

In Figures 7.1 and 7.2 some results for protocol FD have been plotted for 2% packet loss rate and for the number of voice circuits, $N = 20, 80, 1000$ with fixed header size. Figure 7.1 shows the optimal packet size and the corresponding minimum total delay (7.17, 7.18) against useful utilization (v_0). It will be appropriate to compare the points on $N = 20$ with useful utilization four times smaller than those on $N = 80$ curves since they correspond to the same AD rate and channel capacity (see 7.6a). For the corresponding points, it is seen that the minimum total delay for the $N = 20$ is significantly smaller than the $N = 80$. Figure 7.2 shows the total delay against packet size from (7.17) for $N = 20$, and 80 with $v_0 = 0.8$ and 0.2 respectively. In both plots the infinity preceding the minimum is due to queueing delay, while the linear increase following the minimum is due to packetization delay. For $N = 20$, $v_0 = 0.2$, it is seen that at the optimal packet size, the transmission efficiency is very low. This is because of the light traffic level; we can afford to use small packet size to decrease total delay by decreasing packetization delay. However a slight increase in packet size may increase the transmission efficiency substantially with little change in the delay.

In Figure 7.3 we have plotted $\bar{t}_{Dj}$ and $\bar{t}_{Dj}^2$ for protocol TD against packet size rather than minimizing $\bar{t}_{Dj}^2$ wrt t_p . Results are plotted for packets $j = 20$ and $j = 40$, with $\mu = 0$ and fixed header size. Though we have plotted only for $\mu = 0$, we found that $\bar{t}_{Dj}^2$ gets consistently larger as μ gets larger. It would have been

better to do the minimization wrt both t_p and μ , but this needs the use of different weighting factor (b) than unity in equation (7.1). As has been said before a clear choice for b is still a question.

7.6 A Packet Voice System Design

In this section we discuss packet voice system design for protocols FD and TD (CD is very similar to the FD). The designer is assumed to know the channel capacity (C), voice AD rate (R), and the maximum value of the average number of voice circuits in the network ($N_{\max}$). The design involves determining the optimal packet size, total delay, transmission efficiency, packet loss rate for FD (or μ for TD), buffer size, and it is a trade off among these quantities.

Design with the FD protocol may be summarized in the following algorithm:

Step 1: Assuming a small packet loss rate and using $N = N_{\max}$, the optimal packet size, total delay, and transmission efficiency are determined from Figures 7.1-7.2 with appropriate N and packet loss rate. Then the system delay of a packet is calculated as $\tau_s = \frac{t_D}{t_p} - 1$. After rounding τ_s to the next integer, it is used to determine the buffer size as described in section 4.5.

Step 2: If the transmission efficiency is low but the total delay and buffer size are acceptable, step 1 is repeated using a nonoptimal packet size.

Step 3: If the total delay or buffer size is unacceptable steps 1 and 2 are repeated with an acceptable higher packet loss rate,

Step 4: If the packet loss rate becomes unacceptable steps 1-3 are repeated with a lower $N < N_{\max}$.

Design with the TD protocol similarly may be described in an algorithm:

Step 1: Assuming $\mu = 0$, and $N = N_{\max}$ optimal packet size, total delay, and transmission efficiency are determined from Figure 7.3 with appropriate N and μ (further a specific value of a packet number has to be assumed). Then using the resulting total delay and μ , buffer size is determined as described in section 4.5.

Step 2: From Figure 3.7 if the gaps occurring in the talkspurt are unacceptable, the total delay and buffer size are acceptable, then the design is repeated with a higher value of μ .

Step 3: If the transmission efficiency is too low and the total delay and buffer size are acceptable, steps 1 and 2 are repeated with a non-optimal packet size.

Step 4: If the total delay or buffer size becomes unacceptable, steps 1-3 are repeated with $N < N_{\max}$.

After a packet voice system has been designed, packet size will not be changed easily. The transient variations in the network delay due to fluctuations in the traffic will be taken care of by the receiver protocols as described in section 1.2.1.

Figure 7.1 Optimal Packet Size and Minimum Total Delay
for FD Protocol
Packet Loss Rate = 2%

— Total Delay
- - - Normalized Packet Size
N = Number of Voice Circuits

Figure 7.2 Total Delay for FD Protocol against Packet Size
Packet Loss Rate = 2%

N = Number of voice circuits

v_0 = Useful utilization

Figure 7.3 Mean and Second Moment of the Total Delay for TD Protocol against Packet Size
 $N=80$, $v_0=0.8$

— $j=20$

- - - $j=40$

j = Packet sequence-number

Chapter 8

DELAYS IN A CSMA/CD PROTOCOL FOR LOCAL AREA NETWORKS

In the previous chapters, we have studied packetized voice in long-haul packet switched networks. In the present chapter we derive results concerning delays in a local area network which are applicable both to voice and data. A popular local area network uses a shared channel with distributed control based on carrier sense multiple access with collision detection (CSMA/CD) for access to the channel. Several CSMA/CD protocols have been studied which are very close to the protocol implemented in the Ethernet [15,16,17]. So far results derived on these protocols have been limited to the mean delay, throughput and queue size distribution, except for a delay distribution result given by Tobagi [18].

In this chapter, we determine the delay distribution and its moments as well as the queue size distribution for a CSMA/CD protocol slightly different from the one studied by Tobagi. Also, simple asymptotic results are given for the delay distribution and its moments under heavy traffic, and we show that these results are very useful at intermediate traffic levels as well. As we have seen in the earlier chapters, the distribution results are very important for real-time applications such as packetized voice and process control.

The notation to be used in this chapter will be redefined independently of the previous chapters, and is given in a separate table in the list of frequently used symbols at the beginning of the thesis.

8.1 The CSMA/CD Protocol

The protocol that we shall study is the p-persistent protocol [15] with geometrically distributed retransmission delay having adaptive mean value. This protocol has been studied by Johnson and O'Leary in [17], in which they have determined the mean packet delay. Next this protocol is described in some detail.

All users are connected in parallel to a channel, usually a coaxial cable. Let us define a ready user as a user with a packet to transmit, and a contention interval as the time during which the next user to transmit is chosen. A contention interval either (1) begins at the end of a successful transmission if there are any ready users at that instant, or (2) begins at the arrival of the first ready user during an idle period. It ends with the beginning of the next successful transmission.

A contention interval will be assumed to be slotted in time and ready users may transmit only at the beginning of a slot. Then according to this protocol each ready user transmits with probability p_s at the beginning of a contention interval. A user sensing a collision, stops its transmission, and so do the other colliding users. The time required for a user to detect the transmission of another is, in the worst case, the maximum round trip propagation delay of the channel. Any user having a packet to transmit, whether it participated in the collision or not, waits for the channel to clear. Let us define a slot, or collision duration, as the maximum round trip propagation time plus channel clearing time. Following a collision all the ready users again transmit with probability p_s in the next slot. p_s is adjusted dynamically at each slot in such a way to maximize the probability of successful transmission.

Next we derive the distributions of retransmission time and contention interval in slots. Let us make the following definitions:

x = retransmission time, the length of the delay before the next transmission attempt of a ready user, after a collision

c = length of a contention interval

j = number of the users ready to transmit at each slot during a contention interval

q = probability that exactly one user attempts to transmit in a slot during a contention interval and will be called success probability.

Since each ready user transmits with probability p_s at any slot, the retransmission time is geometrically distributed given by:

$$p_x(i) = P\{\text{retransmission time is } i \text{ slots}\} = p_s(1-p_s)^i \quad i=0,1,2,\dots \quad (8.1)$$

Adjustment of p_s dynamically, to be shown shortly, gives rise to the adaptive variation in the retransmission time distribution stated before.

With j users ready to transmit q is:

$$q = jp_s(1-p_s)^{j-1} \quad (8.2a)$$

and the contention interval (c) also has a geometric distribution given by:

$$p_c(n) = P\{\text{contention interval is } n \text{ slots}\} = q(1-q)^n \quad n = 0,1,2,\dots \quad (8.2b)$$

The value of p_s which maximizes q is given in [14,17] as:

$$p_s = \frac{1}{j} \quad (8.3)$$

and the resulting value of success probability (q) is:

$$q = \left(1 - \frac{1}{j}\right)^{j-1} \quad (8.4)$$

As the number of contending users j becomes large, the optimal value of q approaches $1/e$ rapidly. For small values of j the optimal q is always greater than $1/e$. For efficiency this protocol requires each ready user to estimate the number of users trying to transmit, and act accordingly and a number of algorithms to determine this estimate have been suggested in [17,22]. In the following analysis we shall assume that the contention intervals are i.i.d. according to (8.2b). From (8.3) it is seen that at the end of a contention interval one of the ready users is selected at random for transmission.

In the analysis we shall also assume i.i.d. packet transmission times. After a successful transmission all ready users will wait a slot duration for the channel to clear up before the next contention interval starts. Successful transmission time will be defined as to include this extra slot.

We shall model the users by an infinite user population with a Poisson arrival process and arrival rate λ packets/sec into the ready state. This is an approximation to a large finite population in which each user generates packets infrequently. Any newly arriving

packet is assumed to increase the number of ready users by one. All ready users together will be regarded as a single queue. Under the Poisson assumption arrivals are independent of the queue size.

A protocol similar to the one described above has been studied by S.S. Lam [16] who found its queue size distribution and mean delay. The main difference between his protocol and the one described above is that in the first slot of the contention interval all ready users transmit with probability one. While under light traffic this prevents unnecessary waiting, under heavy traffic it adds an extra slot to the length of nearly every contention interval. Next we give a brief description of the protocol studied by Tobagi [18] who found its delay distribution. Following a successful transmission each ready user transmits with probability one, in the case of a collision users defer with geometrically distributed retransmission times with a fixed mean value (fixed p_s). A finite user population has been assumed which results in the arrival rate dependent on the queue size. The PGF of the delay distribution and a recursive formula for moments have been derived using a Markov chain analysis for a constant packet size, however these results are rather complicated.

8.2 Analysis of the CSMA/CD Protocol

We determine the steady-state queue size and waiting time distributions for the CSMA/CD protocol we have described in the previous section. Both parts will make use of the well known results by making certain assumptions. Before we go into detailed analysis let us briefly summarize the protocol:

- Poisson arrivals at constant rate λ
- ready users form a single queue
- successful transmissions are preceded by an i.i.d. contention interval with geometric distribution except for the transmission following an idle period, which has a zero contention interval.

- random selection at the end of a contention interval among all ready users.
- packet transmission times are i.i.d., and they are also independent of the contention interval.

Let us define the following variables:

Δ = slot duration

r = packet transmission time

$a = r + \Delta$ = successful transmission time

$b = a + c$ = service interval

where above c is the contention interval as defined before. Expectations of the random variables r, a, b, c will be defined as $\bar{r}, \bar{a}, \bar{b}, \bar{c}$ and Laplace transforms of their respective pdf's as $T_r(\theta), T_a(\theta), T_b(\theta), T_c(\theta)$. Then we have:

$$T_a(\theta) = T_r(\theta)e^{-\theta\Delta} \quad (8.5a)$$

$$T_b(\theta) = T_a(\theta)T_c(\theta) \quad (8.5b)$$

where from (8.2b):

$$T_c(\theta) = \frac{q}{1 - (1-q)e^{-\theta\Delta}} \quad (8.5c)$$

8.2.1 Queue Size Distribution

Figure 8.1a illustrates how this protocol functions. The service time of each user is considered as the successful transmission time plus contention interval preceding it, except for an arrival following an idle period. For the simplicity of analysis we shall also add a contention interval to the service time of an arrival following an idle period. Under this assumption the service time of all users may be taken as an i.i.d. random variable b with distribution defined in (8.5b) (see Figure 8.1b). With the Poisson arrivals we have an M/G/1 queueing system which may be analysed as usual by imbedding a Markov chain at the end of service intervals as shown in Figure 8.1b. Thus PGF of the steady-state queue size distribution is given by the standard result [23]

$$Q(z) = \frac{T_b[\lambda(1-z)](1-\sigma)(1-z)}{T_b[\lambda(1-z)] - z} \quad (8.6a)$$

where σ is the server (channel) utilization given by:

$$\sigma = \lambda \bar{b} \quad (8.6b)$$

8.2.2 The Waiting Time Distribution

We want to determine the waiting time (w) of a tagged user, t , arriving to the system at steady-state. Following the queue size analysis, we shall make use of the well-known results for the waiting time in an M/G/1 system with random service. At this stage the success of this approach is not assured since we may not be able to treat the

Figure 8.1a. Busy and Idle Periods in the CSMA/CD Protocol

Figure 8.1b. Simplified Busy and Idle Periods in the CSMA/CD Protocol

contention interval plus transmission time as a single variable as in the queue size problem.

When tagged user, τ , arrives the system is either idle or busy with respective probabilities $(1-\sigma)$ and σ . We assume that busy state also includes the contention interval. Let us make the following definitions:

- b^* : length of the service interval during which τ arrives (if the system is busy on his arrival)
- c^* : contention part of b^*
- a^* : successful transmission part of b^*
- u : residual service time (time from the arrival of τ until the end of b^*)
- v : waiting time of τ after b^* if any
- a_τ : successful transmission time of τ

Arrival of τ has been shown in Figures 8.2a,b,c and its waiting time is given by:

$$w = \begin{cases} v & \text{with probability } 1-\sigma \\ u+v & \text{with probability } \sigma \end{cases} \quad (8.7a)$$

The analysis of residual service time (u) is rather complicated because given that tagged user, τ , arrives to a busy system, it may arrive during a contention interval (Figure 8.2b) or successful transmission interval

τ arrives

Figure 8.2a. Waiting Time of the Tagged User Arriving During an Idle Period.

τ arrives

Figure 8.2b. Waiting Time of the Tagged User Arriving During a Busy Period in a Contention Interval

τ arrives

Figure 8.2c. Waiting Time of the Tagged User Arriving During a Busy Period in a Transmission Interval

→ Imbedded Points

(Figure 8.2c). In the former case, τ may be the next user to be served. To simplify the analysis we shall assume that the ratio of slot duration to packet transmission time is very small which makes the contention interval small compared to transmission time. Then we may assume that τ is not served during b^* , while the other arrivals compete for service as usual. Next we consider the waiting time of τ after u , which has been defined above as v : during this time the system cannot go to idle state. Since the next user to be served is chosen at the end of a contention interval the appropriate imbedded points for a Markovian analysis of waiting time are the end of contention intervals. These points have also been shown in Figures 8.2a,b,c and it is seen that time between two successive imbedded points is i.i.d. with b except for the two points for which it is c . Under the assumption made above, we may also neglect the odd contention interval occurring at the beginning of v in Figures 8.2a,b,c. In fact neglecting the contention interval following an idle period is no loss since this was an approximation introduced in the previous section. As a result, if τ arrives to an idle system its waiting time will be zero which is realistic, under these assumptions (8.7a) becomes:

$$w = \begin{cases} 0 & \text{with probability } 1-\sigma \\ u+v & \text{with probability } \sigma \end{cases} \quad (8.7b)$$

Under these assumptions we have redrawn Figures 8.2a,b,c as 8.3a,b, as it is seen b^* has been shown as an uninterrupted service interval. Now, we have reduced the system to an M/G/1 queue with random service with service interval b . This problem has been solved by Kingman [24] and Le Gall [25]. From [24], the Laplace transform of pdf of waiting time is given by:

$$T_w(\theta) = \frac{1-\sigma}{\theta} \int_0^1 \frac{1-x}{\xi(\theta) T_b[\lambda(1-x)] - x} \{T_b[\lambda(1-x)] - T_b[\theta+\lambda(1-x)]\} \left(\frac{dK(x,\theta)}{dx} \right) dx + 1-\sigma \quad (8.8)$$

where $K(x,\theta) = \exp\{-\int_x^1 [y - T_b(\theta+\lambda(1-y))]^{-1} dy\}$

and $\xi(\theta)$ is the solution of $\xi(\theta) = T_b[\theta+\lambda(1-\xi(\theta))]$.

There is no intuitive explanation of the above result and mathematically it is not very tractable, but it does give all the moments of the waiting time. Before we give the moments let us specify the following:

$T_{b^*}(\theta)$ = Laplace transform of the pdf of b^* .

From [23] $T_{b^*}(\theta)$ is given by:

$$T_{b^*}(\theta) = -\frac{1}{b} \frac{dT_b(\theta)}{d\theta} \quad (8.9)$$

Figure 8.3a. Waiting Time of the Tagged User Arriving During an Idle Period

→ Imbedded Points

Figure 8.3b. Simplified Waiting Time of the Tagged User Arriving During a Busy Period

where $T_b(\theta)$ is determined by (8.5b).

u (residual service time) is uniformly distributed over the interval $(0, b^*)$.

$T_u(\theta|b^*)$ = Laplace transform of the pdf of u conditional on b^*

$T_u(\theta|b^*)$ is given by:

$$T_u(\theta|b^*) = \frac{1 - e^{-\theta b^*}}{\theta b^*} \quad (8.10)$$

$E\{u^k|b^*\}$ = k 'th moment of u conditional on b^*

$$E\{u^k|b^*\} = (-1)^k \left. \frac{d^k T_u(\theta/b^*)}{d\theta^k} \right|_{\theta=0} \quad (8.11)$$

m = the number of users at the end of b^*

$M(z|b^*)$ = PGF of the distribution of m conditional on b^* .

From [24], $M(z|b^*)$ is given by:

$$M(z|b^*) = \frac{(1-\sigma)(1-z)e^{-\lambda b^*(1-z)}}{T_b[\lambda(1-z)] - z} \quad (8.12)$$

$E\{m^k|b^*\}$ = k 'th moment of m conditional on b^*

Clearly $E\{m^k|b^*\}$ is given by:

$$E[m^k | b^*] = \left. \frac{d^k M(z | b^*)}{dz^k} \right|_{z=1} \quad (8.13)$$

After these results, the k'th moment of the waiting time is given from [24] as:

$$\overline{w}^k = \sigma E_{b^*} \left\{ E[u^k | b^*] + \frac{(\sigma/\lambda)^k k! E[m^k | b^*]}{\prod_{\ell=0}^{k-1} [1 + \ell(1-\sigma)]} \right\} \quad (8.14)$$

with $E[u^k | b^*]$ and $E[m^k | b^*]$ given in (8.11) and (8.13). The subscript b^* shows that the outer expectation is wrt b^* . Though the above equation gives the moments of waiting time, calculations are in general tedious, and only the first two moments will be given here:

$$\overline{w} = \frac{\lambda \overline{b}^2}{2(1-\lambda \overline{b})} \quad (8.15a)$$

$$\overline{w}^2 = \frac{2}{3} \frac{\lambda \overline{b}^3}{(1-\lambda \overline{b})(2-\lambda \overline{b})} + \frac{4\overline{w}^2}{(2-\lambda \overline{b})} \quad (8.15b)$$

as expected, the first moment of the waiting time is the same as of the first come first served M/G/1 queue [23], and also agrees with the result determined by Johnson and O'Leary [17] for this protocol using a different method.

Next we shall express $\overline{w}$ in terms of λ , moments of packet transmission time ($\overline{r}$, $\overline{r}^2$), slot duration (Δ), and probability of

success in a slot (q). From the definition of a , its moments are given by:

$$\bar{a} = \bar{r} + \Delta \quad (8.16a)$$

$$\overline{a^2} = \overline{r^2} + 2\bar{r}\Delta + \Delta^2 \quad (8.16b)$$

From (8.2) c is geometrically distributed, with its moments given by:

$$\bar{c} = \frac{1-q}{q} \Delta \quad (8.17a)$$

$$\overline{c^2} = \frac{(1-q)(2-q)}{q^2} \Delta^2 \quad (8.17b)$$

From the definition of b its moments are given as:

$$\bar{b} = \bar{a} + \bar{c} = \bar{r} + \frac{\Delta}{q} \quad (8.18a)$$

$$\overline{b^2} = \overline{a^2} + 2\bar{a}\bar{c} + \overline{c^2} \quad (8.18b)$$

$$\overline{b^2} = \overline{r^2} + 2\bar{r}\Delta \left[1 + \frac{1-q}{q} \right] + \frac{2(1-q)}{q} \Delta^2 + \frac{(1-q)(2-q)}{q^2} \Delta^2$$

Let us define the ratio of slot duration to packet transmission time as:

$$\mu = \frac{\Delta}{\bar{r}} \quad (8.19)$$

Also defining the useful server (channel) utilization as:

$$\rho = \lambda \bar{r} \quad (8.20)$$

Substituting from (8.18a,b) into (8.15a) and taking (8.19, 8.20) into account gives the first moment of waiting time as:

$$\bar{w} = \frac{\bar{r} \left[\frac{r^2}{r^2} + 2 \frac{\mu}{q} + \frac{2-q}{q^2} \mu^2 \right]}{2 \left[1 - \rho \left(1 + \frac{\mu}{q} \right) \right]} \quad (8.21)$$

Clearly from the above equation, stability of the system requires:

$$\rho < \frac{q}{q + \mu} \quad (8.22)$$

The second moment of waiting time (8.15b) may be expressed in terms of $\bar{r}$, μ , and q in the same way as the first moment, but the result will not be given here.

8.2.3 Asymptotic Waiting Time Distribution in Heavy Traffic

The waiting time distribution as the useful channel utilization (ρ) approaches to its maximum value given in (8.22) is quite important since it gives the worst case network behavior. (Also the approximations of the previous section becomes more and more reasonable). This distribution has also been derived by Kingman [24], for the random service M/G/1 queue and is given by:

$$P[w > \bar{w}x] = 2\sqrt{x} K_1(2\sqrt{x}) \quad (8.23)$$

where $\bar{w}$ is given by (8.21), and $K_1(x)$ is a modified Bessel function of the second kind of order one [26].

Again from [24], k 'th moment of the waiting time under heavy traffic denoted by $\bar{w}_h^k$ is given as:

$$\bar{w}_h^k = (k!)^2 \bar{w}^k \quad (8.24)$$

Thus the first and second moment under heavy traffic are given by:

$$\bar{w}_h = \bar{w} \quad (8.25a)$$

$$\bar{w}_h^2 = 4\bar{w}^2 \quad (8.25b)$$

From (8.25a), the first moment under heavy traffic is valid for all traffic levels. Also under heavy traffic, the first moment of waiting time of the protocol studied by S.S. Lam [16] approaches to the same expression as (8.15a). We shall not be able to make any comparisons with the protocol studied by Tobagi, because there are not any numerical results presented in [18]. Next, the second moment of waiting time for heavy traffic, $\bar{w}_h^2$ (8.25b) is compared with the second moment for all traffic levels, $\bar{w}^2$, (8.15b). Under heavy traffic ($\lambda\bar{b}$) in (8.15b) approaches to 1, and it is seen that $\bar{w}^2$ consists of $\bar{w}_h^2$ plus another term. Plot of the $\bar{w}^2$, $\bar{w}_h^2$ (see Figure 8.5) show that their values are within 10% of each other for useful utilization levels (ρ) as low

as 0.5. Clearly, in practice the heavy traffic moments (8.24) and the pdf of waiting time (8.23) may be used at the intermediate traffic levels with confidence.

8.3 Numerical Results

The transform of the waiting time distribution (8.8) is a quite complicated expression, and its inversion has not been attempted. All moments of the waiting time may be calculated from (8.14), but this is also involved except the first two moments given in (8.15a,b). On the other hand asymptotic waiting time distribution and its moments (8.23, 8.24) are very easy to calculate. Let us define the total delay of a packet as:

$$d = r + w \quad (8.26)$$

Assuming a constant packet length, then the mean total delay is given by:

$$\bar{d} = \bar{r} + \bar{w} = \bar{r} + \frac{\rho \bar{r} [1 + \frac{2\mu}{q} + \frac{2-q}{q^2} \mu^2]}{2[1 - \rho(1+\mu/q)]} \quad (8.27)$$

where for $\bar{w}$ we have substituted from (8.21). In the following calculations the probability of success in a slot is assumed to be $q = 1/e$.

In Figure 8.4 we have plotted normalized mean total delay (8.27) against useful utilization (ρ) for $\mu = 0.1, 0.01$, and 0.001 . The

smallest value of μ gives the smallest mean delay for the same ρ , because of the smaller contention interval relative to the packet transmission time. Figure 8.5 shows the normalized variance of the delay for heavy traffic as well as for all traffic levels from (8.15b, 8.25b, 8.27). As has been said above, for ρ greater than 0.5 these two sets of curves are very close to each other. Figure 8.6 shows the distribution of the total delay under heavy traffic at different values of ρ . It is seen that as ρ gets closer to its maximum value given by (8.22) in each case delay becomes very large.

8.4 Applications to Packetized Voice

The results of this chapter could be used for any real time applications of a CSMA/CD local area network. Here, as an example, we shall consider the transmission of packetized voice. In contrast to the long-haul packet switched networks we have considered in earlier chapters, in a CSMA/CD local area network packets arrive to their destination in sequence, thus there is no reordering problem. But since the packet delays are random, synchronous delivery of the speech to the listener still is a problem. Similar to the long-haul networks, there may be two possible approaches:

i) Packets are buffered both at the transmitter and receiver, and speech is reconstructed after a delay. No packets are discarded.

ii) A packet is discarded after a certain period, if it still has not been transmitted. This period will be defined as the trial time of a packet.

Here we shall only consider the latter approach with packet trial time taken as the inter-arrival time of the voice packets, this way we need to buffer only one packet at the transmitter. Previously defined as the packet duration (t_p) becomes the trial time of a packet. The objective is to determine the packet loss rate assuming that there are on average N voice circuits active on the local area network. Then from equation (7.4), Poisson arrival rate is given by:

$$\lambda = \frac{N}{2t_p} \quad (8.28)$$

Let us define the ratio of packet trial time (t_p) to mean packet transmission time ($\bar{r}$) as:

$$v = \frac{t_p}{\bar{r}} \quad (8.29)$$

and d^* as the probability that a packet will be discarded, or packet loss rate, then:

$$d^* = P\{w \geq v\bar{r}\} \quad (8.30)$$

where the above probability may be determined from (8.23).

Let us assume that the digitization rate for voice is $R = 16K$ bits/sec, and packet duration, $t_p = 33$ msec. Thus packet length is 528 bits, or with header assumed to be 600 bits. As a conservative value further assume that the network runs at 1M bits/sec. Then the transmission time of a packet is given by:

$$\bar{r} = 0.6 \text{ msec}$$

and the corresponding $v = 56$.

For this value of v , the packet loss rate (d^*) has been tabulated in Table 8.1 for different values of μ and ρ . It is seen that even for useful utilization (ρ) less than ten percent of its maximum value packet loss rate is below 2% limit. Thus we may conclude that performance of a CSMA/CD local area network is quite good in transmitting packetized voice, and these results agree with simulation and experimental studies. [17,27].

Finally results of this chapter may also be used to determine the average number of voice circuits (N) a CSMA/CD local area network may support at each useful utilization level (ρ) at a given delay. Substituting from (8.28, 8.29) into (8.20) and solving for N gives:

$$N = 2vp \quad (8.31)$$

Table 8.1 also gives the value of N at each ρ and packet loss rate (d^*). Clearly, the average number of voice circuits (N) which may be supported increases as ρ increase but at the expense of a higher packet loss rate. As μ decreases, N increases because less time is wasted in contention intervals. But for $\mu < 0.01$ it is seen that N is not sensitive to μ .

Figure 8.4 Mean Delay for the CSMA/CD Protocol
(with fixed packet size)

μ = Slot duration/packet transmission time

Figure 8.5 Variance of the Delay for CSMA/CD Protocol (with fixed packet size)
 — General Result
 ---- Heavy Traffic Result

Figure 8.6 Probability Distribution of the Delay for CSMA/CD Protocol (with fixed packet size)

— $u = 0.001$

- - - $u = 0.1$

μ = slot duration/packet transmission time

ρ = useful utilization

d/r = total delay/packet transmission time

μ	0.1		0.01		0.001	
$\rho_{\max}$	0.78		0.97		0.99	
ρ	0.6	0.7	0.6	0.9	0.6	0.9
%d*	0.02	0.54	0.00	0.87	0.00	0.35
N	67	78	67	100	67	100

$$v = \frac{t_p}{r} = 56$$

Table 8.1 Packet Loss Rate (%d*) and Average Number of Voice Circuits (N) a CSMA/CD local area network may support at a given utilization level (ρ), ($\rho_{\max}$ given by 8.22)

μ = slot duration/packet transmission time

v = packet interarrival time/packet transmission time

Chapter 9

CONCLUSIONS AND FURTHER WORK

9.1 Conclusions

This thesis is concerned with the development of network analysis tools for packetized voice and local area networks.

In the area of packetized voice, the packet delay, buffer size, and optimal packet size problems have been considered for three voice reconstruction protocols (FD, CD and TD) at the receiver. In general FD, CD (protocols discarding late packets) have resulted in smaller delay, and buffer size than TD (which waits for late packets). Results have been derived under the assumption of i.i.d. exponential or geometric network delay.

On buffer size, for a geometrically-distributed single-talkspurt, the buffer distribution, and its mean value were developed for all three protocols. Results for FD and CD were very close while TD resulted in larger values. When averaged over talkspurt lengths the tail of the curves for mean values follow the assumed talkspurt distribution. An approximate analysis of TD was also given with packet virtual circuits (fixed routing of the packets), which resulted in smaller buffer size due to the ordered arrivals. For a constant talkspurt length and a finite buffer length with the FD protocol, the buffer distribution and buffer overflow rate were found. A steady-

state analysis for a talkspurt stream was also done for the FD using a two state Markov chain as the speaker model to determine the long-term buffer occupancy. It was found that the empty buffer probability is quite high due to silent intervals which suggests sharing of the buffer space among several users.

On the delay issue, receiver delay, the total delay, and packet loss rate results were presented for all three protocols. While in the FD and CD these delays are independent of each other, in the TD later delays depend on the former ones. It was also observed that as the total delay is increased to achieve smaller packet loss rate in the FD and CD, and less frequent smaller gaps in the TD, the total delay in all three protocols approach to the same value.

Regarding optimal packet size, the total delay was minimized wrt packet size under a constraint on the packet loss rate (to ensure speech quality). For FD this problem was simplified to minimizing the fixed total delay wrt packet size under a constant packet loss rate. For CD, mean total delay was used instead of fixed delay in FD, and as expected minimization functions had very close forms. For TD because the later delays depend on the earlier ones, the total delay of the j 'th packet in the talkspurt was minimized wrt packet size. As a minimization function mean square total delay was used to account for the effects of both mean and variance. In general it was observed that the optimal packet size may not be a good operating point wrt the transmission efficiency. However a small increase in the packet size

may change the transmission efficiency substantially with little change on the total delay.

An approach for the design of packet voice system with the three protocols was also proposed, involving the trade-off among the system parameters to ensure acceptable total delay, buffer size, packet loss rate and transmission efficiency.

Regarding local area networks, a CSMA/CD protocol has been analysed to find the delay distribution and its moments, using a well known result. Simpler heavy traffic results for the delay distribution and its moments appear to be of practical use. As an application of these results, packetized voice on an Ethernet was considered; packet loss rate and number of voice channels to be supported by an Ethernet were calculated. Results show a small packet loss rate, thus we need to buffer only one packet at the transmitter and receiver.

9.2 Suggestions for Further Work

In the area of packetized voice there are several problems. For packet virtual circuit (fixed routing of packets) approach an approximate analysis was given with TD protocol, the need for a better analysis remains. Though in general the solution methods used in determining buffer distributions and optimal packet size may also be applied with nonexponential network delay distributions, the optimal packet size analysis of the TD protocol (chapter 7) is heavily dependent on the exponential assumption. Further, in determining

the optimal packet size for TD, mean square and variance of the total delay were given equal weights in the minimization function. The relative effects of the mean and variance on the speech quality need to be investigated for a better weighting factor.

On local area networks, heavy traffic behavior of the other CSMA/CD protocols should be compared with the one studied here by simulation.

APPENDIX 1

In this appendix, we derive the unconditional expected number of packets in the buffer for the protocol FD. Substituting (3.15) into (3.16) gives:

$$\begin{aligned} E\{k(n)\} &= \left[1 + \min(n, \tau_S) + \frac{q \min(n, \tau_S) + 1}{1 - q} \right] \sum_{\alpha=\max(0, n-\tau_S)}^{\infty} (1-\rho) \rho^\alpha \\ &\quad - \sum_{\alpha=\max(0, n-\tau_S)}^{n-1} (1-\rho) \rho^\alpha \left[n - \alpha + \frac{q^{n-\alpha}}{1 - q} \right] \\ &\quad - \sum_{\alpha=n}^{\infty} \frac{(1-\rho) \rho^\alpha}{1 - q} \end{aligned} \quad (A1.1)$$

Using the following formula:

$$\sum_{\alpha=\alpha_1}^{\alpha_2} (1-\rho) \rho^\alpha = \rho^{\alpha_1} - \rho^{\alpha_2+1} \quad (A1.2)$$

in the first and the last term above gives:

$$\begin{aligned} E\{k(n)\} &= \left[1 + \min(n, \tau_S) + \frac{q \min(n, \tau_S) + 1}{1 - q} \right] \rho^{\max(0, n-\tau_S)} \\ &\quad - \frac{\rho^n}{1 - q} - (1-\rho) \sum_{\alpha=\max(0, n-\tau_S)}^{n-1} \left[n \rho^\alpha - \alpha \rho^\alpha + \frac{q^n}{1 - q} \left(\frac{\rho}{q} \right)^\alpha \right] \end{aligned} \quad (A1.3)$$

Substituting (A1.2) and

$$\sum_{\alpha=\alpha_1}^{\alpha_2} \alpha \rho^\alpha = \frac{\rho}{(1-\rho)^2} \left[(\alpha_1 \rho^{\alpha_1-1} - (\alpha_2+1) \rho^{\alpha_2})(1-\rho) + \rho^{\alpha_1} - \rho^{\alpha_2+1} \right] \quad (A1.4)$$

into (A1.3), following the simplifications give:

$$E\{k(n)\} = \frac{\rho^{n+1}(1-q)}{(q-\rho)(1-\rho)} + \rho^{\max(0, n-\tau_S)} \left[\frac{1}{1-\rho} + \frac{q}{\rho-q} \frac{\min(n, \tau_S)+1}{\rho-q} \right] \quad (A1.5)$$

APPENDIX 2

In this appendix, we derive the conditional (on α) and unconditional expected number of packets in the buffer for the protocol CD. From (3.21):

$$\begin{aligned} E\{k(n)|\alpha\} = & \left[1 - \max(1, n-\alpha) - \frac{q \max(1, n-\alpha)}{1-q} \right] q^{\max(\mu+1, n-\alpha)-\mu-1} \\ & + \sum_{\tau_S=\max(\mu+1, n-\alpha)}^{\infty} (1-q)q^{\tau_S-\mu-1} \left[\min(n, \tau_S) + \frac{q \min(n, \tau_S)+1}{1-q} \right] \end{aligned} \quad (A2.1)$$

where (A1.2) has been used in simplifying the first term. Second term may be written as follows:

$$\begin{aligned} & \sum_{\tau_S=\max(\mu+1, n+1)}^{\infty} (1-q)q^{\tau_S-\mu-1} \left(n + \frac{q n+1}{1-q} \right) \quad n < \mu+1 \\ & \sum_{\tau_S=\max(\mu+1, n-\alpha)}^n (1-q)q^{\tau_S-\mu-1} \left[\tau_S + \frac{q \tau_S+1}{1-q} \right] \\ & + \sum_{\tau_S=\max(\mu+1, n+1)}^{\infty} (1-q)q^{\tau_S-\mu-1} \left(n + \frac{q n+1}{1-q} \right) \quad n \geq \mu+1 \end{aligned} \quad (A2.2)$$

After simplifying (A2.2) by using (A1.2) and (A1.4), and substituting the final result into (A2.1) gives:

$$E\{k(n)|\alpha\} = \begin{cases} \psi_A & n < \mu+1 \\ \psi_A + \left[\frac{1}{q} \max(\mu+1, n-\alpha) + \frac{1}{1-q} + \frac{q^{\max(\mu+1, n-\alpha)}}{1-q^2} \right] q^{\max(\mu+1, n-\alpha)-\mu} \\ - q^{n-\mu} \left[n + \frac{1}{1-q} \right] - \frac{q^{2n-\mu+2}}{1-q^2} & n \geq \mu+1 \end{cases} \quad (A2.3)$$

where

$$\begin{aligned} \psi_A = & \left[1 - \max(1, n-\alpha) - \frac{q^{\max(1, n-\alpha)}}{1-q} \right] q^{\max(\mu+1, n-\alpha)-\mu-1} \\ & + \left[n + \frac{q^{n+1}}{1-q} \right] q^{\max(0, n-\mu)} \end{aligned}$$

Unconditional expected number of packets in the buffer may be determined either by unconditioning (A2.3) wrt α , or (A1.5) wrt τ_S . The latter is followed for simplicity, and gives:

$$\begin{aligned} E\{k(n)\} = & \sum_{\tau_S=\mu+1}^{\infty} (1-q)q^{\tau_S-\mu-1} \left\{ \frac{\rho^{n+1}(1-q)}{(q-\rho)(1-\rho)} \right. \\ & \left. + \rho^{\max(0, n-\tau_S)} \left[\frac{1}{1-\rho} - \frac{q^{\min(n, \tau_S)+1}}{q-\rho} \right] \right\} \quad (A2.4) \end{aligned}$$

Again making use of (A1.2) and (A1.4) gives:

$$E(k(n)) = \begin{cases} \psi_B, & n \leq \mu+1; \\ \psi_B - \frac{1-q}{(1-\rho)(q-\rho)} (\rho^{n-\mu} - \rho q^{n-\mu-1}) \\ \quad - \frac{(1-q)}{(q-\rho)(\rho-q^2)} (q^{\mu+2} \rho^{n-\mu} - \rho q^{2n-\mu}), & n > \mu+1. \end{cases} \quad (A2.5)$$

where

$$\psi_B = \frac{\rho^{n+1}(1-q)}{(q-\rho)(1-\rho)} + \left[\frac{1}{1-\rho} - \frac{q^{n+1}}{q-\rho} \right] q^{\max(\mu+1, n)-\mu-1}$$

APPENDIX 3

In this appendix we determine the pdf of the receiver delay for protocol TD, using equations (4.8-4.11).

Substituting from (2.4) and (4.8) into (4.10) gives:

$$\begin{aligned} \overline{f(t_R)} = & \rho \int_{\max(\mu t_R)}^{\infty} \lambda e^{-\lambda(t_S - t_R)} \left\{ \frac{1 - \rho}{1 - \rho + \rho e^{-\lambda t_S}} \lambda e^{-\lambda t_S} e^{\lambda \mu} \right. \\ & \left. + \frac{\rho(1-\rho)\lambda e^{-\lambda t_S}}{[1 - \rho + \rho e^{-\lambda t_S}]^2} (1 - e^{-\lambda t_S} e^{\lambda \mu}) \right\} dt_S \end{aligned} \quad (A3.1)$$

The above integral is evaluated by change of variables to:

$$u = 1 + \frac{\rho}{1-\rho} e^{-\lambda t_S}$$

which results in:

$$\begin{aligned} \overline{f(t_R)} = & \lambda e^{\lambda t_R} (1-\rho) \left(1 + \frac{1-\rho}{\rho} e^{\lambda \mu t_p} \right) \left[\ln \left(1 + \frac{\rho}{1-\rho} e^{-\lambda \max(\mu t_p, t_R)} \right) \right. \\ & \left. - \frac{\rho e^{-\lambda \max(\mu t_p, t_R)}}{1 - \rho + \rho e^{-\lambda \max(\mu t_p, t_R)}} \right] \end{aligned} \quad (A3.2)$$

After evaluating (4.11) and substituting with (A3.2) into (4.9) gives the pdf of the receiver delay as:

$$\begin{aligned}
 p_R(t_R) = (1-\rho) \left(1 + \frac{1-\rho}{\rho} e^{\lambda \mu t_p} \right) & \left\{ \ln \left(1 + \frac{\rho}{1-\rho} e^{-\lambda \mu t_p} \right) \delta(t_R) \right. \\
 + \lambda e^{\lambda t_R} & \left[\ln \left(1 + \frac{\rho}{1-\rho} e^{-\lambda \max(\mu t_p, t_R)} \right) \right. \\
 - \frac{e^{-\lambda \max(\mu t_p, t_R)}}{1-\rho+\rho e^{-\lambda \max(\mu t_p, t_R)}} & \left. \left. \right] \right\} \quad (A3.3)
 \end{aligned}$$

REFERENCES

- [1] P.T. Brady, "A Model for Generating on-off Speech Patterns in two way Conversation", B.S.T.J., Vol. 48, pp. 2445-2472, 1969.
- [2] H. Frank, I. Gitman, "Economic Analysis of Integrated DOD Voice and Data Networks", National Telecommunications Conference, Dec. 1978.
- [3] G.J. Covielle, D.L. Lake, and G. Redinbo, "System Design Implications of Packetized Voice", Proc. 1977 Int. Conf. on Commun., p. 38.3.49 - 38.3.53.
- [4] J.W. Forgie, "Speech Transmission in Packet-Switched Store and Forward Networks", National Computer Conference, 1975, pp. 137-142.
- [5] G.J. Covielle, "Comparative Discussion of Circuit vs. Packet Switched Voice", IEEE Trans. on Comm., Vol. Com-27, pp. 1153-1159, 1979.
- [6] B. Gold, "Digital Speech Networks", Proc. IEEE, Vol. 65, pp. 1636-1658, 1977.
- [7] D. Minoli, "Packet Switched Networks I", Australian Electronics Engineering, pp. 48-52, April 1979.
- [8] A.W.F. Huggins, "Effect of Lost Packets on Speech Intelligibility", NSC Note No. 78, 1976.
- [9] D. Minoli, "Optimal Packet Length for Packet Voice Transmission", IEEE Trans. on Comm., Vol. Com-27, pp. 607-611, 1979.
- [10] G. Barberis and D. Pazzaglia, "Analysis and Optimal Design of a Packet Voice Receiver", IEEE Trans. on Comm., Vol. Com-28, pp. 217-227, 1980.
- [11] G. Barberis, "Buffer Sizing of a Packet Voice Receiver", IEEE Trans. on Comm., Vol. Com-29, pp. 152-156, 1981.
- [12] M. Gopal, J. Wong, J.C. Majithia, "An Evaluation of Playout Strategies for Voice Transmission in Packet Networks", Proc. of Computer Networking Symposium, Dec. 1981.
- [13] M. Suzuki, J. Kashio, R. Toki and T. Kato, "An Adaptive Control on Length of Jitter Absorbing Buffer in Packet Speech System", ICC. 1981, pp. 41.6.1-41.6.5

- [14] R.M. Metcalfe and D.R. Boggs, "Ethernet: Distributed Packet Switching for Local Computer Networks", Communications of the ACM, Vol. 19, pp. 395-403, 1976.
- [15] F. Tobagi and V.B. Hunt, "Performance Analysis of Carrier Sense Multiple Access with Collision Detection", Computer Networks, Vol. 4, pp. 245-259, 1980.
- [16] S.S. Lam, "A Carrier Sense Multiple Access Protocol for Local Networks", Computer Networks, Vol. 4, pp. 21-32, 1980.
- [17] D.H. Johnson and G.C. O'Leary, "A Local Access Network for Packetized Digital Voice Communication", IEEE Trans. on Comm., Vol. Com-29, pp. 679-688, 1981.
- [18] F.A. Tobagi, "Distributions of Packet Delay and Interdeparture Time in Slotted Aloha and Carrier Multiple Access", Journal of the ACM, October 1982.
- [19] T. Bially, A.J. McLaughlin, and C.J. Weinstein, "Voice Communication in Integrated Digital Voice and Data Networks", IEEE Trans. on Comm. Vol. Com-28, pp. 1478-1490, 1980.
- [20] F. Kamoun, L. Kleinrock, R. Muntz, "Queueing Analysis of the Ordering Issue in a Distributed Database Concurrency Control Mechanism", Second International Conference on Distributed Computing Systems, pp. 13-22, France, 1981.
- [21] C.S. Beightler, D.T. Phillips, D.J. Wilde, "Foundations of Optimization", Prentice-Hall, 1979.
- [22] J.A. Field and J.W. Wong, "A Carrier Sense Multiple Access (Collision Detection) System with Global Information", Comcon Fall 1982.
- [23] L. Kleinrock, "Queueing Systems, Volume 1: Theory", J. Wiley and Sons, 1975.
- [24] J.F.C. Kingman, "On Queues in Which Customers are Served in Random Order", Proceedings of the Cambridge Philosophical Society (58), pp. 79-91, 1961.
- [25] P. LeGall, "Les Systemes avec ou sans attente et les processus stochastiques", Tome 1 Paris Dunod, 1962. See for a summary: J.W. Cohen, The Single Server Queue, pp. 427-431, New York: Wiley Interscience, 1969.

- [26] A. Erdelyi, W. Magnus, F. Oberhettinger and F.G. Tricomi, "Table of Integral Transforms", Vol. 1, New York, 1954.
- [27] J.F. Shoch, "Carrying Voice Traffic Through an Ethernet Local Network - A General Overview", Xerox Palo Alto Research Center.
- [28] W.E. Naylor, L. Kleinrock, "Stream Traffic Communication in Packet Switched Networks: Destination Buffering Considerations", IEEE Trans. on Comm., Vol. Com-30, pp. 2527-2534, 1982.

END

0	5	1	0	9	1	8	4
---	---	---	---	---	---	---	---

FIN